Normalizing flows for acceleration of transcranial ultrasound computed tomography with uncertainty quantification

Rafael Orozco[1] ROROZCO@GATECH.EDU
Mathias Louboutin[1] MLOUBOUTIN3@GATECH.EDU
Ali Siahkoohi[1] ASIAKOOHI@GATECH.EDU
Gabrio Rizzuti[1] @.EDU
Tristan van leeuwen[1] T@GATECH.EDU
Felix Herrmann[1] FELIX.HERRMANN@GATECH.EDU
[1] *Georgia Institute of Technology*

Editors: Under Review for MIDL 2022

Abstract

We present a novel approach to transcranial ultrasound computed tomography (TUCT) that utilizes normalizing flows to improve the speed of imaging and provide uncertainty quantification. Our method combines physics-based methods with data-driven methods to accelerate the reconstruction of the final image. The neural network is based on normalizing flows that allows us to calculate Bayesian uncertainty in the resulting TUCT images. We introduce the concept of a physical summary statistic to compress incoming observations. This compression enables efficient training of the neural network and also standardizes the size of the data regardless of imaging configurations. We evaluate our approach with synthetic experiments and demonstrate that it can significantly improve the speed of TUCT while giving uncertainty quantification. We validate the quality of our image reconstructions and the recovered uncertainty with quality metrics and different method baselines.

Keywords: Invertible Networks, Medical Imaging, Bayesian, Uncertainty quantification, Physics and Machine Learning Hybrid

1. Problem Statement

Transcranial ultrasound computed tomography (TUCT) is a non-invasive, non-toxic imaging technique that can aims to create high resolution images of clinically relevant structures in the brain by transmission and reception of acoustic wavefeilds. Traditional b-mode has been studied but does not show promise. Instead more recent approaches Guasch et al. (2020) have shown that modeling all aspects of the wave enables higher resolution. Current wave equation based imaging methods are computationally expensive since they require the application of forward and gradient operators. This limits the clinical use of these methods since they can take up 36 hours Guasch et al. (2020) to form an image. In addition, the imaging process is affected by incomplete measurements, noise and other sources of uncertainty, that can limit the accuracy and reliability of TUCT. In this paper, we propose a novel, data-driven approach to TUCT that utilizes normalizing flows to improve the speed of imaging without sacrificing accuracy and provide uncertainty quantification.

The rest of this paper is organized as follows. In Section 1, we review the related literature and describe the background and context of our work. In Section 3, we present

© 2022 R. Orozco, M. Louboutin, A. Siahkoohi, G. Rizzuti, T.v. leeuwen & F. Herrmann.

our proposed approach and provide details about our experimental setup. In Section 4, we present our results and discuss their implications. Finally, in Section 5, we conclude with a summary of our findings and suggestions for future work.

2. Related work

Gaush et al Guasch et al. (2020) pioneered using FWI methods for transcranial ultrasound computed tomography

Cudeiro-Blanco et al. (2022) Discuss design decisions needed to recover soft tissue and bone.

Taskin et al. (2020) Explain the importance of a good background model for normal Least square approaches based on linearization.

Donaldson and He (2021) Learn a cnn to post process after a single iteration of FWI

Tong et al. (2022) learned some decompositions. Showed they could accelerate FWI and image blot clots.

Marty et al. (2021) demonstrate different receiver geometries favor transmission vs reflection.

Closely related to our work (**?**) used the mean-field approximation to calculate uncertainty assuming. Our work, accomplishes similar results while making no underlining assumptions on the posterior/prior distributions and requires only one set of forward/gradient calls.

3. Methods

3.1. Transcranial ultrasound computed tomography

Transcranial ultrasound computed tomography (TUCT) To model the propagation of the acoustic waves, we use the second-order wave equation in time that solves the partial differential equation (PDE):

$$\frac{1}{\rho(x,y)c(x,y)^2}\frac{\partial^2}{\partial t^2}u(x,y,t) - \nabla \cdot \frac{1}{\rho(x,y)}\nabla u(x,y,t) = q(t,x,y). \tag{1}$$

In equation 1, $\rho(x,y)$ represents the field of density as a function of two coordinates of space, $c(x,y)$is acoustic velocity, $u(x,y,t)$ is the acoustic pressure measured at a point in space and time, ∇ is the derivative in space and $q(t,x,y)$ is the acoustic source that is defined by the transducer source. We can express the forward process (solving the PDE 1) by the non-linear operator $\mathcal{F}$ parameterized by the acoustic values $\mathbf{x} = \{c(x,y), \rho(x,y)\}$:

$$\mathbf{y} = \mathcal{F}(\mathbf{x}) + \epsilon \tag{2}$$

We include ϵ to include effects in the experiment that lead to additive noise in the observed acoustic data. Figure shows an example of acoustic observations. The goal of (TUCT) is to form a high-resolution image of $\mathbf{x}$ from acoustic measurements. For imaging purposes, the unknown parameter of interest is acoustic impedance $\rho(x,y)c(x,y)$. This imaging process is an inverse problem that we will solve in a Bayesian framework so that we can measure uncertainty in the receiver geometry, modeling errors and noise.

3.2. Bayesian approach to inverse problem

Tomographic methods for medical imaging are fundamentally an inverse problem. Where the goal is to estimate the internal structures x that caused the forward operator F to produce measurements y from the injected energy () inpinging on the internal structures. This forward process, is normally called the forward problem and is described by the following equation:

Upon receiving indirect observations y, the goal of Bayesian inverse problem solving is to calculate the conditional distribution on x given y. This conditional distribution $p(x|y)$ is called the posterior distribution. It contains all the information needed to estimate x and the uncertainty surrounding that estimate. The posterior $p(x|y)$ encodes two important prior assumptions that are given by writing out Bayes theorem:

The first is $p(x)$ that describes the prior assumptions we make on the ground truth x. Next is the conditional distribution $p(y|x)$ that describes the forward process we assume is used to create the observation y. In general, the posterior distribution $p(x|y)$ is expensive to calculate. In fact, traditional methods like MCMC would require many thousand iterations. This makes these methods inpractical for clinical use that require fast results. In this paper, we suggest a viable method is conditional normalizing flows a variational inference method that takes advantage of the distribution learning capabilities of deep learning models.

3.3. Conditional normalizing flows for posterior distribution

Scale exponentially with the dimension of the data and image. But normalizing flows scale favorably with dimension making our task tractable for approximate distribution calculation (NEED CITATINO!!). One practical explanation for the efficiency of for large dimensions is Multiscale splitting introduced in Real-NVP (Dinh et al., 2016) as this helps normalizing flows to scale to high-dimensional data (Bond-Taylor et al., 2021).

Are trained with the following objective:

$$\hat{\theta} \ = \arg\min_{\theta} \frac{1}{N} \sum_{n=1}^{N} \left(\|f_\theta(\mathbf{x}^{(n)}; \mathbf{y}^{(n)})\|_2^2 - \log |\det \mathbf{J}_{f_\theta}| \right) \tag{3}$$

As is common in Normalizing flows: where networks are constructed such that their inverse are easy to calculate, requiring the same cost as the forward pass. Also the Jacobian determinant is calculated exactly with linear cost in the size of parameters of x. Sampling the posterior distribution from a conditional normalizing flow is fast as it only requires an inverse pass on a neural network.

3.4. Score as a summary statistic

The observation y is large, in our case is acoustic data of the size $R_t^N * N_r * N_s$ (about 16 million). To make our algorithm practical, we would like to compress the data [cite the simulation based inference paper] such that it retains information relevant to the inference task at hand. In other words, how much can we compress y such that it does not lose information that could inform the estimation of x?. This is an important consideration when streaming data could put strain on the Input/output system. In our data driven case, this is of particular concern since the raw data is too large to fit in consumer GPU for

practical training purposes. We propose to follow [cite massively compressed] that shows that the score is asymptotically maximally informative of x. The score in this paper is defined as the gradient of the log likeilihood with respect to x calculated at a fiduciary point x_0. Importantly, the score achieves asymptotically refers to two things, firstly how close the fiduciary point x_0 is to the ground truth x and secondly how close the assumed likelihood is to the ground truth one. Deviations from these two assumptions, will produce a summary statistic that is uninformative. In the results section, we will qualitatively explore this how the uninformativeness of the is reflected in the quality of the inference. In section of conclusions, We will also discuss future work that could allow for larger deviations from this assumption.

In figure **??** we show the assumed fiduciary point x0 for an example slice, the resulting score and the ground truth model. We acknowledge that this preimaging step requires our algorithm to not be completely realtime. This gradient is well known in the seismic imaging community as the reverse time migration (cite rtm papers (could use some help here from the more seismic people))

The ground truth likihood in our synthetic case is related to the noise model that we used to simulate our forward data. We used colored noise that was made by bandpassing the noise with the frequency content of the wavelet. This modeling corresponds to a non-isotropic gaussian given by the convolution with the wavelet. The likelihood $p()$ we assume to calculate the score is of an isotropic Gaussian. This is a deviation from the true likelihood but we did not notice a degradation in quality.

$$\hat{\theta},\, \hat{\phi} = \arg\min_{\theta\,\phi} \frac{1}{N} \sum_{n=1}^{N} \left(\|f_\theta(\mathbf{x}^{(n)}; h_\phi(\bar{\mathbf{y}}^{(n)}))\|_2^2 - \log|\det \mathbf{J}_{f_\theta}| \right) \tag{4}$$

Can cite SPIE paper to show that it reduces training time.

4. Experiments and Results

4.1. Training dataset

To train our method, we require samples from the prior distribution $p_{\mathbf{x}}(\mathbf{x})$ on ground truth brain acoustic values. To make these samples, we start from the FASTMRI dataset (Zbontar et al., 2018) that contains MRI images of human brains. There is no immediate relationship between MRI intensity and acoustic speed values. As a heuristic, we took the acoustic values of the main brain tissues (Brain Grey Matter=1.505, Cerebellum White Matter=1.552, Blood Veins=1.5782]), and used k-means to assign acoustic values to MRI intensities. This process is automatic and we generated 2D slices of acoustic values from 250 brains. To validate how realistic the models are we compared with the MIDA model (Iacono et al., 2015). The MIDA model represents a gold standard since it was created by a team of experts. Since the MIDA model requires so much expert time, there is only one model which is not enough to train a data-driven method. In appendix, figure we compare a slice from the MIDA model with one of our models derived from MRI intensities. We also compare the acoustic data simulated in both slices. These appear to be similar enough to merit using our MRI derived models. In future work, we will explore more complicated

workflows to produce these acoustic models required for training. 250 brains Out of these we use 90% to train 5% for validation and and 5% for testing.

We simulated the forward wave propagations using Devito (Louboutin et al., 2019) and JUDI (Witte et al., 2019). Devito is a Domain Specific Language that allows to define PDE equations that are automatically compiled into high-performant C stencils. JUDI, The combination of these two software makes our method quick to prototype while insuring the high performance needed for fast clinical results.

To algorithmically find the skull we used BET software (Smith, 2002).

Algorithm for pretraining: need: N_{train} samples from the prior on ground truth models p(x) Forward simulation and Gradients step 1: make data y=F(x)+e step 2: calculate gradient $\bar{y}$ step 3: train conditional normalizing flow

Algorithm on inference: Need starting model x0 step 1: calculate gradient $\bar{y}$ step 2: sample N_{post} Gaussian normal noise z step 3: pass N_{post} z through Inverse of conditional normalizing flow to generate posterior samples.

4.2. Conditional normalizing flow training

The conditional normalizing flow was implemented in the julia package InvertibleNetworks.jl(Witte et al., 2020). The learned summary network h_ϕ, we chose was a UNet (Ronneberger et al., 2015). We pregenerated the physical summary statiscs $\bar{y}$ to reduce computational cost of training.

We trained on a 32GB A100. Each epoch took 9 minutes and we trained until the objective stopped decreasing on a leave out validation set. Total training time was 4.5 hours. We did not observe severe overfitting as shown in the training/validation curve Supplemental figure.

4.3. Evaluating image reconstruction

The computational cost is dominated by the calculation of the physical summary statistic $\bar{y}$. This calculation took on our hardware 1 second (). This calculation only needs to be calculated once per observation, then many samples can be generated from the posterior at the cost of one inverse network evaluation (10ms per sample). The total time required for N_{post} is $\approx$ 1+N$_{post}$ $\times$ 0.01 for a total of 1.64seconds with our hardware.

Since we can sample from the posterior, we can generate different point estimates such as the posterior mean, or the MAP. To calculate the posterior mean we take the average of 64 samples of the posterior. See supplemental section for an analysis of the quality of the posterior mean (NEED TO DEFINE POSTERIOR MEAN) as the number of posterior samples increase. In this work, we concentrate on the posterior mean because it is proven to be the estimator with less error on expectation (Whang et al., 2021).

As a comparison point for the posterior mean we show the FWI result and UNet. Note that the FWI is computationally intensive, requiring 40 calls to the forward and gradient of physics model to converge while our method only requires one gradient call. The UNet also only requires one gradient call. Both FWI and UNet only provide the shown point estimate while our method also gives uncertainties.

By testing the methods an unseen slice, in figure (Figure showing images of reconstructions compared to ground truth, FWI, and unet) we can see that our method is visually

closer to the ground truth than the other compared methods. We show PSNR, SSIM and the MSE as quality metrics between the methods. Importantly, note that in subfigure, out method can plot the posterior samples that assigns a 1-σ confidence band to the error that our method makes.

For a more quantitative comparison, we take 50 unseen test slices and compute the quality metrics. From the table, we can see that

4.4. Evaluating sensitivity to model of skull

An important consideration of our method is that since it is gradient based, we need a fiduciary point at which to calculate the gradient. we assume that we have access to an acoustically correct model of the skull. Practically, we envisage that by using x-ray based computed tomography (CT) we will build an acoustic model of the patients brain that we can then use to invert for the acoustic properties of the inside brain tissue. The process of recovering acoustic properties of bone from CT measurements is well documented [cite the main one and also cite costas]. Acoustic properties of bone are well recovered by CT but the acoustic properties of soft tissue (the imaging goal of our method) are not recovered well by CT.

Since our wave physics model is non-linear, the process will be sensitive to the fiduciary point used to calculate the gradient. A phenomena well understood in the seismic imaging community where much work is dedicated to designing good fiduciary points (background or starting models). Here we study the effect of this starting model on the result of our method. We numerically study this effect in two scenarios: firstly with constant deviations from our recovered model vs the true model used to make acoustic observations, we plot the error made in +- 0.1m/s. Secondly, we study the effect of speckle noise type errors.

Unsurprisingly, our method degrades in quality as the fiduciary point degrades. This behaviour is expected and unavoidable since our method is gradient based and non-linear. However, in the future work section we will discuss potential avenues to ameliorate this degradation.

4.5. Evaluating sensitivity to noise

Gaush et al observed SNR of 34db in a lab setting so we need to show that our method can be robust to noise up to that level. During training we add noise to our data in that level and up to 15snr. We plot in figure **??** the quality of the conditional mean reconstruction of our method vs increasing the noise level from 35db to 15db. As expected there is a decrease in quality but still well within reason. Importantly, the certainty of our estimate also decreases pointing to correct calibration of our uncertainty quantification.

4.6. TASK: tumor segmentation

To further validate the practical use of our method, we generate brain slices with synthetic tumors. Our tumor models are based on [cite photoacoustic tumor paper] by generating tumor phantoms that have an edema of acoustic value and a tumor part with acoustic value. Note that these phantoms are not seen by the learned methods on training time. They are only seen on test time. This provides important insight on if the learned model is learning a bias towards healthy brains that is detrimental to downstream tasks on unhealthy brains.

4.7. TASK: Anomaly detection

Normalizing flows provide two purposes, they can sample from distributions and we can do exact likelihood evaluation. Using the likelihood we can detect anomalous events in the observed data y. Since our method was trained on healthy brains then pathalogies (ie tumor, embolism, hemmorages) will be out of distribution and detectable by our method. For validation purposes, we create a dataset of synthetic tumors. In Figure we plot the density histograms of a leave-out set of healthy brains as compared to an unseen set of tumor brains. We see considerable separation of the datasets pointing to the ability of our learned likelihoods to create a criterion for detection of anomalous brains.

4.8. Evaluating uncertainty

We evaluate the calibration. The size of our problem its non-linearity and the non-gaussianity of our prior prevents us from a comparison with any ground truth posterior. Instead, we follow the literature and evaluate our method using a variety of metrics designed the analyse the validity of the posterior from a Bayesian sense and also a practical sense. As a comparison point, we use MC-dropout (**?**). We note that the statistical literature is currently debating whether this technique is truly Bayesian (**??**). Nonetheless, we use MC-dropout method as a comparison since its computational cost is similar to our method. Both methods require the evaluation of a trained neural network for each posterior sample.

We use the following metrics to evaluate the quality of our uncertainty: The ECE calibration: On expectation, the error made by our method should correlate with the uncertainty.

The dice overlap of the error and uncertainty made by our models (**?**)

Bayesian contraction: a Bayesian method needs to show contraction on the ground truth as more data is observed. Here contraction means that more data should decrease the uncertainty. Not only that, the error made should also decrease. Qualitatively we confirm this behaviour in figure where increasing the number of source experiments decreases the overall uncertainty. As a quantitative measure we use the sum of variance for all parameters. Figures shows that over the test set, increasing sources shows the proper Bayesian contraction. Also, increasing the number of receivers does the same as seen in Figure.

Results and conclusions:

References

Sam Bond-Taylor, Adam Leach, Yang Long, and Chris G Willcocks. Deep generative modelling: A comparative review of vaes, gans, normalizing flows, energy-based and autoregressive models. *arXiv preprint arXiv:2103.04922*, 2021.

Javier Cudeiro-Blanco, Carlos Cueto, Oscar Bates, George Strong, Tom Robins, Matthieu Toulemonde, Mike Warner, Meng-Xing Tang, Oscar Calderón Agudo, and Lluis Guasch. Design and construction of a low-frequency ultrasound acquisition device for 2-d brain imaging using full-waveform inversion. *Ultrasound in Medicine & Biology*, 48(10):1995–2008, 2022.

Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real nvp. *arXiv preprint arXiv:1605.08803*, 2016.

Robert W Donaldson and Jiaze He. Instantaneous ultrasound computed tomography using deep convolutional neural networks. In *Health Monitoring of Structural and Biological Systems XV*, volume 11593, pages 396–405. SPIE, 2021.

Lluís Guasch, Oscar Calderón Agudo, Meng-Xing Tang, Parashkev Nachev, and Michael Warner. Full-waveform inversion imaging of the human brain. *NPJ digital medicine*, 3 (1):1–12, 2020.

Maria Ida Iacono, Esra Neufeld, Esther Akinnagbe, Kelsey Bower, Johanna Wolf, Ioannis Vogiatzis Oikonomidis, Deepika Sharma, Bryn Lloyd, Bertram J Wilm, Michael Wyss, et al. Mida: a multimodal imaging-based detailed anatomical model of the human head and neck. *PloS one*, 10(4):e0124126, 2015.

Mathias Louboutin, Michael Lange, Fabio Luporini, Navjot Kukreja, Philipp A Witte, Felix J Herrmann, Paulius Velesko, and Gerard J Gorman. Devito (v3. 1.0): an embedded domain-specific language for finite differences and geophysical exploration. *Geoscientific Model Development*, 12(3):1165–1187, 2019.

Patrick Marty, Christian Boehm, and Andreas Fichtner. Acoustoelastic full-waveform inversion for transcranial ultrasound computed tomography. In *Medical Imaging 2021: Ultrasonic Imaging and Tomography*, volume 11602, pages 210–229. SPIE, 2021.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

Stephen M Smith. Fast robust automated brain extraction. *Human brain mapping*, 17(3): 143–155, 2002.

Ulas Taskin, Kjersti Solberg Eikrem, Geir Nævdal, Morten Jakobsen, Dirk J Verschuur, and Koen WA Van Dongen. Ultrasound imaging of the brain using full-waveform inversion. In *2020 IEEE International Ultrasonics Symposium (IUS)*, pages 1–4. IEEE, 2020.

Junkai Tong, Xiaocen Wang, Jiahao Ren, Min Lin, Jian Li, He Sun, Feng Yin, Lin Liang, and Yang Liu. Transcranial ultrasound imaging with decomposition descent learning-based full waveform inversion. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 69(12):3297–3307, 2022.

Jay Whang, Erik Lindgren, and Alex Dimakis. Composing normalizing flows for inverse problems. In *International Conference on Machine Learning*, pages 11158–11169. PMLR, 2021.

Philipp Witte, Gabrio Rizzuti, Mathias Louboutin, Ali Siahkoohi, and Felix Herrmann. Invertiblenetworks. jl: A julia framework for invertible neural networks, 2020.

Philipp A Witte, Mathias Louboutin, Navjot Kukreja, Fabio Luporini, Michael Lange, Gerard J Gorman, and Felix J Herrmann. A large-scale framework for symbolic implementations of seismic inversion algorithms in juliaa symbolic seismic inversion framework. *Geophysics*, 84(3):F57–F71, 2019.

Jure Zbontar, Florian Knoll, Anuroop Sriram, Tullie Murrell, Zhengnan Huang, Matthew J Muckley, Aaron Defazio, Ruben Stern, Patricia Johnson, Mary Bruno, et al. fastmri: An open dataset and benchmarks for accelerated mri. *arXiv preprint arXiv:1811.08839*, 2018.