Super-resolution microbubble localization in unprocessed ultrasound RF signals using a 1D dilated CNN

Nathan Blanken[1] N.BLANKEN@UTWENTE.NL
Jelmer M. Wolterink[2] J.M.WOLTERINK@UTWENTE.NL
Hervé Delingette[3] HERVE.DELINGETTE@INRIA.FR
Christoph Brune[2] C.BRUNE@UTWENTE.NL
Michel Versluis[1] M.VERSLUIS@UTWENTE.NL
Guillaume Lajoinie[1] G.P.R.LAJOINIE@UTWENTE.NL
[1] *Physics of Fluids, MESA+ Institute for Nanotechnology, Technical Medical Centre, University of Twente, Enschede, The Netherlands*
[2] *Applied Mathematics, Technical Medical Centre, University of Twente, Enschede, The Netherlands*
[3] *EPIONE, INRIA, Sophia Antipolis, France*

Editors: Under Review for MIDL 2022

Abstract

We present a super-resolution ultrasound approach based on direct deconvolution of single-channel ultrasound radio-frequency (RF) signals with a one-dimensional dilated convolutional neural network (CNN). Data are generated with a physics-based simulator that simulates the echoes from a dense cloud of monodisperse microbubbles and captures the full, nonlinear response of resonant, lipid-coated microbubbles. The network is trained with a novel dual-loss function, which features elements of both a classification loss and a regression loss and improves the detection-localization characteristics of the output. The potential of the presented approach to super-resolution ultrasound imaging is demonstrated with a delay-and-sum reconstruction with deconvolved ultrasound data. The resulting image shows an order-of-magnitude gain in axial resolution compared to a delay-and-sum reconstruction with unprocessed element data.

Keywords: Contrast-enhanced ultrasound, deconvolution, dilated convolutional neural network, microbubbles, super-resolution.

1. Introduction

Recently, super-resolution ultrasound imaging with ultrasound localization microscopy (ULM) has received much attention (Errico et al., 2015). In ULM, the microvasculature is imaged by tracking sparsely distributed microbubbles (contrast agents) flowing through the capillaries. The low bubble concentration ensures the separation of their point spread functions but results in long acquisition times. As ULM often relies on reconstructed images, much of the information content in the raw ultrasound data is unused. Here, we present an alternative approach based on direct deconvolution of unprocessed ultrasound RF data (i.e. *before* image reconstruction). This is an extended abstract of a full paper that was recently accepted for publication in *IEEE Transactions on Medical Imaging* (Blanken et al., 2022).

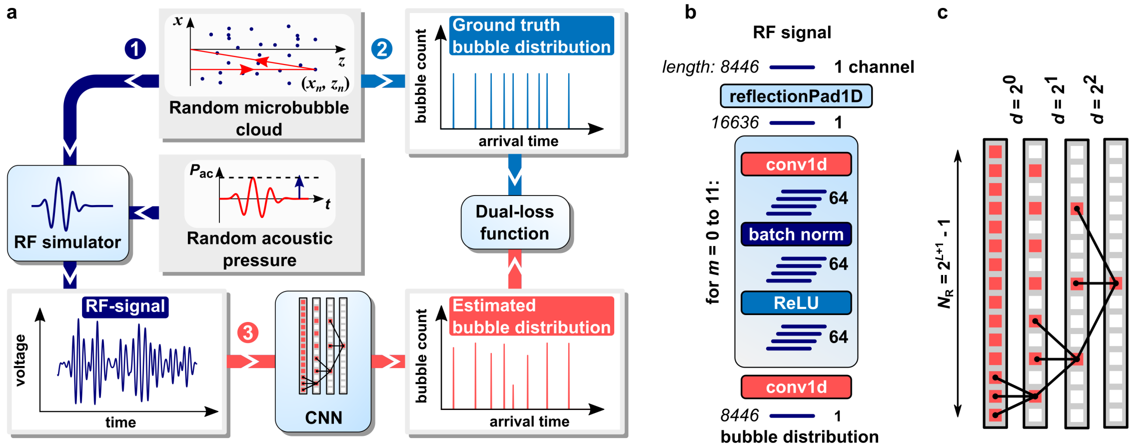

Figure 1: **a** Generation of (1) 1D RF signals and (2) ground truth arrival times of bubble echoes and (3) CNN training. **b** CNN architecture. **c**. Exponential expansion of the receptive field N_R as a function of the number of convolutional layers L.

2. Methods

Figure 1 summarizes the methodology. We generate data for training, validation, and testing with a simulator that simulates the nonlinear propagation of a plane wave emitted by a multi-element transducer (1.7 MHz) and the resulting nonlinear response from a dense microbubble cloud. The arrival times of the individual microbubble echoes serve as 1D ground truth data. The high sampling rate of 62.5 MHz allows for super-resolved microbubble localization. An RF signal U can be regarded as a convolution of the ground truth φ with an individual microbubble echo (a variable convolution kernel). The RF signals contain up to 1000 microbubbles echoes, corresponding to an average echo overlap of 94%.

To recover φ from U, we use a dilated CNN (Figure 1). The fine signal sampling results in long signals (8446 grid points) with large-scale features, requiring a large receptive field. A dilated CNN preserves translational equivariance and has a receptive field that increases exponentially with network depth, without loss of resolution or coverage (Yu and Koltun, 2016). The neural network is trained with a dual-loss function $\mathcal{L} = \varepsilon_1 L_1(\varphi^*, \hat{\varphi}^*) + \varepsilon_2 \mathrm{DL}(\varphi^\dagger, \hat{\varphi}^\dagger)$, which is a linear combination of an L_1 regression loss and a Dice loss DL. Here, φ^* are *soft* labels, generated by convolving φ with a Gaussian kernel, and $\varphi^\dagger$ are *hard* labels, a binary version φ. The hats denote predictions, and ε_1 and ε_2 are tunable proportionality constants. L_1 enforces a high degree of bubble localization on the predictions, whereas DL enforces a high bubble detection rate.

3. Results

Figure 2a shows a typical example of an input RF signal, its ground truth bubble distribution φ, and the output $\hat{\varphi}$ of a trained model for this signal. For a quantitative analysis of the