Journal of *Imaging*

MDPI

Article

Confidence Estimation for Machine Learning-Based Quantitative Photoacoustics

Janek Gröhl [1,2,*], Thomas Kirchner [1,3], Tim Adler [1] and Lena Maier-Hein [1,2,*]

1 Division of Computer Assisted Medical Interventions (CAMI), German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany; t.kirchner@dkfz-heidelberg.de (T.K.); t.adler@dkfz-heidelberg.de (T.A.)
2 Medical Faculty, Heidelberg University, 69120 Heidelberg, Germany
3 Faculty of Physics and Astronomy, Heidelberg University, 69120 Heidelberg, Germany
* Correspondence: j.groehl@dkfz-heidelberg.de (J.G.); l.maier-hein@dkfz-heidelberg.de (L.M.-H.)

Received: 30 October 2018; Accepted: 6 December 2018; Published: 10 December 2018

Abstract: In medical applications, the accuracy and robustness of imaging methods are of crucial importance to ensure optimal patient care. While photoacoustic imaging (PAI) is an emerging modality with promising clinical applicability, state-of-the-art approaches to quantitative photoacoustic imaging (qPAI), which aim to solve the ill-posed inverse problem of recovering optical absorption from the measurements obtained, currently cannot comply with these high standards. This can be attributed to the fact that existing methods often rely on several simplifying a priori assumptions of the underlying physical tissue properties or cannot deal with realistic noise levels. In this manuscript, we address this issue with a new method for estimating an indicator of the uncertainty of an estimated optical property. Specifically, our method uses a deep learning model to compute error estimates for optical parameter estimations of a qPAI algorithm. Functional tissue parameters, such as blood oxygen saturation, are usually derived by averaging over entire signal intensity-based regions of interest (ROIs). Therefore, we propose to reduce the systematic error of the ROI samples by additionally discarding those pixels for which our method estimates a high error and thus a low confidence. In silico experiments show an improvement in the accuracy of optical absorption quantification when applying our method to refine the ROI, and it might thus become a valuable tool for increasing the robustness of qPAI methods.

Keywords: confidence learning; uncertainty estimation; quantitative photoacoustic imaging; error analysis; deep learning

1. Introduction

Photoacoustic imaging (PAI) has been shown to have various medical applications and to potentially benefit patient care [1–3]. It is a non-invasive modality that offers the ability to measure optical tissue properties, especially optical absorption μ_a, both locally resolved and centimeters deep in tissue. Knowledge of these properties allows for deriving functional tissue parameters, such as blood oxygenation SO_2, which is a biomarker for tumors and other diseases [4]. The photoacoustic (PA) signal is a measure of the pressure waves arising from the initial pressure distribution p_0, which depends mainly on μ_a, the Grüneisen coefficient Γ, and the light fluence ϕ, which is shaped by the optical properties of the imaged tissue [5]. Because of this dependence, the measured p_0 is only a qualitative indicator of the underlying μ_a, because even if the initial pressure distribution could be recovered perfectly, estimation of the light fluence is an ill-posed inverse problem that has not conclusively been solved [6].

In order to derive quantitative information from initial pressure p_0 reconstructions of photoacoustic images, one has to account for the light fluence and solve the optical inverse problem. Most methods model the distribution of optical absorption coefficients by iteratively updating the

distribution after computing the solution of a forward model (cf., e.g., [7–14]) with inclusion of the acoustic inverse problem [15,16]. Alternatively, in multispectral photoacoustic imaging applications, the functional parameters are approximated directly by using a variety of spectral unmixing techniques (cf., e.g., [17–19]). Recently, machine learning-based methods for quantitative PAI (qPAI) have been proposed. These encompass end-to-end deep learning on 2D images [20] or the estimation of voxel point estimates with Context Encoding qPAI (CE-qPAI) [21], which incorporates the 3D p_0 context around each voxel into a single feature vector that is used to learn the fluence at that particular voxel. Some of the listed approaches to qPAI have been shown to work in ideal in silico conditions or on specific datasets. At the same time, they have proven difficult to use in clinical applications, which can be attributed to a lack of robustness caused by a priori assumptions that are made regarding, e.g., illumination, probe design, calibration factors, or scattering properties [22]. Developing tools to estimate systematic errors and gain information on the quantification of uncertainties in PAI could thus be of great benefit and could be utilized to improve quantification accuracy.

Uncertainty quantification and compensation is an essential research objective in computer sciences and has been studied extensively in various fields, including image-guided navigation (cf., e.g., [23,24]), multi-modal image registration (cf., e.g., [25,26]), and lesion detection [27]. Current approaches to obtaining confidence intervals for neural network estimates include, e.g., dropout sampling (cf., e.g., [28–31]), probabilistic inference (cf., e.g., [32–34]), sampling from latent variables (cf., e.g., [35–37]), or using ensembles of estimators (cf., e.g., [38,39]). The exploration of such uncertainty quantification methods in the field of PAI, however, has only just started (cf., e.g., [40–43]).

In a recent publication [44], we showed a method for uncertainty quantification for the CE-qPAI method. A key result was that the practice of evaluating PA images over a purely input noise-based (aleatoric) region of interest (ROI) can be improved when also taking into account model-based (epistemic) uncertainty. To achieve this, we combined both sources of uncertainty into a joint uncertainty metric and used this to create an ROI mask on which to compute the statistics. A limitation to that approach could be seen in the fact that we used an uncertainty model specially tailored toward the CE-qPAI method. To overcome this bottleneck, we expand on our prior work in this contribution and present a method that yields confidence estimates by observing the performance of an arbitrary qPAI algorithm and uses the estimates to refine an ROI that was defined based on aleatoric uncertainty.

For validation in the context of qPAI, we applied this methodology to different PA signal quantification algorithms to investigate whether the approach is applicable in a general manner. We hypothesize that an estimated error metric is indicative of the actual absorption quantification error and that we can consequently improve on μ_a estimations by evaluating on an ROI that is further narrowed down with a confidence threshold (CT).

2. Materials and Methods

This section gives an overview of the confidence estimation approach, the experiments, and the used dataset and briefly introduces the different qPAI methods that are being used.

Method for Confidence Estimation. Our approach to confidence estimation can be applied to any qPAI method designed to convert an input image I (p_0 or raw time-series data) into an image of optical absorption I_{μ_a}. In order to not restrict the qPAI method to a certain class (e.g., a deep learning-based method), we made the design decision to base the confidence quantification method on an external observing method. For this, we use a neural network, which is presented tuples of input image I and absorption quantification error e_{μ_a} in the training phase. When applying the method to a previously unseen image I, the following steps are performed (cf. Figure 1).

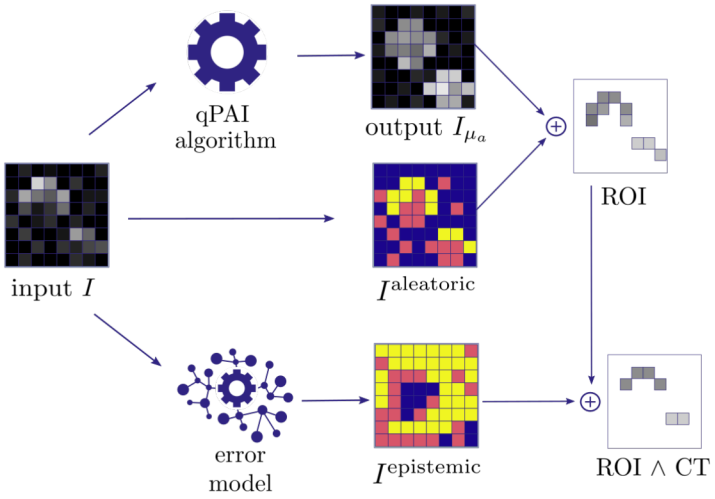

Figure 1. Visualization of the proposed method for confidence estimation using an observing neural network as an error model. The estimator generates an output for a given input and the error model is used to obtain an estimate of the quantification error from the same input data. The region of interest (ROI), which is based on the aleatoric uncertainty $I^{\text{aleatoric}}$ extracted from the input data, can then be refined using the error estimates $I^{\text{epistemic}}$ of the error model as a confidence threshold (CT).

1. Quantification of aleatoric uncertainty: I is converted into an image $I^{\text{aleatoric}}$ reflecting the aleatoric uncertainty. For this purpose, we use the contrast-to-noise-ratio (CNR), as defined by Welvaert and Rosseel [45], as $\text{CNR} = (S - \text{avg}(b))/\text{std}(b)$, with S being the pixel signal intensity, and $\text{avg}(b)$ and $\text{std}(b)$ being the mean and standard deviation of all background pixels in the dataset. Using this metric, we predefine our ROI to comprise all pixels with $\text{CNR} > 5$.
2. Quantification of epistemic confidence: I is converted into an image $I^{\text{epistemic}}$ reflecting the epistemic confidence. For this purpose, we use the external model to estimate the quantification error e_{μ_a} of the qPAI algorithm.
3. Output generation: A threshold over $I^{\text{aleatoric}}$ yields a binary image with an ROI representing confident pixels in I_{μ_a} according to the input signal intensity. We then proceed to narrow down the ROI by applying a confidence threshold (CT) which removes the n% least confident pixels according to $I^{\text{epistemic}}$.

Deep Learning Model. As our external observing network, we used an adapted version of a standard U-Net [46] implemented in PyTorch [47]. The model uses 2×2 max pooling for downscaling, 2×2 transpose convolutions for upscaling, and all convolution layers have a kernel size of 3×3 and a padding of 1×1. We modified the skip connections to incorporate a total of three convolution layers and thus generate an asymmetric U-Net capable of dealing with different input and output sizes (cf. Figure 2). This is necessary to enable the network to directly output reconstructed initial pressure or optical absorption distributions when receiving raw time-series data as input, as the data have different sizes on the y-axis. Specifically, the second of these convolutions was modified to have a kernel size of 3×20, a stride of 1×20, and a padding of 1×9, effectively scaling down the input along the y-axis by a factor of 20. To be more robust to overfitting, we added dropout layers with a dropout rate of 25% to each convolutional layer of the network. Note that a recent study [48] suggests that the U-net architecture is particularly well-suited for medical imaging applications because of its ability to generate data representations on many abstraction levels.

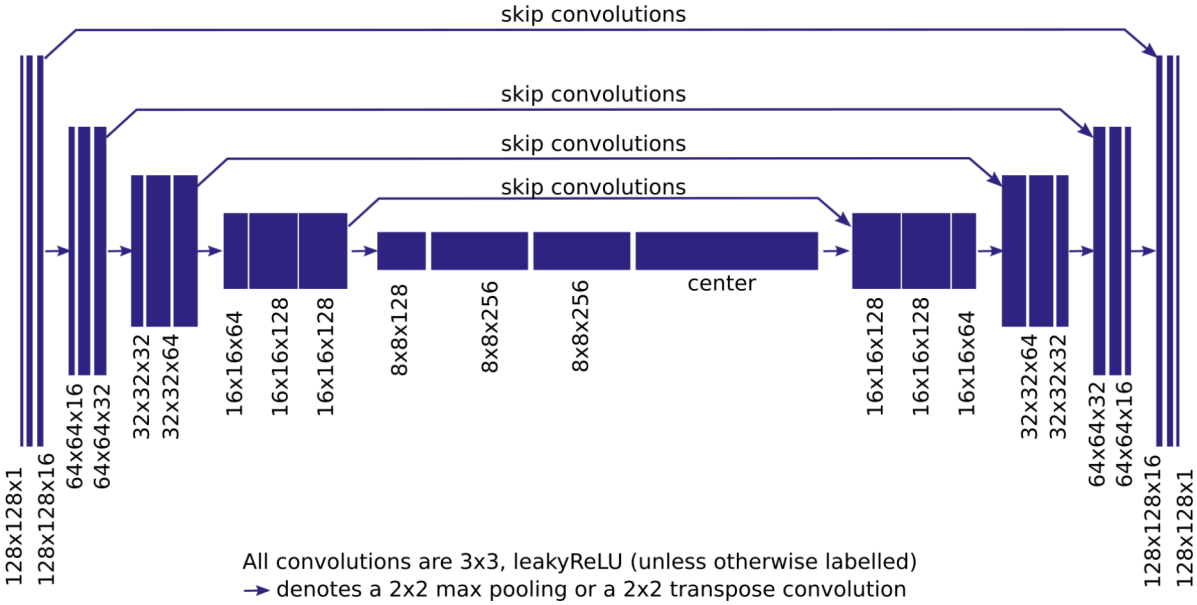

Figure 2. Visualization of the deep learning model used in the experiments: a standard U-Net structure with slight modifications to the skip connections. The (x, y, c) numbers shown represent the x and y dimensions of the layers, as well as the number of channels c. Specifically, in this figure, they show the values for a 128 × 128 input and 128 × 128 output. The center consists of an additional convolution and a skip convolution layer to enable different input and output sizes.

Quantitative PAI Methods. We applied our approach to confidence estimation to three different qPAI methods; a naïve quantification method, as well as two different deep learning-based approaches (cf. Figure 3). The three methods are detailed in the following paragraphs.

Naïve Fluence Correction: As a naïve quantitative PAI (qPAI) reference method that does not use a deep learning model to reconstruct a quantitative absorption estimate $\hat{\mu}_a$, we performed fluence compensation similar to, e.g., [49,50]. To achieve this, we used a simple Monte Carlo fluence simulation ϕ_h based on the same hardware setup as used in the dataset, without any vascular structures inside the volume but instead with a homogeneous absorption coefficient of 0.1 cm^{-1} and a reduced scattering coefficient of 15 cm^{-1}. To quantify optical absorption with this method, we corrected the simulated p_0 images with ϕ_h by calculating $\hat{\mu}_a = p_0/\phi_h$.

Fluence Correction: Fluence correction refers to a two-stage algorithm, where the initial pressure is corrected by an estimation of the underlying light fluence. The quantification model for this method uses two deep neural networks, N_1 and N_2, with adapted U-Net architectures, as described above (cf. Figure 2). N_1 yields an estimation of the underlying fluence $\hat{\phi}$ from the input data S: $N_1(S) = \hat{\phi}$, and N_2 is used to obtain the initial pressure distribution from the input data $\hat{p_0}$: $N_2(S) = \hat{p_0}$ with the aim of also reducing the noise of S. When using PA raw time-series data as the input, it has to be considered that the decoder and encoder sections of N_2 are asymmetric and the described U-Net adaptation has to be used. The optical absorption coefficients are then estimated from the results of N_1 and N_2 by computing $\hat{\mu}_a = N_2(S)/N_1(S) = \hat{p_0}/\hat{\phi}$ (for a visual representation of the method see Figure 3a).

Direct Absorption Estimation: Recently, a deep learning method for end-to-end estimation of absorption and derived functional parameters from p_0 distributions was proposed [20]. In a similar fashion, we also use a deep learning model N with a modified U-Net architecture (cf. Figure 2) to directly estimate μ_a from the input signal S: $N(S) = \hat{\mu}_a$ (for a visual representation of the method see Figure 3b). This time, μ_a estimation is done without the intermediate step of fluence correction, which might be more sensitive to errors due to error propagation or the presence of artifacts and noise.

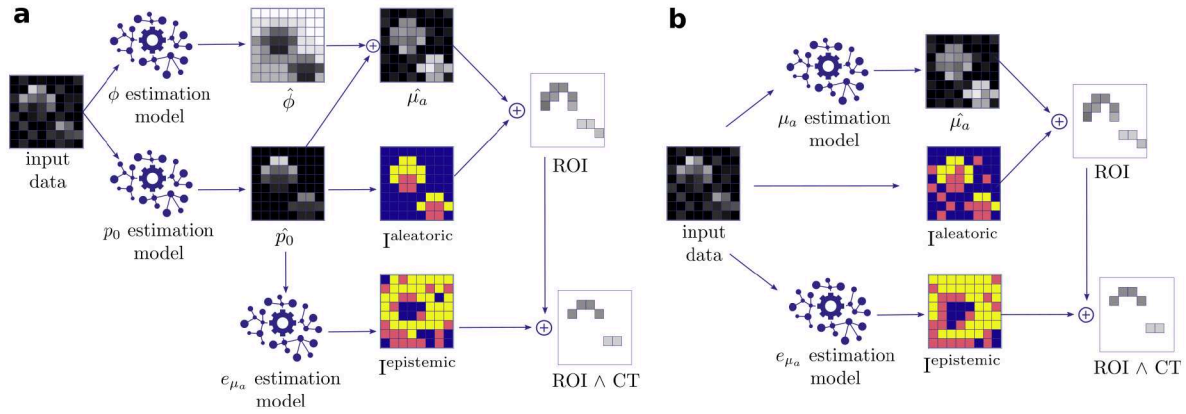

Figure 3. Visualization of the two methods for absorption quantification with subsequent confidence estimation. (**a**) A quantification approach based on fluence estimation. In our implementation, the denoised initial pressure $\hat{p_0}$ and the fluence $\hat{\phi}$ distributions are estimated using deep learning models. These are used to calculate the underlying absorption $\hat{\mu_a}$. An error model then estimates the quantification error and, in combination with an aleatoric uncertainty metric, a region of interest is defined. (**b**) An approach in which one model is used to directly estimate $\hat{\mu_a}$ from the input data. $I^{\text{aleatoric}}$ is calculated on the basis of the input data and $I^{\text{epistemic}}$ is estimated with a second model.

Validation Data. We simulated an in silico dataset containing 3600 training samples, 400 validation samples, as well as 150 calibration and test samples which were used in all experiments. Each data sample consists of the ground truth optical tissue parameters, the corresponding light fluence, and the initial pressure distribution simulated with the mcxyz framework [51], which is a Monte Carlo simulation of photon propagation where we assume a symmetric illumination geometry with two laser outputs. The data sample also comprises raw time-series data simulated using the k-Wave [52] toolkit with the first-order 2D k-space method, assuming a 128-element linear array ultrasound transducer with a central frequency of 7.5 MHz, a bandwidth of 60%, and a sample rate of 1.5×10^{-8} s. The illumination and ultrasound geometry are depicted in Figure 4. Each tissue volume sample comprises 1–10 tubular vessel structures, whose absorption coefficients μ_a range from 2 to 10 cm^{-1} in vessel structures and are assumed constant with 0.1 cm^{-1} in the background. We chose a constant reduced scattering coefficient of 15 cm^{-1} in both background and vessel structures. Additional details on the simulation parameters can also be found in our previous work [53]. The raw time-series data was noised after k-space simulation with an additive Gaussian noise model of recorded noise of our system [54]. For the experiments where we directly use p_0, we noise the initial pressure distribution with a Gaussian additive noise model, as also described in [7]. In our case, the noise model comprised an additive component with $(5 \pm 5)\%$ of the mean signal and a multiplicative component with a standard deviation of 20% to simulate imperfect reconstructions of p_0. The data used in this study is available in a Zenodo repository [55].

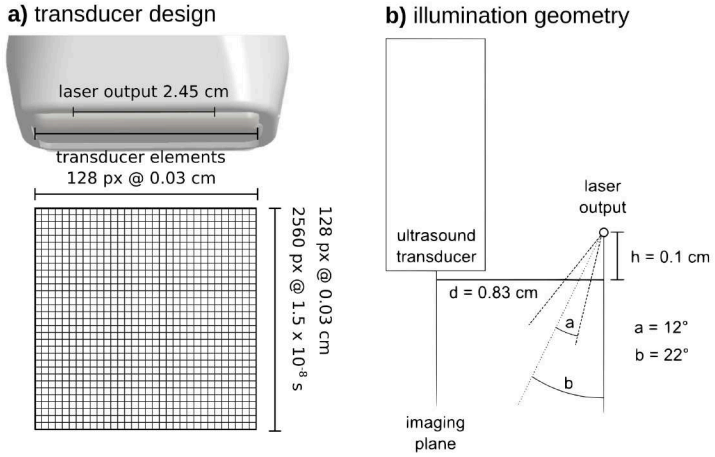

Figure 4. Depiction of the illumination geometry and the transducer design which is based on our Fraunhofer DiPhAS photoacoustic imaging (PAI) system [54]. (**a**) The ultrasound transducer design with the position of the laser output and the transducer elements, as well as the imaging plane; (**b**) one-half of the symmetric transducer design, where the laser outputs are in parallel left and right to the transducer elements over a length of 2.45 cm.

Experimental Design. We performed five in silico experiments to validate our approach to confidence estimation in qPAI—one experiment with naïve fluence correction applied to p_0 data, as well as our different configurations for quantification with deep learning. Both methods shown in Figure 3 are applied to initial pressure p_0, as well as raw time-series data. We used the Trixi [56] framework to perform the experiments. All qPAI models were trained on the training set and the progress was supervised with the validation set. We also used the validation set for hyperparameter optimization of the number of training epochs and batch sizes. We trained for 50 epochs, showing the network 10^4 randomly drawn and augmented samples from the training set; each epoch had a learning rate of 10^{-4}, used an L_1 loss function, and augmented every sample with a white Gaussian multiplicative noise model and horizontal mirroring to prevent the model from overfitting. Afterward, we estimated the optical absorption $\hat{\mu}_a$ of the validation set, calculated the relative errors $e_{\mu_a} = |\hat{\mu}_a - \mu_a|/\mu_a$, and trained the external observing neural networks on the errors of the validation set with the same hyperparameters, supervising the progression on the calibration set. For better convergence, we used the weights of the p_0 estimation model as a starting point for the e_{μ_a} estimation deep learning models. All results presented in this paper were computed on the test set.

3. Results

We report the relative absorption quantification error e_{μ_a} at various different confidence thresholds (CT). To this end, we evaluated only the top n percent most confident estimates by excluding all estimates below a certain CT. We performed this evaluation over all ROI image samples of the respective dataset and examined the relative changes in e_{μ_a} compared with the evaluation over all ROI pixels. This was done in five different in silico experiments, corresponding to the qPAI methods when applied to initial pressure, as well as raw time-series input data. Our findings are summarized in Figure 5.

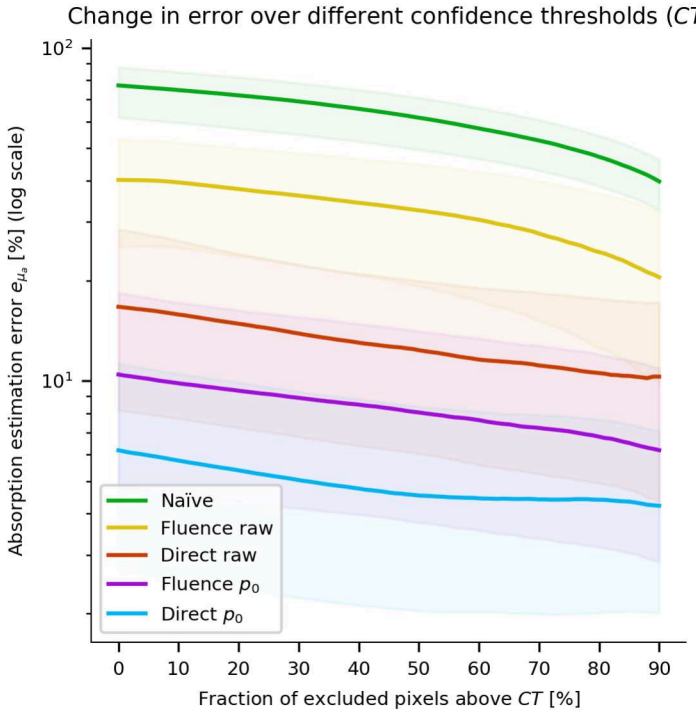

Figure 5. Quantification error as a function of the confidence threshold (CT) for five different quantification methods. The line shows the median relative absorption estimation error when only evaluating with the most confident estimates regarding CT, and the transparent background shows the corresponding interquartile range. Naïve: Fluence correction with a homogeneous fluence estimate; Fluence raw/p_0: Deep learning-based quantification of the fluence applied to p_0 and raw time-series input data and subsequent estimation of μ_a; Direct raw/p_0: End-to-end deep learning-based quantification of μ_a applied to p_0 and raw time-series input data.

Figure 5 shows the absorption quantification error e_{μ_a} for all five experiments for confidence thresholds CT ranging from 10 to 100%. In all cases, the error decreases when excluding more pixels with a higher estimated error. When only considering the top 50% most confident estimates, our method yields a decrease in the error e_{μ_a} of up to approx. 30% (increasing to up to about a 50% improvement when evaluating only the top 10% of the most confident estimates). Figure 6 shows violin plots visualizing the changes in the distribution of e_{μ_a} when applying different confidence thresholds CT. The results reveal a meandering of the distribution toward lower error values. Note that the results are only given for the ROI. When computing the statistics over the entire image (thus decreasing the difficulty of the problem due to the homogeneous background), the median quantification error drops to about 0.1% (direct estimation), 5–10% (fluence correction), and 20% (naïve approach).

Figure 7 shows representative images of the experiment corresponding to the end-to-end direct μ_a quantification method applied to p_0 data. It shows samples of the test set comprising the best, median, and worst result when applying a 50% CT to narrow down the ROI. The results show a maximum increase of nearly 80% in accuracy in the best case while achieving a median improvement of about 29% and worsening the quantification results by 28% in the worst case from the test set. Analogous illustrations for the other experiments can be found in Appendices A–D.

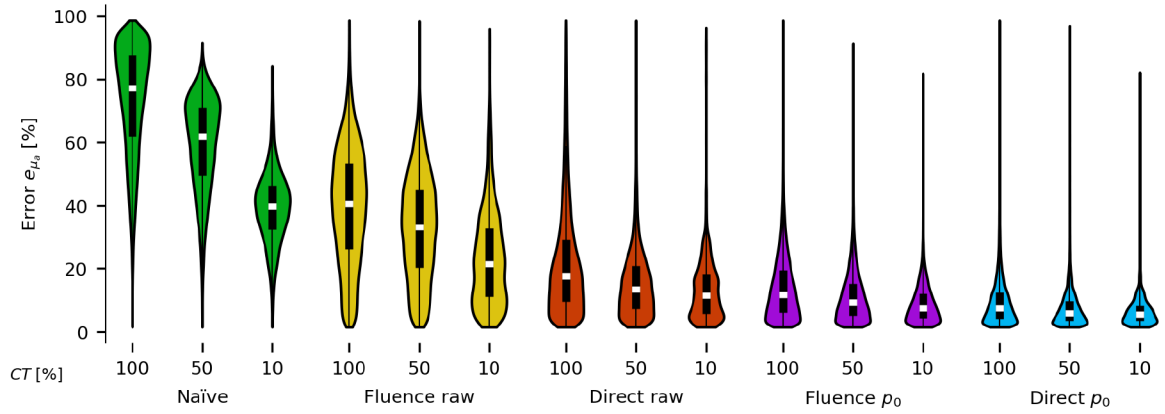

Figure 6. Visualization of the changes in the distribution of the absorption quantification error e_{μ_a} when applying different confidence thresholds $CT = \{100\%, 50\%, 10\%\}$. The plot shows results for all five conducted experiments. The white line denotes the median, and the black box corresponds to the interquartile range. Outliers of e_{μ_a} with a value greater than 100% have been omitted from this plot.

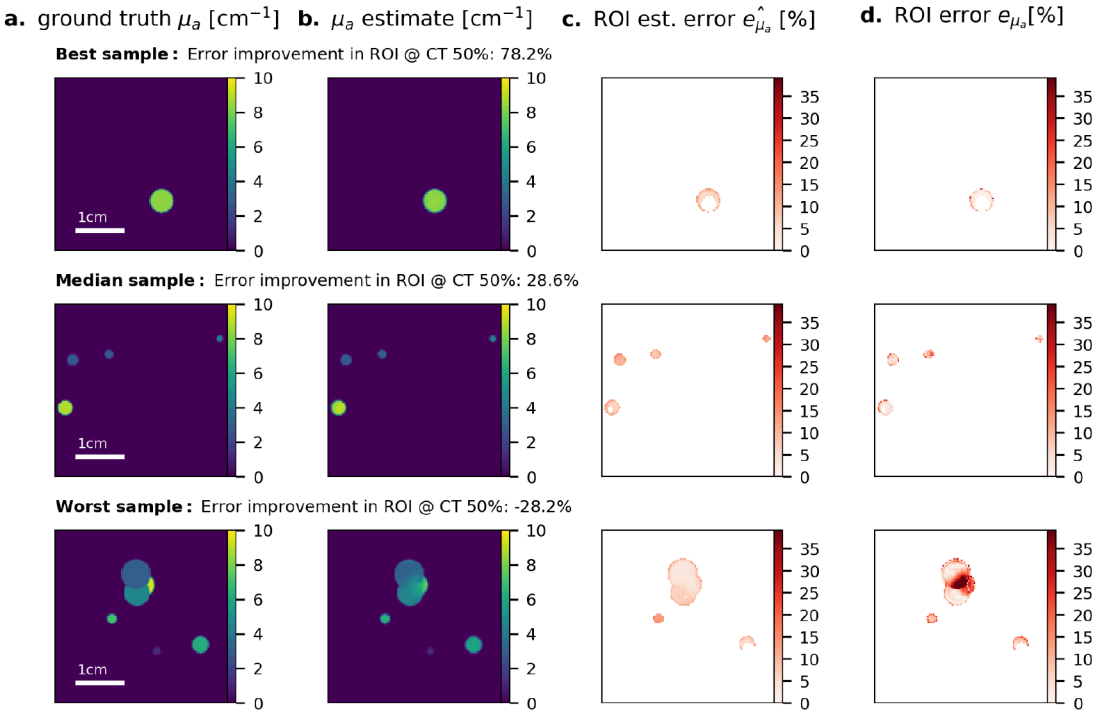

Figure 7. Sample images of the end-to-end direct μ_a quantification method applied to p_0 data, showing the best, the worst, and the median performance of our method when considering only the 50% most confident quantification estimations. All images show the (**a**) ground truth absorption coefficients (**b**) reconstructed absorption, (**c**) error estimate from external model, (**d**) the actual quantification error.

4. Discussion

In this work, we present a method which uses estimated confidences in the context of photoacoustic signal quantification to increase the accuracy of the quantification algorithms. In theory, the proposed method is independent of the underlying qPAI method, as it uses a deep learning model to observe the errors resulting from the quantification method in order to provide confidence estimates. While the application of our method to other state-of-the-art qPAI methods is the subject of future work, we aimed to show the general applicability of our method by also incorporating a

naïve fluence compensation method into the experiments. Our results suggest that using a method to estimate confidence information to refine a region of interest for subsequent computations might be a valuable tool for increasing the robustness of qPAI methods and could be easily integrated in future qPAI research.

We hypothesized that our confidence metric is indicative of e_{μ_a}. In the experiments, we showed that a deep learning model is able to learn a representation of the errors of the quantification method leading to error improvements of 10–50% in region-of-interest structures and yielding up to 5-fold improvements in background structures. Furthermore, Figure 5 shows that the absorption estimation error does not decrease monotonously, especially for the qPAI methods that yield more accurate results. One reason for this might be that the confidence estimates are not correlated to the quantification error and that low confidences might still correspond to low errors. One has to point out the worse performance of the quantification methods when applied directly to raw time-series data. One reason for this might be that the addition of the acoustic inverse problem and the inclusion of a realistic noise model greatly increased the complexity of the problem and reduced the amount of information in the data due to, e.g., limited-view artifacts. At the same time, we did not increase the number of training samples or change the methodology to account for this.

The dataset simulated for the experiments was specifically designed such that out-of-plane fluence effects cannot occur, as the in silico phantoms contain only straight tubular vessel structures that run orthogonal to the imaging plane. Additionally, other a priori assumptions of the parameter space were made, such as a constant background absorption, an overall constant scattering coefficient, and a fixed illumination geometry. Due to the homogeneous nature of the background structure, the errors observed in our study are highly specialized to our dataset. This is especially apparent with the direct estimation method, as here, e_{μ_a} is never greater than 0.2%. As such, we focus on reporting the errors in the ROI, as only reporting the results of the entire images would be misleading. In order for the method to generalize to more complex or in vitro datasets and yield similar μ_a and confidence estimation results, more elaborate and diverse datasets would need to be simulated. Nevertheless, the experiments demonstrate that applying an ROI threshold based on the estimation of the quantification error can lead to an increase in accuracy for a given dataset regardless of the underlying qPAI method.

From a qPAI perspective, end-to-end deep learning-based inversion of PA data is feasible in specific contexts and for specific in silico datasets, as shown previously [20] and in this work. However, PA signal quantification cannot be regarded as solved in a general manner. One of the main reasons is the large gap between simulated in silico data and in vivo recordings. In order for deep learning to tackle this problem, either highly sophisticated unsupervised domain adaptation methods have to be developed, or a large number of labeled correspondences between the simulation domain and real recorded images need to be provided, which is not currently feasible due to the lack of methodology to reliably measure ground truth optical properties in in vivo settings. However, with the promising progress in PA image reconstruction from limited-view geometries with deep learning techniques (cf., e.g., [57,58]), it might be possible to start bridging the gap and to improve on the current methods for qPAI.

Author Contributions: Conceptualization, J.G., T.K. and L.M.-H.; methodology, J.G., T.K. and T.A.; software, J.G.; data curation, J.G.; writing—original draft preparation, J.G.; writing—review and editing, J.G., T.K., T.A. and L.M.-H.; supervision, L.M.-H.

Funding: This project has received funding from the European Union's Horizon 2020 research and innovation programme through the ERC starting grant COMBIOSCOPY under grant agreement No. ERC-2015-StG-37960.

Acknowledgments: The authors would like to thank David Zimmerer for letting us tap into his well of knowledge on deep learning, Clemens Hentschke and the SIDT group for their support on the state of the art in non deep learning uncertainty quantification, Niklas Holzwarth for proofreading the manuscript, and the ITCF of the DKFZ for the provision of their computing cluster for data simulation.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CNR	Contrast-To-Noise-Ratio
PA	Photoacoustic
PAI	Photoacoustic Imaging
qPAI	quantitative PAI
ROI	Region of Interest
CT	Confidence Threshold

Appendix A. Results for the Naïve Fluence Compensation Method

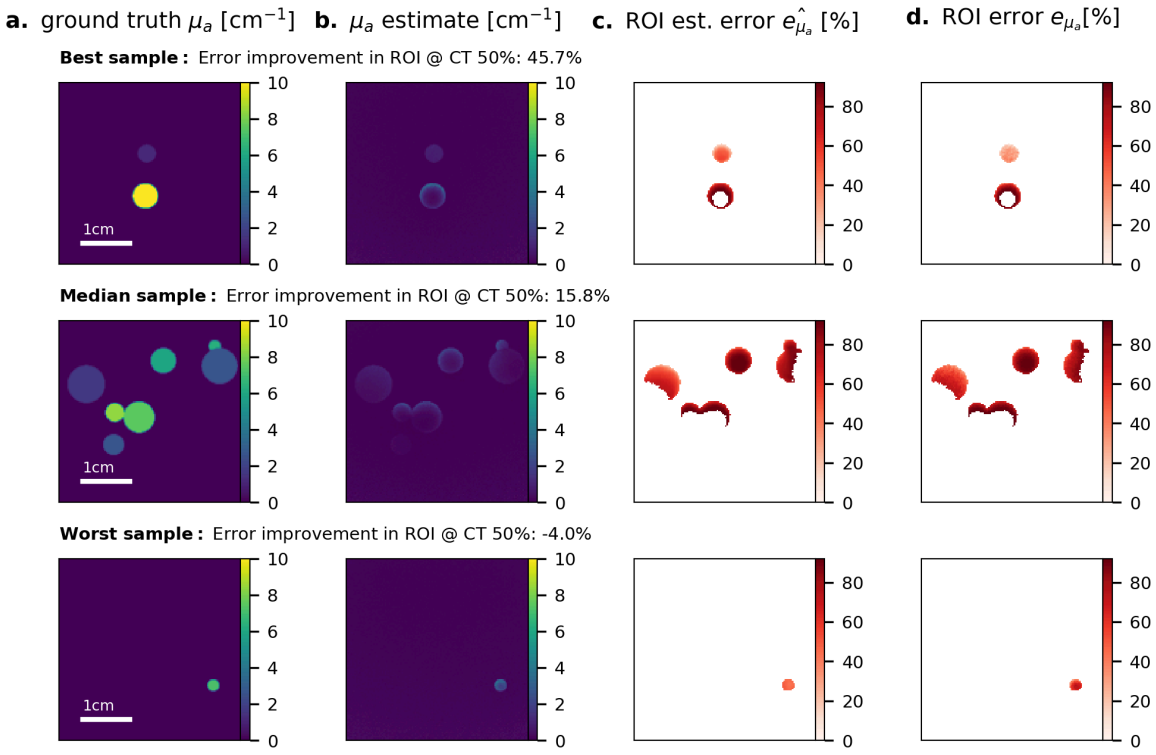

Figure A1. Sample images showing the best, the worst, and the median performance of our method when considering only the 50% most confident quantification estimations. All images show the (**a**) ground truth absorption coefficients (**b**) reconstructed absorption, (**c**) error estimate from external model, (**d**) the actual quantification error.

Appendix B. Results for Fluence Correction on p_0 Data

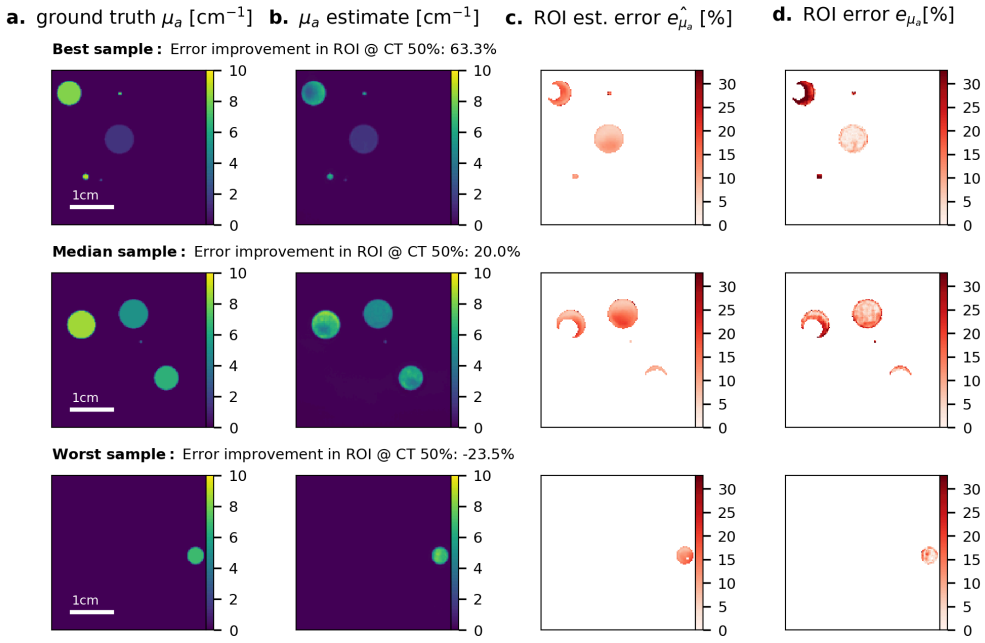

Figure A2. Sample images showing the best, the worst, and the median performance of our method when considering only the 50% most confident quantification estimations. All images show the (**a**) ground truth absorption coefficients (**b**) reconstructed absorption, (**c**) error estimate from external model, (**d**) the actual quantification error.

Appendix C. Results for Fluence Correction on Raw PA Time Series Data

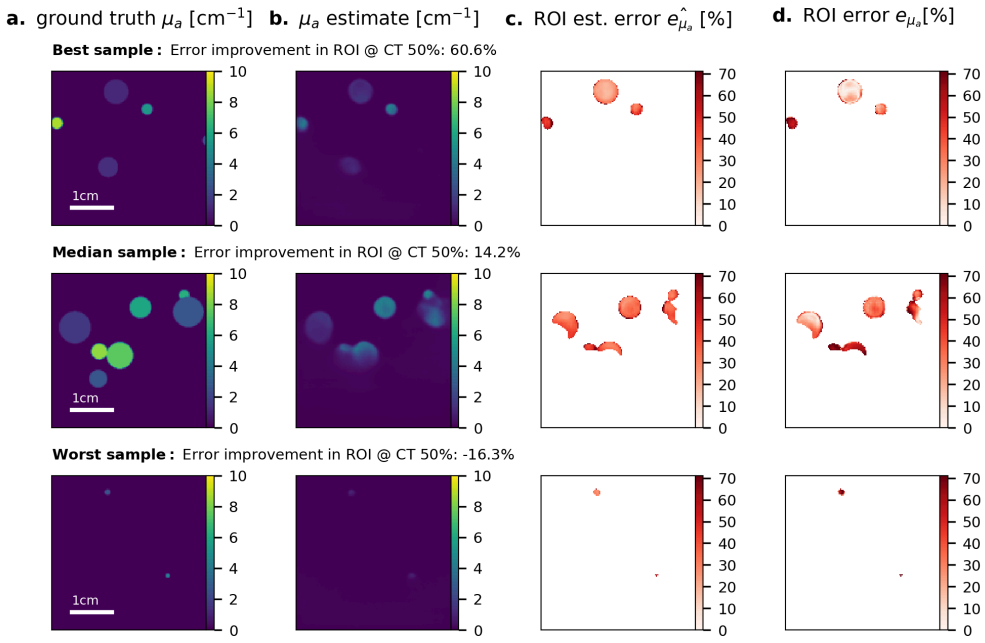

Figure A3. Sample images showing the best, the worst, and the median performance of our method when considering only the 50% most confident quantification estimations. All images show the (**a**) ground truth absorption coefficients (**b**) reconstructed absorption, (**c**) error estimate from external model, (**d**) the actual quantification error.

Appendix D. Results for Direct μ_a Estimation on Raw PA Time Series Data

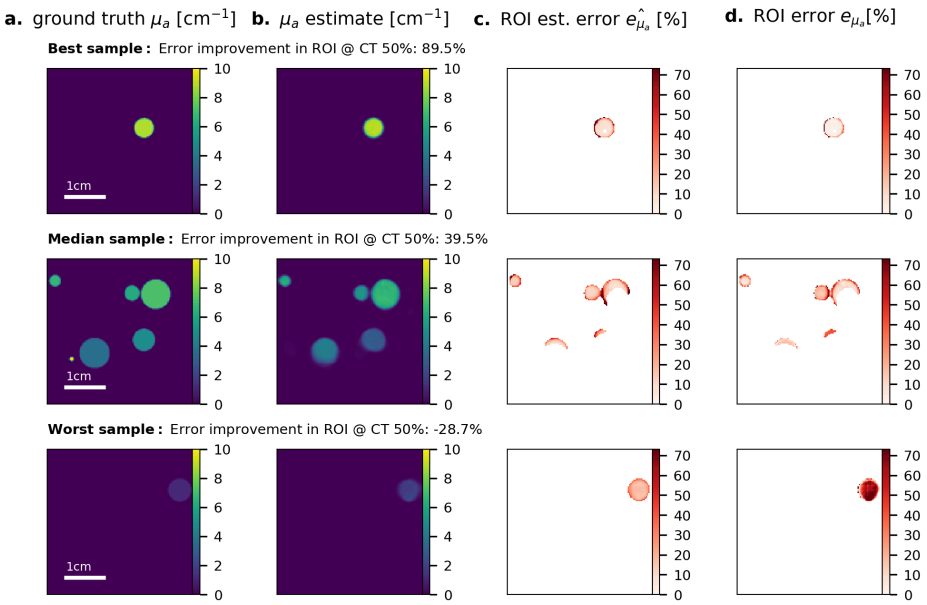

Figure A4. Sample images showing the best, the worst, and the median performance of our method when considering only the 50% most confident quantification estimations. All images show the (**a**) ground truth absorption coefficients (**b**) reconstructed absorption, (**c**) error estimate from external model, (**d**) the actual quantification error.

References

1. Valluru, K.S.; Willmann, J.K. Clinical photoacoustic imaging of cancer. *Ultrason* **2016**, *35*, 267–280. [CrossRef] [PubMed]
2. Knieling, F.; Neufert, C.; Hartmann, A.; Claussen, J.; Urich, A.; Egger, C.; Vetter, M.; Fischer, S.; Pfeifer, L.; Hagel, A.; et al. Multispectral Optoacoustic Tomography for Assessment of Crohn's Disease Activity. *N. Engl. J. Med.* **2017**, *376*, 1292–1294. [CrossRef] [PubMed]
3. Laufer, J. Photoacoustic Imaging: Principles and Applications. In *Quantification of Biophysical Parameters in Medical Imaging*; Springer: Cham, The Netherlands, 2018; pp. 303–324. [CrossRef]
4. Mitcham, T.; Taghavi, H.; Long, J.; Wood, C.; Fuentes, D.; Stefan, W.; Ward, J.; Bouchard, R. Photoacoustic-based SO_2 estimation through excised bovine prostate tissue with interstitial light delivery. *Photoacoust* **2017**, *7*, 47–56. [CrossRef] [PubMed]
5. Jacques, S.L. Optical properties of biological tissues: A review. *Phys. Med. Boil.* **2013**, *58*, R37. [CrossRef] [PubMed]
6. Tzoumas, S.; Nunes, A.; Olefir, I.; Stangl, S.; Symvoulidis, P.; Glasl, S.; Bayer, C.; Multhoff, G.; Ntziachristos, V. Eigenspectra optoacoustic tomography achieves quantitative blood oxygenation imaging deep in tissues. *Nat. Commun.* **2016**, *7*, 12121. [CrossRef] [PubMed]
7. Cox, B.T.; Arridge, S.R.; Köstli, K.P.; Beard, P.C. Two-dimensional quantitative photoacoustic image reconstruction of absorption distributions in scattering media by use of a simple iterative method. *Appl. Opt.* **2006**, *45*, 1866–1875. [CrossRef] [PubMed]
8. Cox, B.; Laufer, J.; Beard, P. The challenges for quantitative photoacoustic imaging. Photons Plus Ultrasound: Imaging and Sensing. *Int. Soc. Opt. Photonics* **2009**, *7177*, 717713.
9. Cox, B.; Laufer, J.G.; Arridge, S.R.; Beard, P.C. Quantitative spectroscopic photoacoustic imaging: A review. *J. Biomed. Opt.* **2012**, *17*, 061202. [CrossRef]
10. Yuan, Z.; Jiang, H. Quantitative photoacoustic tomography: Recovery of optical absorption coefficient maps of heterogeneous media. *Appl. Phys. Lett.* **2006**, *88*, 231101. [CrossRef]
11. Yuan, Z.; Jiang, H. Quantitative photoacoustic tomography. *Philos. Trans. R. Soc. Lond. A Math. Phys. Eng. Sci.* **2009**, *367*, 3043–3054. [CrossRef]

12. Wang, Y.; He, J.; Li, J.; Lu, T.; Li, Y.; Ma, W.; Zhang, L.; Zhou, Z.; Zhao, H.; Gao, F. Toward whole-body quantitative photoacoustic tomography of small-animals with multi-angle light-sheet illuminations. *Biomed. Opt. Express* **2017**, *8*, 3778–3795. [CrossRef] [PubMed]
13. Saratoon, T.; Tarvainen, T.; Cox, B.; Arridge, S. A gradient-based method for quantitative photoacoustic tomography using the radiative transfer equation. *Inverse Probl.* **2013**, *29*, 075006. [CrossRef]
14. Tarvainen, T.; Pulkkinen, A.; Cox, B.T.; Arridge, S.R. Utilising the radiative transfer equation in quantitative photoacoustic tomography. *Photons Plus Ultrasound Imaging Sens.* **2017**, *10064*, [CrossRef]
15. Haltmeier, M.; Neumann, L.; Rabanser, S. Single-stage reconstruction algorithm for quantitative photoacoustic tomography. *Inverse Probl.* **2015**, *31*, 065005. [CrossRef]
16. Kaplan, B.A.; Buchmann, J.; Prohaska, S.; Laufer, J. Monte-Carlo-based inversion scheme for 3D quantitative photoacoustic tomography. Photons Plus Ultrasound: Imaging and Sensing. *Int. Soc. Opt. Photonics* **2017**, *10064*, 100645J.
17. Tzoumas, S.; Ntziachristos, V. Spectral unmixing techniques for optoacoustic imaging of tissue pathophysiology. *Philos. Trans. R. Soc. A* **2017**, *375*, 20170262. [CrossRef] [PubMed]
18. Perekatova, V.; Subochev, P.; Kleshnin, M.; Turchin, I. Optimal wavelengths for optoacoustic measurements of blood oxygen saturation in biological tissues. *Biomed. Opt. Express* **2016**, *7*, 3979–3995. [CrossRef]
19. Glatz, J.; Deliolanis, N.C.; Buehler, A.; Razansky, D.; Ntziachristos, V. Blind source unmixing in multi-spectral optoacoustic tomography. *Opt. Express* **2011**, *19*, 3175–3184. [CrossRef]
20. Cai, C.; Deng, K.; Ma, C.; Luo, J. End-to-end deep neural network for optical inversion in quantitative photoacoustic imaging. *Opt. Lett.* **2018**, *43*, 2752–2755. [CrossRef]
21. Kirchner, T.; Gröhl, J.; Maier-Hein, L. Context encoding enables machine learning-based quantitative photoacoustics. *J. Biomed. Opt.* **2018**, *23*, 056008. [CrossRef]
22. Fonseca, M.; Saratoon, T.; Zeqiri, B.; Beard, P.; Cox, B. Sensitivity of quantitative photoacoustic tomography inversion schemes to experimental uncertainty. SPIE BiOS. *Int. Soc. Opt. Photonics* **2016**, *9708*, 97084X.
23. Maier-Hein, L.; Franz, A.M.; Santos, T.R.D.; Schmidt, M.; Fangerau, M.; Meinzer, H.; Fitzpatrick, J.M. Convergent Iterative Closest-Point Algorithm to Accomodate Anisotropic and Inhomogenous Localization Error. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 1520–1532. [CrossRef] [PubMed]
24. Alterovitz, R.; Branicky, M.; Goldberg, K. Motion Planning Under Uncertainty for Image-guided Medical Needle Steering. *Int. J. Robot. Res.* **2008**, *27*, 1361–1374. [CrossRef] [PubMed]
25. Sykes, J.R.; Brettle, D.S.; Magee, D.R.; Thwaites, D.I. Investigation of uncertainties in image registration of cone beam CT to CT on an image-guided radiotherapy system. *Phys. Med. Boil.* **2009**, *54*, 7263–7283. [CrossRef] [PubMed]
26. Risholm, P.; Janoos, F.; Norton, I.; Golby, A.J.; Wells, W.M. Bayesian characterization of uncertainty in intra-subject non-rigid registration. *Med. Image Anal.* **2013**, *17*, 538–555. [CrossRef] [PubMed]
27. Nair, T.; Precup, D.; Arnold, D.L.; Arbel, T. Exploring Uncertainty Measures in Deep Networks for Multiple Sclerosis Lesion Detection and Segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Granada, Spain, 16–20 September 2018; Springer: Berlin, Germany, 2018; pp. 655–663.
28. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
29. Gal, Y.; Ghahramani, Z. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In Proceedings of the International Conference on Machine Learning, New York, NY, USA, 19–24 June 2016; pp. 1050–1059.
30. Li, Y.; Gal, Y. Dropout Inference in Bayesian Neural Networks with Alpha-divergences. *arXiv* **2017**, arXiv:1703.02914.
31. Leibig, C.; Allken, V.; Ayhan, M.S.; Berens, P.; Wahl, S. Leveraging uncertainty information from deep neural networks for disease detection. *Sci. Rep.* **2017**, *7*, 17816. [CrossRef]
32. Feindt, M. A Neural Bayesian Estimator for Conditional Probability Densities. *arXiv* **2004**, arXiv:Phys./0402093.
33. Zhu, Y.; Zabaras, N. Bayesian deep convolutional encoder-decoder networks for surrogate modeling and uncertainty quantification. *J. Comput. Phys.* **2018**, *366*, 415–447. [CrossRef]
34. Kohl, S.A.; Romera-Paredes, B.; Meyer, C.; De Fauw, J.; Ledsam, J.R.; Maier-Hein, K.H.; Eslami, S.; Rezende, D.J.; Ronneberger, O. A Probabilistic U-Net for Segmentation of Ambiguous Images. *arXiv* **2018**, arXiv:1806.05034.

35. Kingma, D.P.; Welling, M. Auto-encoding variational bayes. *arXiv* **2013**, arXiv:1312.6114.
36. Mescheder, L.; Nowozin, S.; Geiger, A. Adversarial variational bayes: Unifying variational autoencoders and generative adversarial networks. *arXiv* **2017**, arXiv:1701.04722.
37. Ardizzone, L.; Kruse, J.; Wirkert, S.; Rahner, D.; Pellegrini, E.W.; Klessen, R.S.; Maier-Hein, L.; Rother, C.; Köthe, U. Analyzing Inverse Problems with Invertible Neural Networks. *arXiv* **2018**, arXiv:1808.04730.
38. Lakshminarayanan, B.; Pritzel, A.; Blundell, C. Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles. In *Advances in Neural Information Processing Systems 30*; Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2017; pp. 6402–6413.
39. Smith, L.; Gal, Y. Understanding Measures of Uncertainty for Adversarial Example Detection. *arXiv* **2018**, arXiv:1803.08533.
40. Pulkkinen, A.; Cox, B.T.; Arridge, S.R.; Goh, H.; Kaipio, J.P.; Tarvainen, T. Direct estimation of optical parameters from photoacoustic time series in quantitative photoacoustic tomography. *IEEE Trans. Med. Imaging* **2016**, *35*, 2497–2508. [CrossRef] [PubMed]
41. Pulkkinen, A.; Cox, B.T.; Arridge, S.R.; Kaipio, J.P.; Tarvainen, T. Estimation and uncertainty quantification of optical properties directly from the photoacoustic time series. Photons Plus Ultrasound: Imaging and Sensing 2017. *Int. Soc. Opt. Photonics* **2017**, *10064*, 100643N.
42. Tick, J.; Pulkkinen, A.; Tarvainen, T. Image reconstruction with uncertainty quantification in photoacoustic tomography. *J. Acoust. Soc. Am.* **2016**, *139*, 1951–1961. [CrossRef] [PubMed]
43. Tick, J.; Pulkkinen, A.; Tarvainen, T. Photoacoustic image reconstruction with uncertainty quantification. In *EMBEC & NBC 2017*; Eskola, H., Väisänen, O., Viik, J., Hyttinen, J., Eds.; Springer: Singapore, 2018; pp. 113–116.
44. Gröhl, J.; Kirchner, T.; Maier-Hein, L. Confidence estimation for quantitative photoacoustic imaging. Photons Plus Ultrasound: Imaging and Sensing 2018. *Int. Soc. Opt. Photonics* **2018**, *10494*, 104941C. [CrossRef]
45. Welvaert, M.; Rosseel, Y. On the Definition of Signal-To-Noise Ratio and Contrast-To-Noise Ratio for fMRI Data. *PLoS ONE* **2013**, *8*, e77089. doi:10.1371/journal.pone.0077089. [CrossRef] [PubMed]
46. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*; Springer: Berlin, Germany, 2015; Volume 9351, pp. 234–241.
47. Paszke, A.; Gross, S.; Chintala, S.; Chanan, G.; Yang, E.; DeVito, Z.; Lin, Z.; Desmaison, A.; Antiga, L.; Lerer, A. Automatic Differentiation in PyTorch. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017.
48. Isensee, F.; Petersen, J.; Klein, A.; Zimmerer, D.; Jaeger, P.F.; Kohl, S.; Wasserthal, J.; Koehler, G.; Norajitra, T.; Wirkert, S.; et al. nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation. *arXiv* **2018**, arXiv:1809.10486.
49. Bauer, A.Q.; Nothdurft, R.E.; Erpelding, T.N.; Wang, L.V.; Culver, J.P. Quantitative photoacoustic imaging: Correcting for heterogeneous light fluence distributions using diffuse optical tomography. *J. Biomed. Opt.* **2011**, *16*, 096016–096016. [CrossRef]
50. Daoudi, K.; Hussain, A.; Hondebrink, E.; Steenbergen, W. Correcting photoacoustic signals for fluence variations using acousto-optic modulation. *Opt. Express* **2012**, *20*, 14117–14129. [CrossRef] [PubMed]
51. Jacques, S.L. Coupling 3D Monte Carlo light transport in optically heterogeneous tissues to photoacoustic signal generation. *Photoacoust* **2014**, *2*, 137–142. [CrossRef] [PubMed]
52. Treeby, B.E.; Cox, B.T. k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields. *J. Biomed. Opt.* **2010**, *15*, 021314. [CrossRef] [PubMed]
53. Waibel, D.; Gröhl, J.; Isensee, F.; Kirchner, T.; Maier-Hein, K.; Maier-Hein, L. Reconstruction of initial pressure from limited view photoacoustic images using deep learning. Photons Plus Ultrasound: Imaging and Sensing 2018. *Int. Soc. Opt. Photonics* **2018**, *10494*, 104942S. [CrossRef]
54. Kirchner, T.; Wild, E.; Maier-Hein, K.H.; Maier-Hein, L. Freehand photoacoustic tomography for 3D angiography using local gradient information. *Photons Plus Ultrasound Imaging Sens.* **2016**, *9708*, 97083G. [CrossRef]
55. Gröhl, J.; Kirchner, T.; Adler, T.; Maier-Hein, L. *In Silico 2D Photoacoustic Imaging Data*; Zenodo: Meyrin, Switzerland, 2018. [CrossRef]

56. Zimmerer, D.; Petersen, J.; Koehler, G.; Wasserthal, J.; Adler, T.; Wirkert, A. *MIC-DKFZ/Trixi: Pre-Release*; Zenodo: Meyrin, Switzerland, 2018. doi:10.5281/zenodo.1345137.
57. Hauptmann, A.; Lucka, F.; Betcke, M.; Huynh, N.; Adler, J.; Cox, B.; Beard, P.; Ourselin, S.; Arridge, S. Model-Based Learning for Accelerated, Limited-View 3-D Photoacoustic Tomography. *IEEE Trans. Med. Imaging* **2018**, *37*, 1382–1393. [CrossRef]
58. Antholzer, S.; Haltmeier, M.; Schwab, J. Deep learning for photoacoustic tomography from sparse data. *Inverse Probl. Sci. Eng.* **2018**. [CrossRef]