Solving the visibility problem in photoacoustic imaging with a deep learning approach providing prediction uncertainties

Guillaume Godefroy\textsuperscript{a,}\textsuperscript{*}, Bastien Arnal\textsuperscript{a}, Emmanuel Bossy\textsuperscript{a}

\textsuperscript{a}Univ. Grenoble Alpes, CNRS, LIPhy, 140 rue de la Physique, CS 40700, 38058 Grenoble CEDEX 9, FRANCE

Abstract

Photoacoustic imaging is a promising biomedical imaging modality providing optical contrasts at depth. Conventional reconstructions suffer from the limited view and bandwidth of the ultrasound transducers. As a result, structures elongated in the axis of the probe or large compared to the point spread function of the system are not fully recovered. A deep learning approach is proposed to handle these problems and is demonstrated both in simulations and in experiments on a multi-scale model of leaf skeleton. We employed an experimental approach to build the training and the test sets using photographs of the samples as ground truth images. Reconstructions produced by the neural network show a greatly improved image quality as compared to conventional approaches. In addition, this work aimed at quantifying the reliability of the neural network predictions. To achieve this, the dropout Monte-Carlo procedure is applied to estimate a pixel-wise degree of confidence on each predicted picture. This degree of confidence is well correlated to the variability observed in certain regions, among a stack of subsequently acquired frames. Last, we address the possibility to use transfer learning with simulated data in order to drastically limit the size of the experimental dataset.

Keywords: Photoacoustic imaging; Deep learning; Visibility artefacts; Monte Carlo dropout; Bayesian neural network

1. Introduction

Photoacoustic (PA) imaging is an emerging biomedical modality based on the generation of acoustic waves by light absorption. This modality is promising, as it enables imaging at depth with high spatial and temporal resolution, and can provide images of the optical absorption \[1\] with specific molecular contrast which can be enhanced by spectroscopy.

In conventional PA imaging, a short nanosecond laser pulse is sent into the medium and the emitted ultrasonic waves are collected by a conventional ultrasound (US) probe. At the US propagation time scale, the object illumination is quasi instantaneous as the speed of light is several orders of magnitude higher than the speed of sound, resulting in the emission of strongly coherent acoustics waves \[2\]. These waves will interfere constructively or destructively depending of the structure of the object, often leading to two well-known artefacts on the reconstructed image: the limited bandwidth and the limited view artefacts \[3\].

In the first case, when a large object compared to the bandwidth of the transducer is illuminated, PA signals are produced with a strong low frequency component that is filtered out by the transfer function of the probe. In the second, for a structure elongated in the axis of the probe, the produced waves interfere constructively perpendicularly to the probe but mostly destructively throughout the elongation. As a result, very few signals are collected by linear or matrix array probes due to their limited angular view. Both artefacts will further be referred to as the visibility problem in this paper.

The limited view problem has been addressed in several studies. The most intuitive approach is to either rotate the object relatively to the probe \[4\] (or vice versa \[5\]) or use ring shaped transducer arrays \[6\]\[3\] in order to cover all angles. However, a clinical implementation would benefit from a handheld real-time system as currently used in ultrasound imaging.

Other approaches rely on the introduction of a spatial modulation of optical absorption of the sample, either through injection of sparse absorbing particles \[7\], by a modulation of physical properties \[8\], or by computing statistical properties of the PA signal generated by fluctuating sources in the medium. The fluctuations of the PA signal can be obtained by using random optical speckle pattern illuminations \[9\] or from flowing red blood cells \[10\], naturally present in the blood vessels. Nevertheless, these methods require long acquisition times in order to get significant statistical properties, which decreases temporal resolution.

In this work, a deep learning approach is proposed to overcome the visibility problem and improve the image...
quality in a real-time single shot configuration. A neural network can be viewed as an algorithm composed of many parameters, called weights, designed to faithfully derive given input data into a desired form of output \[11\]. This algorithm is trained over multiple examples to obtain the best representation of the studied phenomenon. After training, the network transforms the input raw image into an output image that should resemble the ground truth. The training set consists of multiple raw data/ground truth pairs that will be used to optimize the weights of the network.

Convolutional neural network networks (CNN) \[12\] are the most popular category of deep learning algorithms (DLA), and have reached state of the art performances in several imaging problems including segmentation \[13\], classification \[14\], artefacts removal \[15\] or denoising \[16\]. CNN have been introduced recently in biomedical imaging, showing impressive results in various tasks \[17\] \[18\] \[19\]. Over the past two years, some groups already investigated deep learning applied to PA imaging for several purposes including direct reconstruction of the initial pressure \[20\], handling artefacts coming from sparse data \[21\] \[22\] \[23\], reflection artefacts removal \[24\], point source localization \[25\] \[26\] and quantitative measurements \[27\] \[28\]. The correction of the limited bandwidth problem was also investigated on very simple objects \[29\]. Some of these studies \[30\] \[22\] \[23\] showed that deep learning can also reduce the limited view artefacts although results were either numerical or obtained with non-conventional imaging devices. A linear array was experimentally used in \[31\] but a ground truth was missing to assess the success of the approach. Finally, in most of the cited studies, experimental results were predicted from models trained only on simulation data, producing a less accurate reconstruction.

In this issue, we focus on the correction of the whole visibility problem, induced both by the limited view and limited bandwidth of a conventional linear US probe. The originality of our approach resides in the design of a dedicated model object and a method to create an experimental training dataset. The method is used to clearly assess the capacity of a neural network to remove these artefacts on experimental images that were not used during the training. In this study, the ground truth is known for the test set, which consisted of some of those unseen images. Thus, quantitative evaluation of the quality of the reconstruction can be done.

Despite the impressive performance of DLA to reconstruct PA images, errors can be made by the algorithm which may misinterpret the data. This is one of the main limitations of neural network approaches in the medical field: the lack of confidence in the results. In this work, we estimate the uncertainty in our prediction through Bayesian machine learning framework. We followed the approach proposed by Ghahramani and Gal \[32\], referred as Monte Carlo dropout (MC dropout), which has been recently applied for phase imaging \[33\]. Our CNN is converted into a Bayesian neural network to introduce randomness in the prediction process, which makes the prediction no longer deterministic: the model will predict different outputs for the same input. Then, for a given input, several outputs are generated and are interpreted as samples of a probabilistic distribution, from which parameters can be estimated, such as the mean value and the confidence measure.

The quantification of this uncertainty could be very helpful for real-time navigation as a feedback for the user, who may eventually choose to display only the reliable parts of the images.

We also studied the DLA performance over different input data types. Usually, a conventionally reconstructed image is used as input instead of the pressure time series. The DLA will thus focus on learning to correct the artefacts instead of encoding the PA forward operation. This prior reconstruction can be obtained by applying delays and summation (DAS) on the radiofrequency (RF) signals. Envelope images are usually displayed for the end user, since it better represents the object. The demodulation can be performed by generating the quadrature component of the RF time signals using an Hilbert transform before applying DAS and the modulus of the obtained complex image corresponds to the envelope image. While the input of the DLA for PA image reconstruction is usually the envelope image, we choose to train our network with the RF image. The RF image (Fig 1.a) is obtained by applying DAS directly on the real-valued RF time signals. The RF image is modulated by the impulse response of the transducer, resulting in axial oscillations. We show the RF image carries more information than the envelope image. The two approaches are compared in Supplementary Materials.

Finally, we investigated the design of the training sets. Indeed, processing experimental data with CNN trained solely on simulated data seems to produce poor reconstruction \[23\] which we confirm here, while constructing a large experimental dataset is complex and time consuming. We varied the relative sizes of the combined experimental and simulated datasets and observed its impact on the reconstruction performances.

2. Material and methods

2.1. Conventional reconstruction methods

For comparison purpose, DAS envelope image and L2 deconvolution image are provided. DAS is fast and robust whereas deconvolution methods are more computational and more complex to implement since the knowledge of
Figure 1: a, Creation of the experimental training set. A linear probe is coupled to a water tank containing the leaf through a window composed of a tight Mylar membrane. The leaf is in the imaging plane of the probe. The laser beam is shined from the top and the RF data are acquired. A photograph of the leaf was previously taken. The RF PA image of the ROI is reconstructed and the photograph is processed to extract the same area. b, Uncertainty prediction: Several images are generated using the same input. The mean and the standard deviation (std) of these samples are estimated pixel by pixel. The prediction is unstable in the marked area, resulting in a high std.

As shown in Fig. 1.a, the leaf is maintained horizontal into an agarose gel which stands in a tank filled with degassed and deionized water. Through a side window composed of a frame tightening a Mylar membrane, an ultrasonic transducer array (15.6 MHz central frequency, L22-8v, Verasonics, USA), connected to a multi-channel acquisition system (Vantage 256 High Frequency, Verasonics, USA) is coupled with echographic gel. Thus, the leaf is in the imaging plane of the probe. It is illuminated from the top with 5 ns laser pulses at 10 Hz repetition rate (λ = 532 nm) from a frequency-doubled Nd:YAG laser (Surrelite, Continuum, USA). For each laser shot, PA signals are acquired and RF images are reconstructed using DAS assuming a homogeneous medium and neglecting the presence of the agarose gel. Ground truth images are generated from photographs of the leaf taken with a CMOS camera (X-E2, Fujifilm, Japan). These photographs are converted to gray scale (8 bits) and pixels below a threshold are set to 0 to suppress background noise.

2.2. Creation of the experimental dataset

A model of leaf skeleton was chosen as imaging sample (see Fig. 1.a). To obtain a strong photoacoustic signal, the veins of leafs skeleton were been tainted with black ink and the limbs were dissolved by chemical treatment. The smallest veins of the leaf are finally manually cut to remove unsolvable details. Each pair of the dataset is constituted of a RF image (input of the network) and the corresponding photograph (ground truth) of a 5.12×5.12 mm² area of the leaf.

As shown in Fig. 1.a, the leaf is maintained horizontal into an agarose gel which stands in a tank filled with degassed and deionized water. Through a side window composed of a frame tightening a Mylar membrane, an ultrasonic transducer array (15.6 MHz central frequency, L22-8v, Verasonics, USA), connected to a multi-channel acquisition system (Vantage 256 High Frequency, Verasonics, USA) is coupled with echographic gel. Thus, the leaf is in the imaging plane of the probe. It is illuminated from the top with 5 ns laser pulses at 10 Hz repetition rate (λ = 532 nm) from a frequency-doubled Nd:YAG laser (Surrelite, Continuum, USA). For each laser shot, PA signals are acquired and RF images are reconstructed using DAS assuming a homogeneous medium and neglecting the presence of the agarose gel. Ground truth images are generated from photographs of the leaf taken with a CMOS camera (X-E2, Fujifilm, Japan). These photographs are converted to gray scale (8 bits) and pixels below a threshold are set to 0 to suppress background noise.

Registration between the PA image and the corresponding photograph is needed. The magnitude of the transformation to apply (rotation, translation and scaling) are found by minimizing the correlation coefficient between the PA image (the reference) and the transformed photograph.

630 pairs of images from two leafs are obtained, split
into the experimental training set and the experimental validation set with 550 and 80 pairs respectively. The validation set is used during the training to assess that the optimization process is over. An experimental test set of 15 pairs is then constituted from a part of a leaf that is not present in the training set. It will be used to evaluate the performance of our approach.

2.3. Creation of the simulation dataset

The same photographs are used to create the simulation dataset. Data augmentation is applied on those photographs to increase the size of the training set, by applying rotations, mirror transformations, horizontal or vertical shear, and center expansion or compression. Then, 1×1 cm images are extracted to simulate their PA signals. The method used to simulate PA imaging is described in our previous work [35]. In brief, the imaging system response is experimentally measured for a single source at one spatial location and the synthesis of the RF data coming from a whole object is obtained by summing the contributions of each pixel of the object, that is recovered by applying delays on the initial response. The medium is assumed to be homogeneous with a constant speed of 1500 m.s\(^{-1}\). DAS is then applied to reconstruct an RF image of 5.12×5.12 mm\(^2\) area and the photograph is cropped to match the dimensions. Propagating PA waves from a larger area (1×1 cm\(^2\)) than the one viewed by the network (5.12×5.12 mm\(^2\)) enable to take into account the presence of surrounding structures which can affect reconstruction during experiments.

A series of 1400 pairs of images are obtained. Around five days are needed to compute the dataset on Matlab with a single computer.

2.4. Deep learning framework

UNET [36] is a widely used CNN in the medical field. A slightly modified architecture, presented in supplementary materials is implemented with the open source library Keras. Dropout [37] and batch normalization [38] layers are added to limit the over-fitting of the model. The last layer contains only one filter instead of two in the original version, as the expected output is a single image. The last activation function is also suppressed as the prediction is no longer a binary image. It is worth mentioning that several modifications supposed to improve the result including skip connections between input and output [17], residual blocks [39] and fully densely connected blocks [40] have been tested without significant improvement of the prediction. The cost function is the classical mean squared error, and an Adam optimizer is used with a learning rate of 5.10\(^{-4}\) and momentum of 0.9 [40] with batch size of 8 images. An early stopping approach based on the validation loss was chosen to limit under- and over-fitting [41].

The prediction phase must be random to model uncertainty. In the MC dropout approach, noise is injected in the model by activation of the dropout layers (dropout rate of 50%) both during training and prediction. In this study, 20 inferences are generated from forward passes through the model with a different dropout mask. The different resulting predictions allow to further estimate the distribution mean and its standard deviation which gives a map of uncertainty (see Fig. 1.b).

The training and the evaluation of the network, composed by around 3,000,000 neurons, are performed on a NVIDIA Quadro P2000. Around 30 min are needed for the training on the simulation set and 20 min on the experimental set.

2.5. Quantitative assessment of the network performance

As mentioned previously, the same leaves are used to create the simulation and the experimental dataset. It means that from the same object (ie an area of the leaf), we will be able to obtain either the experimental RF data or the simulated one. For comparison purpose, the reconstructed objects shown in the figures are the same for both the simulation and the experimental case. A third example is used for the MC dropout results, the estimated uncertainties of the two previous examples being described in supplementary materials.

To evaluate the accuracy of a trained neural network, the normalized 2D cross-correlation [12] (NCC) and the scaled and shifted structured similarity index (sSSIM) are computed between each output and ground truth. The first one uses local sum to normalize the cross correlation for feature matching. SSIM [43] is a well-used metric to evaluate the perceived quality of an image. It is computed over several small windows of the image, quantifying the structure, contrast and luminescence similarities. The sSSIM [44] is used for obtaining a scaled and unbiased score which was not disadvantaging for the other reconstruction methods. For an overall performance estimation of the network, the mean and standard deviation values among the test set are presented.

On each result, we evaluate an uncertainty map with the MC dropout method. The standard deviation over time is additionally computed throughout 50 frames to compare a regular way to assess the local variability in the reconstructions with the MC dropout uncertainty estimation. In this case random noise is induced by the acquisition process, resulting in small changes in the predicted images. We also computed the absolute truth error of the reconstruction.

3. Results

3.1. Simulation results

Fig. 2. shows a comparison between the reconstructed image from simulated data provided by the DLA (d,h),
the envelope image (DAS, b,f), a L2-regularized deconvolution (c,g) and the photograph of the object (ground truth, a,e). L2 minimization and DAS clearly do not provide vertical structures, ie the structures elongated in the axis of the probe. Veins having inclinations of more than 45 degrees are missing due to the limited view problem. The inside of the thicker vessel is also missing and the thickness of the thinner vessel is underestimated for DAS reconstruction and overestimated for the L2 minimization (arrows). The deep learning reconstruction yields in contrast an almost artefact-free reconstruction with errors located only on the smallest appendages, resulting from the manual cutting, and on few vertical structures which are not completely recovered (stars).

The performances of the different reconstruction methods and their standard deviation, evaluated over the 15 pairs of the simulation test set with the metric described in 2.5 are shown in Tab. 1 These numbers clearly confirm these qualitative assessments: when the DLA is used, the NCC and sSSIM are around three times higher compared to the simple DAS. Scores for the deconvolution method, not shown here, are on the same order than DAS’s one.

Quantitative performances are shown in Tab. 1 where we can again see a huge improvement for the deep learning approach comparing to the DAS, although lower than simulation results. Both the sSSIM (0.81) and the correlation coefficient (0.77) are significantly enhanced. It should also be noted that the DAS performs better on experimental data than on simulation data.

### 3.2. Experimental results

Fig. 3 shows a comparison between the reconstructed image from the experimental data provided by the DLA (d,h), the conventional DAS reconstruction (b,f), a L2-regularized deconvolution (c,g) and the photograph of the object (ground truth, a,e). Similarly, the DAS approach and the L2-minimization both fail to recover vertical structures as well as to provide a good rendering of the vein thicknesses by filling the inside of the thicker ones. In contrast, the DLA trained on the experimental data yields a reconstruction with most of the vertical structures recovered and a correct thickness of the veins (arrows). A few structures are again not recovered, and some mistakes occurred especially for the reconstruction of vertical veins (stars). The orientation is not always perfectly respected, and in some places, some veins appear when there are none in reality.

#### 3.3. Uncertainty estimation

Results of the MC dropout procedure are presented in Fig. 4. A low standard deviation indicates a good robustness of the technique: the prediction remains stable over several realizations. The estimated uncertainty (Fig. 4.e) is well correlated to the absolute error (Fig. 4.d) and to the standard deviation over time (Fig. 4.f). Most of the errors in the prediction are captured like the small vessel at the bottom, that was not fully recovered, or the central one which was reconstructed but in a curved shape instead of a straight one (arrows).

---

### Table 1: Quantitative measurement of reconstruction quality with the normalized 2D cross-correlation and the scaled and shifted structured similarity index

|                | Simulation          | Experiment         |
|----------------|---------------------|--------------------|
|                | DAS                 | DLA                | DAS  | DLA  |
| NCC            | 0.31 ± 0.02         | 0.89 ± 0.01        | 0.44 ± 0.06 | 0.81 ± 0.03 |
| sSSIM           | 0.29 ± 0.02         | 0.87 ± 0.01        | 0.38 ± 0.05 | 0.77 ± 0.03 |

---

![Figure 2: Results on the simulated test set built from simulated RF data, two examples. a, e, Ground truth: photograph of the object. b, f, Envelope image, delay and sum. c, g, L2-regularized deconvolution. d, h, Prediction of the deep learning algorithm.](image1)

![Figure 3: Results on the experimental test set built from experimentally acquired RF data, two examples. a, e, Ground truth: photograph of the object. b, f, Envelope image, delay and sum. c, g, L2-regularized deconvolution. d, h, Prediction of the deep learning algorithm.](image2)
3.4. Impact of the pretraining and the size of the training set on the performance

In this part, the efficacy of a pretraining session with a simulation dataset is investigated for improving the general performance and for reducing the size of the experimental training set. The uncertainty prediction was not studied in this configuration.

The DLA was trained with experimental datasets of different sizes, from 10 to 550 pairs (the entire dataset). For each size, the training is repeated 30 times with, for each of them, a training and validation set composed of different pairs randomly chosen. This is needed to limit the influence of individual pairs on the training set size (for example, it is likely that 10 examples very close to the test set will provide a better prediction than 20 very distant ones), especially for small set sizes. The displayed sSSIM values are therefore an average over all the test sets from the 30 different weights. To evaluate pretraining, we repeated this procedure with weights initialized by those obtained at the end of a training session on a simulation dataset composed of 1400 pairs.

The results are shown in Fig. 5. As expected, the performance increases with the experimental dataset size. Below 200 pairs, errors remain present and the veins thickness is not always faithfully represented. From 200 pairs, the image quality seems visually stable, although the sSSIM value still slowly increases.

With pretraining (blue curve) convergence is faster. When the full dataset is used, pretraining only slightly increases the performance of the network (sSSIM improved from 0.76 to 0.78). For a smaller size such as 50 pairs, the score improvement is better (from 0.63 to 0.72). A reconstructed image comparable to the one obtained with the total experimental training set is almost reached from this experimental dataset constituted of 50 pairs. In this situation, a pretraining session therefore enables to decrease the size of the experimental training set by a factor 4.

4. Discussion

The algorithm trained with simulated data is able to produce, from simulated data, images that are free from visibility artefacts. When trained with simulated data, the algorithm applied to experimental data fails in achieving a good image quality. When both training and prediction are made with experimental data, while a few errors may remain, most vessels are well recovered. A fundamental difference between simulation and experiment is the nature of the ground truth. In simulation, the ground truth and the data are proportional to the same physical quantity, as the ground truth is the optical absorption of the entire two-dimensional object and the PA data are derived from it. In the experiment, the ground truth of a three-dimensional object is captured from the top with a camera with bright field illumination (in transmission). It then represents the integration of the optical absorption which is rather opaque. Consequently, this picture is not quantitatively representing the sample absorption while the PA data is sensitive to the thickness of each branches and then remains proportional to the optical absorption of the object. Therefore, our method is not supposed to provide a quantitative reconstruction of the PA image as the network is forced to learn another representation of the object.

The use of RF image as input of the network improves the performance both quantitatively and qualitatively compared to envelope image input (see supplementary materials). In fact, the RF image, despite being more different from the true physical structure of the object than the envelope image, and thus from the ground truth, carries more information to be captured by the network. Our results show that errors are often located at the edges of the reconstructed image. Indeed, in these area, less information about the surrounding structures is available. One way to limit these artefacts could be to reconstruct on a larger area and crop the edges. Most of the errors remaining on experimental images are located where the manual cutting was performed resulting in small appendices. These structures, which do not belong to the initial object, turned out to mislead the network which seeks to elongate them to join all the veins together. It is reasonable to think that the number of errors would have been lower on a more regular object. More broadly speaking, these results can be enhanced by improving the quality of the training set.
Nevertheless, the errors in the reconstructed image, especially the invented structures are problematic for end users (clinicians, biologists...). The MC dropout approach proposed in this article helps locating most of them. Importantly, estimated uncertainty remains low where reconstruction is correct, leading to a clear distinction of the suspicious areas: false alarm, which could mislead the clinician, are rare. It is worth mentioning that if the method helps locating the invented structures or incorrect reconstruction, (false positive), it is less efficient when capturing missed structure (false negative), illustration can be found in supplementary materials. This is understandable, as these errors are mostly related to a lack of information in the data. The standard deviation over time was computed to compare our result with an uncertainty estimation. The two maps are well correlated, showing that MC dropout can provide a correct estimation based on an image reconstructed from a single acquisition. This feature is promising in the context of moving tissue imaging and real time navigation.

We explored other methods such as Deep ensemble [45] and Dropout ensemble [33]. In our study, the results were better for MC dropout, in particular for the mean estimation. This difference could be explained by the required modification of the loss function for the other methods, involving a decrease of the overall performance. To obtain these results, our model was trained on an experimental dataset. However in clinical context, large experimental datasets are complex to build. Using only simulated data to train a model and reconstruct experimental images from this model could be considered. In our study, this approach produces unsatisfactory results, as we illustrate in supplementary material: predictions on experimental data provided by a DLA trained either on simulated data or experimental one are shown. Indeed, although the conditions of simulation and experiments were close, the similarity between those conditions was not sufficient, starting with the difference in the nature of the ground truth. Consequently, many artefacts remain on the experimental reconstructed image. Theses results are in agreement with observations made by Davoudi et al. [23].

However, as shown in the previous section, the incorporation of simulated data through a transfer learning approach allows reducing drastically the size of the experimental dataset. The algorithm only needs to update its parameters with the difference between the simulation and the experiment, which is easier than learning the overall procedure. In the medical field, such a pretraining session could be useful for reducing the number of patients necessary to create a training set.

Figure 5: Performance of a deep learning algorithm trained on an experimental dataset only (red) versus a deep learning algorithm pretrained on 1400 simulated data then trained on the same experimental dataset (blue), for different number of experimental data.
Several challenges must be taken into account in order to apply our method in a more general context. As obtaining an optical image of an object is not feasible in depth in an opaque medium such as a biological environment, alternatives methods must be developed to construct the training set in order to predict PA images in vivo with a neural network.

The influence of noise on RF data should also be studied to assess the validity of this approach in a noisier environment. In this work, the signal to noise ratio (SNR) on RF data is about 60. Note that this value represents the SNR of signals produced by horizontal structures, while this work mainly focuses on the reconstruction of vessels affected by the visibility problem, for which the signal is almost nonexistent. Besides, the background of our DAS images is mostly polluted by clutter, an artefact located around the object originating from the lack of information for reconstruction. The amplitude is often higher than those of the signal generated by vertical structures. Finally, the quality of the prediction is strongly influenced by the class of the object to reconstruct. The relative homogeneity of the studied dataset is one of the reason the DLA performs well. In a more general context, the capacity to generalize properly would be crucial.

Aside from increasing the quality of the reconstructed image, DLA offers other interesting features. Only 10 ms is needed to make a prediction using a regular graphic card which is much lower than the reconstruction time for the deconvolution method. Real time reconstruction during navigation could be achieved. Besides, once trained, the network does not need any parameters to be set by the user unlike in the deconvolution approach where the regularization parameter has to be chosen carefully and in a rather subjective way.

5. Conclusion

The possibility of removing visibility artefacts with a neural network has been demonstrated both in simulation and experiments on a class of complex objects. Vertical parts of objects and the inside of large structures, missing on conventional reconstruction approaches, are recovered. These qualitative assessments are confirmed by quantitative metrics, which are far better for the DLA approach compared to than for the conventional reconstruction methods. However, some errors may still present in the reconstructed image, although their number might be reducible by improving the experimental protocol. To locate these errors, a MC dropout approach was proposed and successfully applied: the generated uncertainty map is well correlated with the true error. Besides, it has been shown that pretraining the network with simulated data enables to reduce the size of the experimental training set by a factor of 4 while maintaining a similar quality of reconstruction.

6. Funding Information

This project has received funding from the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation program (grant agreement No 681514-COHERENCE).

References

[1] P. Beard, Biomedical photoacoustic imaging, Interface focus 1 (4) (2011) 602–631.
[2] Z. Guo, L. Li, L. V. Wang, On the speckle-free nature of photoacoustic tomography, Medical physics 36 (9Part1) (2009) 4084–4088.
[3] X. L. Deán-Ben, D. Razansky, On the link between the speckle free nature of optoacoustics and visibility of structures in limited-view tomography, Photoacoustics 4 (4) (2016) 133–140.
[4] R. A. Kruger, W. L. Kiser Jr, D. R. Reinecke, G. A. Kruger, Thermoacoustic computed tomography using a conventional linear transducer array, Medical physics 30 (5) (2003) 856–860.
[5] D. Yang, D. Xing, S. Yang, L. Xiang, Fast full-view photoacoustic imaging by combined scanning with a linear transducer array, Optics express 15 (23) (2007) 15566–15575.
[6] J. Xia, M. R. Chatni, K. I. Maslov, Z. Guo, K. Wang, M. A. Anastasio, L. V. Wang, Whole-body ring-shaped confocal photoacoustic computed tomography of small animals in vivo, Journal of biomedical optics 17 (5) (2012) 050506.
[7] X. L. Dean-Ben, D. Razansky, Localization optoacoustic tomography, Light: Science & Applications 7 (4) (2018) 18004–18004.
[8] L. Wang, G. Li, J. Xia, L. V. Wang, Ultrasonic-heating-encoded photoacoustic tomography with virtually augmented detection view, Optica 2 (4) (2015) 307–312.
[9] T. Chaigne, B. Arnal, S. Vilov, E. Bossy, O. Katz, Super-resolution photoacoustic imaging via flow-induced absorption fluctuations, Optica 4 (11) (2017) 1397–1404.
[10] S. Vilov, G. Godefroy, B. Arnal, E. Bossy, A unified framework for photoacoustic fluctuation imaging. application to visibility enhancement with fluctuations induced by blood flow, arXiv preprint arXiv:2006.09341 (2020).
[11] Y. Bengio, I. Goodfellow, A. Courville, Deep learning, Vol. 1, Citeseer, 2017.
[12] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, et al., Gradient-based learning applied to document recognition, Proceedings of the IEEE 86 (11) (1998) 2278–2324.
[13] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431–3440.
[14] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al., Imagenet large scale visual recognition challenge, International journal of computer vision 115 (3) (2015) 211–252.
[15] C. Dong, Y. Deng, C. Change Loy, X. Tang, Compression artifacts reduction by a deep convolutional network, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 576–584.
[16] J. Xie, L. Xu, E. Chen, Image denoising and inpainting with deep neural networks, in: Advances in neural information processing systems, 2012, pp. 3431–3440.
[17] Z. Akkus, A. Galimzianova, A. Hoogi, D. L. Rubin, B. J. Erickson, Deep learning for brain mri segmentation: state of the art and future directions, Journal of digital imaging 30 (4) (2017) 449–459.
[18] K. H. Jin, M. T. McCann, E. Froustey, M. Unser, Deep convolutional neural network for inverse problems in imaging, IEEE Transactions on Image Processing 26 (9) (2017) 4509–4522.
[19] T. Liu, Y. Wang, X. Yang, B. Lei, L. Liu, S. X. Li, D. Ni, T. Wang, Deep learning in medical ultrasound analysis: a review, Engineering (2019).
[20] D. Waibel, J. Gröhl, F. Isensee, T. Kirchner, K. Maier-Hein, L. Maier-Hein, Reconstruction of initial pressure from limited view photoacoustic images using deep learning, in: Photons Plus Ultrasound: Imaging and Sensing 2018, Vol. 10494, International Society for Optics and Photonics, 2018, p. 104942S.

[21] A. Hauptmann, F. Lucka, M. Betcke, N. Huynh, J. Adler, B. Cox, P. Beard, S. Ourselin, S. Arridge, Model-based learning for accelerated, limited-view 3-d photoacoustic tomography, IEEE transactions on medical imaging 37 (6) (2018) 1382–1393.

[22] S. Antholzer, M. Haltmeier, J. Schwab, Deep learning for photoacoustic tomography from sparse data, Inverse problems in science and engineering 27 (7) (2019) 987–1005.

[23] N. Davoudi, X. L. Deán-Ben, D. Razansky, Deep learning optoacoustic tomography with sparse data, Nature Machine Intelligence 1 (10) (2019) 453–460.

[24] S. Govinahallisathyanarayana, B. Ning, R. Cao, S. Hu, J. A. Hossack, Dictionary learning-based reverberation removal enables depth-resolved photoacoustic microscopy of cortical microvasculature in the mouse brain, Scientific reports 8 (1) (2018) 1–12.

[25] A. Reiter, M. A. L. Bell, A machine learning approach to identifying point source locations in photoacoustic data, in: Photons Plus Ultrasound: Imaging and Sensing 2017, Vol. 10064, International Society for Optics and Photonics, 2017, p. 100643I.

[26] D. Allman, A. Reiter, M. A. L. Bell, Photoacoustic source detection and reflection artifact removal enabled by deep learning, IEEE transactions on medical imaging 37 (6) (2018) 1464–1477.

[27] C. Cai, K. Deng, C. Ma, J. Luo, End-to-end deep neural network for optical inversion in quantitative photoacoustic imaging, Optics letters 43 (12) (2018) 2752–2755.

[28] T. Kirchner, J. Gröhl, L. Maier-Hein, Context encoding enables machine learning-based quantitative photoacoustics, Journal of biomedical optics 23 (5) (2018) 056008.

[29] S. Gu, V. S. Kadimesetty, S. K. Kalva, M. Pramanik, S. Ganapathy, P. K. Yalavarthy, Deep neural network-based bandwidth enhancement of photoacoustic data, Journal of biomedical optics 22 (11) (2017) 116001.

[30] S. Guan, A. Khan, S. Sikdar, P. Chitnis, Fully dense unet for 2d sparse photoacoustic tomography artifact removal, IEEE journal of biomedical and health informatics (2019).

[31] M. W. Kim, G.-S. Jeng, I. Pelivanov, M. O'Donnell, Deep-learning image reconstruction for real-time photoacoustic system, IEEE Transactions on Medical Imaging (2020).

[32] Y. Gal, Z. Ghahramani, Dropout as a bayesian approximation: Representing model uncertainty in deep learning, in: international conference on machine learning, 2016, pp. 1050–1059.

[33] Y. Xue, S. Cheng, Y. Li, L. Tian, Reliable deep-learning-based phase imaging with uncertainty quantification, Optica 6 (5) (2019) 618–629.

[34] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, SIAM journal on imaging sciences 2 (1) (2009) 183–202.

[35] S. Vilov, B. Arnal, E. Hojman, Y. C. Eldar, O. Katz, E. Bossy, Super-resolution photoacoustic and ultrasound imaging with sparse arrays, Scientific reports 10 (1) (2020) 1–8.

[36] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, Springer, 2015, pp. 234–241.

[37] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: A simple way to prevent neural networks from overfitting, The journal of machine learning research 15 (1) (2014) 1929–1958.

[38] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, arXiv preprint arXiv:1502.03167 (2015).

[39] F. I. Diakogiannis, F. Waldner, P. Caccetta, C. Wu, Resunet-a: a deep learning framework for semantic segmentation of remotely sensed data, ISPRS Journal of Photogrammetry and Remote Sensing 162 (2020) 94–114.

[40] D. P. Kingma, J. Ba, Adam: A method for stochastic optimiza-

tion, arXiv preprint arXiv:1412.6980 (2014).

[41] L. Prechelt, Early stopping-but when?, in: Neural Networks: Tricks of the trade, Springer, 1998, pp. 55–69.

[42] J.-C. Yoo, T. H. Han, Fast normalized cross-correlation, Circuits, systems and signal processing 28 (6) (2009) 819.

[43] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE transactions on image processing 13 (4) (2004) 600–612.

[44] J. Schwab, S. Antholzer, R. Nuster, M. Haltmeier, Real-time photoacoustic projection imaging using deep learning, arXiv preprint arXiv:1801.06693 (2018).

[45] B. Lakshminarayanan, A. Pritzel, C. Blundell, Simple and scalable predictive uncertainty estimation using deep ensembles, in: Advances in neural information processing systems, 2017, pp. 6402–6413.
Solving the visibility problem in photoacoustic imaging with a deep learning approach providing prediction uncertainties: supplementary materials

Guillaume Godefroy, Bastien Arnal, Emmanuel Bossy
Univ. Grenoble Alpes, CNRS, LIPhy, 140 rue de la Physique, CS 40700, 38058 Grenoble CEDEX 9, FRANCE

This document provides supplementary information to "Solving the visibility problem in photoacoustic imaging with a deep learning approach providing prediction uncertainties". Included are a schematic representation of the deep learning algorithm (DLA), a comparison of the prediction of the DLA based on either modulus image or radiofrequency (RF) image as input, the prediction from experimental data for a DLA trained with simulation data and the uncertainty estimation of the instances presented in the main text.

1. CNN architecture

Unet is a well known network architecture first developed for segmentation task. Our implementation is shown in Fig. 1. It is a convolutional neural network composed of two paths: the contracting and expanding path. The first one, called the encoder, is a traditional stack of convolutional and pooling layers where the network extracts more and more complex features. The second one, called the decoder, is the symmetric expanding path where pooling operation are replaced by upsampling operator to recover at the output the resolution of the input. Context information is propagated from the encoder to the decoder through skip connections to provide local information to the global information while upsampling (black arrows).

The weights are initialised with samples from a truncated normal distribution centered on 0 with standard deviation depending of the number of units in the weight tensor. Dropout layers are added to this architecture. Dropout is a popular regularization technique to limit overfitting. A certain set of neurons, chosen randomly, are disabled at each training step. This prevents units from co-adapting too much and forces the network to learn more robust features. Batch normalization normalizes the output of the previous activation layer by subtracting the batch mean and dividing by the batch standard deviation. It helps to speed up the learning and also reduces overfitting by adding some noise, similarly as dropout.

![CNN architecture](image)

**Figure 1**: CNN architecture.

2. Envelope or real valued image as input of the network

The input of the network is obtained from the delays and sum algorithm (DAS) applied to time signals. When applied to real time signals, DAS provides a real valued RF image. When applied to complex signals obtained with a Hilbert
transform, DAS provides a complex image whose modulus is the envelope image. The RF image and the envelope image are the two types of input that we consider here. One DLA was trained for each type of input, the corresponding predictions are displayed in Fig 2. The algorithm performs better on the RF image, leading to a scaled and shifted structured similarity index (sSSIM) of 0.77 instead of 0.72. The prediction from the envelope image suffers from more artefacts (arrows) and the DLA fails to recover the true vessels thicknesses, which are over estimated. The RF image, despite being more different from the true physical structure of the object than the envelope image, and thus from the ground truth, carries more information to be captured by the network.

![Figure 2](image)

Figure 2: Deep learning algorithm (DLA) prediction on experimental data for envelope PA image or RF PA image as input, two examples. a, f, Envelope image. b, g, RF image. c, h, Prediction with DLA trained on envelope data. d, i, Prediction on experimental data with DLA trained on RF data. e, j, Ground truth.

3. Reconstruction of experimental data with DLA trained on simulation data

Predictions from neural networks trained with simulation datasets or experimental datasets are presented in Fig 3. Although the DLA trained on simulation data still manages to find several vertical structures not visible on the DAS image, the predicted image is polluted by a lot of artefacts. Clearly, experimental data are necessary to train efficiently the model. However, as shown in the main text, pretraining the network on a simulation dataset allows reducing the size of the experimental training set.

![Figure 3](image)

Figure 3: Deep learning algorithm (DLA) prediction on experimental data for DLA trained on simulated data and DLA trained on experimental data, two examples. a, d, Ground truth. b, e, Prediction with DLA trained on simulated data. c, f, Prediction on experimental data with model trained on experimental data.
4. Uncertainty estimation

Uncertainty estimation of the two previous examples are presented in Fig 4. Similarly to the example in the main text, the standard deviation (std) map helps to locate errors in the reconstruction such as invented structures and incorrect orientation or position (arrows). One can notice that although the value of the std is higher at location of some missing veins (star), some of them are not even displayed on the std (circle). The lack of information in the data may be more important for these structures, misleading the DLA.

Figure 4: Uncertainty estimation on experimental RF data.

a, g, Ground truth: photograph of the object. b, h, Mean image of the object computed over 20 inferences generated from a unique acquisition with the deep learning algorithm, dropout activated.

c, i, Corresponding STD. d, j, Absolute error between the ground truth (a,g) and the mean (b,h).

e, k, Corresponding STD. f, l, Corresponding STD