Preprocessing of Freehand Ultrasound Synthetic Aperture Measurements using DNN

Rick Pandey*†, Jan Kirchhof*†, Fabian Krieg*† ‡, Eduardo Pérez*†, Florian Römer†
*Technische Universität Ilmenau, Ilmenau, Germany
†Fraunhofer Institute for Nondestructive Testing IZFP, Saarbrücken, Germany
‡htw saar, Saarbrücken, Germany

***Abstract*—Manual ultrasonic inspection is a widely used Nondestructive Testing (NDT) technique due to its simplicity and compatibility with complex structures. However, in contrast to the data acquired using a robotic positioner, manual measurements suffer from perturbations caused by a variable coupling and a varying scanning density. Imaging techniques like the synthetic aperture focusing technique rely on an unperturbed dense measurement from an equidistant measurement grid. Consequently, imaging based on freehand measurements leads to artifacts. This work aims at reducing such artifacts by preprocessing the manual measurements using Deep Neural Networks (DNN). The training of a DNN requires a large set of labeled measurements which is difficult to obtain in NDT. In this work, we present a technique to train the DNN using only synthetic data. We show that the resulting DNN generalizes well on real measurements. We present an improvement in Generalized Contrast to Noise Ratio by a factor of 20 and 3 compared to omitting the preprocessing for synthetic and measurement data, respectively.**

***Index Terms*—Deep Neural Network, Ultrasound NDT, Freehand Measurements,**

I. Introduction and State of the Art

Ultrasound Nondestructive Testing (NDT) is widely used to evaluate material properties and detect harmful defects without destroying or impairing the component being tested [1]. A common way of investigation is to compute images from single channel synthetic aperture measurements using the Synthetic Aperture Focusing Technique (SAFT) [2]. Typically, SAFT is only employed on automated measurements taken from a regular equidistant grid. However, in many scenarios measurements need to be taken manually by an engineer as employing an automated setup would not be cost effective and requires a large effort in programming and configuration to be compatible with the specimen. By using an assistance system to track the scanning positions during these manual measurements, imaging becomes possible [3]. SAFT can be formulated in a progressive manner to create a reconstruction image that is updated with each newly acquired scan. Naturally, these manual scanning positions will not form a regular equidistant grid leading to artifacts in the reconstruction caused by the unbalanced scanning density [4]. In addition to this, further artifacts arise if there are variations in the scans due to varying coupling between the transducer and the specimen surface, e.g., due to surface roughness. Still, a valid approach is to view the freehand scans as sparse perturbed measurements taken on a denser grid. This grid can for example arise from the discrete output of the tracking system.

In the medical community, freehand 3D ultrasound imaging systems have been developed for decades [5]. However, in contrast to our work, a standard approach is to interpolate the 3D volumes from 2D beamforming images [6] instead of employing a 3D reconstrution scheme. In NDT, the Region of Interest (ROI) is usually much larger than in medical scenarios, making synthetic aperture approaches more attractive than the use of arrays. Further, since the goal is to detect small discontinuities in a large ROI, a progressive reconstruction scheme that locally updates the 3D image based on incoming measurements is required in contrast to computing a complete image of a single target object as typical in medicine. This has the additional benefit that NDT scenarios allow for a much simpler acquisition setup, only requiring a single channel transducer and a 2D surface tracking system that could be implemented using a single camera [4].

Deep learning is a leading edge of artificial intelligence that allows computer programs to train themselves in order to process and learn from available data [7]. Recently, it has been shown that the image quality can be maintained by using a Deep Neural Network (DNN) to perform imaging from undersampled synthetic aperture data [8], [9] compared to conventional imaging of the fully sampled data. In [10] a DNN is used to estimate the motion of the probe of a freehand ultrasound system based on the ultrasound measurement data.

In more related work, DNNs have been employed to interpolate measurements of irregularly subsampled seismic data [11]. A U-Net that is trained on randomly subsampled data is used to reconstruct the missing measurements. In [12], this method was extended to the reconstruction of 3D volumes using 3D convolutional neural networks.

In this paper, we propose a DNN based preprocessing technique which aims at the prediction of missing scans in the proximity of the current acquired measurement leading to a locally fully sampled and adjusted data set that is then used as input data to compute a SAFT image. A natural way to train such a DNN would be to manually subsample densely recorded data and then train the network to interpolate the missing samples. However, this requires a large set of

Eduardo Pérez and Fabian Krieg are funded by DFG under the project "CoSMaDU" - GA 2062/5-1. This work was supported by the Fraunhofer Internal Programs under Grant No. Attract 025-601128.

labeled measurements which are difficult to obtain in NDT. Hence, in contrast to [11], [12], we train our model using a data set comprising only synthesized measurements using a rather simplified forward model. The different propagation modes are separable, enabling techniques that address a single mode at a time [13]. Due to this, multiple scattering can be incorporated via ray tracing, where each ray can be assigned a refraction angle based on the desired mode. Furthermore, when the geometry is simple and scatterers are well spaced from each other and from the specimen boundaries, a single scattering event suffices. We demonstrate that the DNN model in fact generalizes well on real measurement data, resulting in fewer artifacts in the SAFT images based on the pre-processed measurements compared to the SAFT images using only the sparse measurement data. In addition, the local ratio between measured and predicted scans can be used as a further feedback to the engineer to point out regions with high uncertainty (i.e. few measurements).

II. SAFT AND FREEHAND MEASUREMENTS

Assume that a single ultrasound transducer is placed at position $(x, y, z = 0)$ on the surface of a specimen. In a handheld scenario, the position of the transducer is tracked by a tracking system, e.g. by using a camera. The specimen is assumed to be homogeneous and isotropic with constant speed of sound c. The reflections of an inserted pulse measured by that same transducer can then be modeled as

$$g_{x,y}(t) = \sum_{d=1}^{D} b_d h(t - \tau_{x,y}(x_d, y_d, z_d)) + n_{x,y}(t). \quad (1)$$

Here, $h(t)$ is the inserted ultrasound pulse shape, D is the number of reflectors, $\tau_{x,y}(x_d, y_d, z_d)$ is the time of flight computed as

$$\tau_{x,y}(x_d, y_d, z_d) = \frac{2}{c}\sqrt{(x - x_d)^2 + (y - y_d)^2 + z_d^2},$$

b_d the reflection coefficient, and $n_{x,y}(t)$ comprises the measurement noise. The specimen is scanned from a set of measurement positions $\mathcal{M}$ distributed on the specimen's surface. From these measurements, the 3-D SAFT image $\boldsymbol{r}$ is then computed as [4]

$$[\boldsymbol{r}]_{x_r,y_r,z_r} = \sum_{\forall x,y \in \mathcal{M}} a_{x,y}(x_r, y_r, z_r) g_{x,y}\left(\tau_{x_y}(x_r, y_r, z_r)\right), \quad (2)$$

with $a_{x,y}(x_r, y_r, z_r)$ being an apodization function that down-weights scans with increasing distance from the assumed defect position (x_r, y_r, z_r). In this work, a Gaussian apodization as in [14] is used. Eq. (2) can be implemented as an iterative scheme that updates $\boldsymbol{r}$ with every newly acquired measurement. In an automated setting, the positions in $\mathcal{M}$ form an equidistant grid. In contrast, in a handheld measurements, the scanning positions will be unevenly distributed on the surface as illustrated in Fig. 1, tracked only with the limited accuracy of the employed tracking system. Further, additional variations between scans caused by varying coupling can occur due to the manual movement of the transducer. Images reconstructed by SAFT will contain artifacts due to these perturbations [3]. These manual measurement challenges are very subtle and cannot be avoided even with extreme care. Hence, it is necessary to preprocess the acquired measurements to improve the robustness, which results in artifact reduction in the reconstructed image.

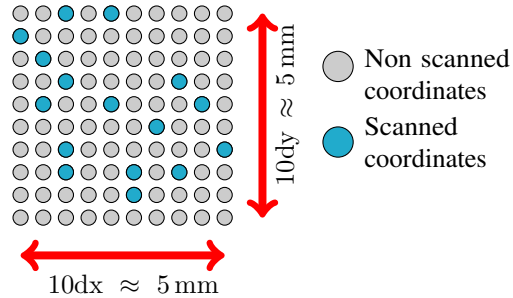

Fig. 1: Illustration of handheld ultrasonic measurements in a local sub-region

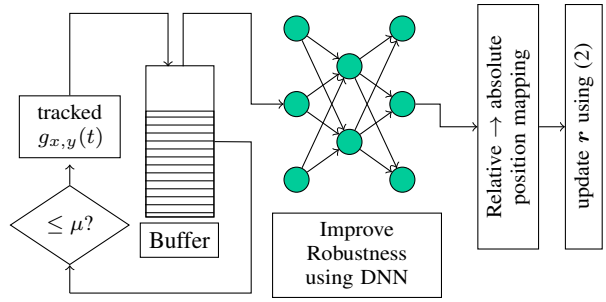

Fig. 2: Proposed processing chain

III. PROPOSED DNN MODEL

To account for the challenges discussed in Sec. II, we introduce a preprocessing step using a DNN. In order to address these, we need more than one measurement in a local neighborhood. Hence, a buffer is used to collect the measurements from the acquisition system. Once a sufficient number of scans are available in a defined region, the DNN is used to interpolate the missing scans as well as correct the time shifts in the acquired measurements. To track sufficient measurements in a local region, we introduce a threshold μ. Fig. 2 represents a block diagram of the approach used in this work. If the buffer accumulates μ scans, the measurements with their relative positions in the defined local region are passed on to DNN. The trained model replicates a U-net architecture [7, chap. 14] trained at regions with and without defects. The model solves a regression problem in time domain by learning a function $f : \mathbb{R}^{n_x \times n_y \times n_t} \rightarrow \mathbb{R}^{n_x \times n_y \times n_t}$, where the input to the function is a sparse 3D array containing the available measurement data on a $n_x \mathrm{d}x \times n_y \mathrm{d}y$ grid and the output of the function is a dense 3D array with missing measurements predicted. Here, $\mathrm{d}x$ and $\mathrm{d}y$ are the grid spacing in x and y direction, respectively, and n_t is the number of time samples per scan. Fig. 1 provides an illustration of a typical handheld measurement scenario in a local region. We assume that the prediction can be performed relative to the current local measurements and is therefore independent of the absolute position on the specimen's surface, i.e. a single DNN is used

globally for the prediction. The scans predicted by the DNN are fed to the online SAFT reconstruction.

IV. Data Preparation and Training of DNN

The training of a DNN requires careful planning and data preparation. The high level training parameters used in this

Training Cluster	TU Ilmenau Cluster , on CPU
Deep Learning Library	Keras
Optimizer	Adam
Loss Function	Mean squared error
Validation metrics	Accuracy
Batch Size	10, 30, 100
Weight Initialization	Random (uniform distribution)

TABLE I: Training parameters for proposed DNN model

work are listed in Table I, and the procedure is detailed next.

A. Simulated Scans

In order to train the proposed model we need a large dataset of 3D measurements of a given dimension. It is difficult to acquire sufficient and versatile enough real measurements for such a scenario. DNNs easily overfit if the training data do not contain the full spectrum of situations that may occur in real measurements. For example, the model can be overfit for a defect location as we cannot fabricate defects at all locations possible in real scenarios. Hence, to ensure that the training dataset is suitable, we synthesize measured scans following the model in (1) for $h(t)$ defined as

$$h(t) = e^{-\alpha(t)t^2} \cos(2\pi f_c t + \phi), \tag{3}$$

where f_c, is the carrier frequency, $\phi \in [0, \pi]$ is the phase, and $\alpha(t) = f_c^2(1 - \mathrm{sgn}(t) \cdot \gamma)$. The parameter $\gamma \in [0, 1]$ steers the asymmetry of the envelope. As $\gamma \to 1$ the scan becomes more asymmetric. An example is depicted in Fig. 3. Through observations from real measurements and industrial expertise we arrived at a conclusion that 0.7 should be a good fit for γ.

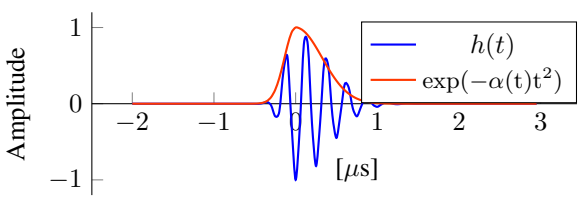

Fig. 3: Simulated ultrasound echo

B. Simulated Training Data sets

Based on the 1-D model specified in the previous section, we synthesize M 3-D arrays $\boldsymbol{X}_m \in \mathbb{R}^{n_x \times n_y \times n_t}$, $m = 1, 2, \ldots, M$ as our training data set. Each data set contains various defects positioned at random but mutually distant locations, such that there are no overlapping echoes in any scan. Each measurement is then perturbed with additive white Gaussian noise as well as random offsets along the z-axis to simulate phase shifts due to varying coupling to form a perturbed measurement $\boldsymbol{X}'_m$. To simulate the manual measurements, we assume that freehand measurements will lead to scanning positions distributed over the whole (sub)-region, but with varying density.

Recalling Fig. 1, this excludes a scenario where only the grid positions at the edge of the region are scanned (forming a perfect square) and all remaining points are untouched, and any other worst-case scenario, where μ scans are concentrated at a single location. In contrast, a valid scenario comprises for example perturbed scans at the blue coordinates that are input to the DNN with the goal to predict the unperturbed scans at the gray coordinates. The prediction can only work properly for features of the specimen present in a sufficient number of physical measurements. In theory, to detect a single point source in 3-D space, 4 scans are enough. However, μ needs to ensure that echoes from this point source are present in at least 4 scans taken in a subregion. Its concrete value therefore depends on the minimum defect size that needs to be detected as well as the employed grid spacing.

Based on this, we simulated the manual measurements by subsampling every $\boldsymbol{X}'_m$ along x and y uniformly at random to keep only μ measurements that are input to the DNN by setting everything else to zero. The network is then trained in a supervised manner using the corresponding fully sampled $\boldsymbol{X}_m$ as ground truth, i.e., it is trained such that the mean squared error between its output $\hat{\boldsymbol{X}}_m = f(\boldsymbol{X}'_m)$ and the true $\boldsymbol{X}_m$ is minimized. During the training, we used 70% of M as training data and 30% as validation data.

V. Results

To test our proposed setup, we trained a U-net with the layer wise structure as depicted in Fig. 4. All the convolutional layers

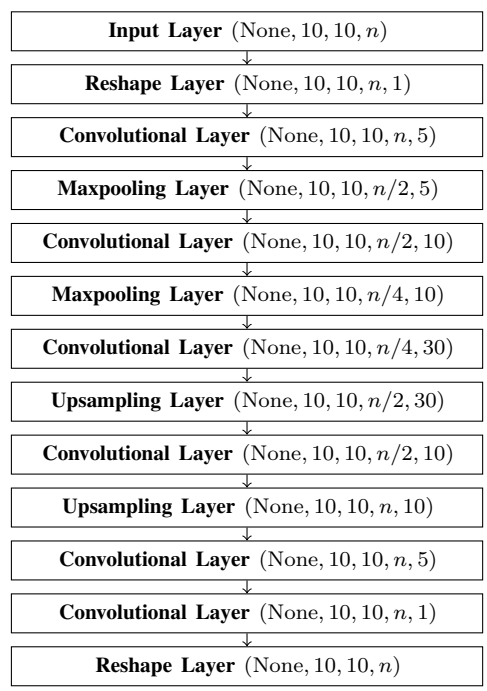

Fig. 4: Proposed U-net.

used in the model use convolution kernels of size $10 \times 10 \times 3$ and a *ReLU* as activation function except the last convolutional

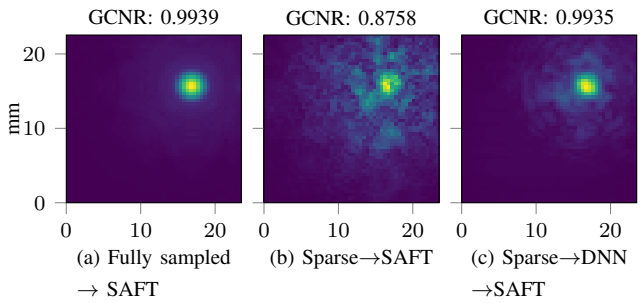

Fig. 5: Comparison of top view of reconstructions among fully sampled data, perturbed measurements and preprocessed measurements using the proposed DNN.

layer which uses no activation. Maxpooling is applied to the temporal axis, since ultrasound measurements are commonly oversampled and additionally admit a sparse representation under the model in (1). Each scan therefore contains only small number of relevant features. For the training, the $\boldsymbol{X}_M$ are created assuming aluminium as the material and based on the parameters listed in Table II.

(x_d, y_d, z_d)	$\mathcal{U}(0, x_n), \mathcal{U}(0, y_n), \mathcal{U}(0, z_n)$ m^3
M	69603
training data	70% of M
validation data	30% of M
epochs	5
$(n_x \times n_y \times n_t)$	$10 \times 10 \times 590$
dx, dy	$\mathcal{U}(0.4, 0.6)$ mm^2
time-shifts	$\mathcal{U}(0.084, 1.26)$ mm
occurence of z-offsets	$\mathcal{U}(25, 75)$ %
μ	$[4, 8, 12, 16]$ scans per 10×10 grid
computational time	14.1 hours

TABLE II: Data set parameters to train the proposed DNN. The data set comprises only simulated data generated using the forward model defined in (1) and (3).

The trained DNN is then applied to both synthetic and real measurements. Fully sampled synthetic measurements are generated for a single point source marking the defect using the same model as above and then artificially corrupted with noise and varying coupling. Based on this, we simulated the manual measurements by dividing the full data-set into sub-regions of size $n_x = n_y = 10$ and subsampling each sub-region uniformly at random keeping only μ scans per sub-region. The artificially perturbed dataset is then preprocessed using DNN. SAFT reconstruction is carried out for the fully sampled, the perturbed and the perturbed and additionally preprocessed measurements, respectively. To determine μ, we trained several networks on different threshold levels and compared their performance using the described procedure for several randomly generated synthetic data sets. By this, the lowest level leading to reliable performance was determined to be 16 for our settings. Fig. 5 represents the top view of these three reconstructions of the synthetic data. Clearly, the preprocessed measurements using DNN outperforms the reconstruction from perturbed data. We further evaluate the results using the GCNR [15]. The GCNR is a generalization of the standard Contrast-to-Noise-Ratio (CNR),

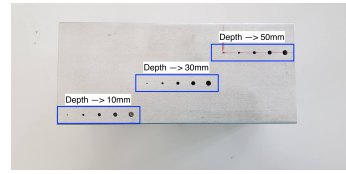

Fig. 6: Investigated aluminium specimen

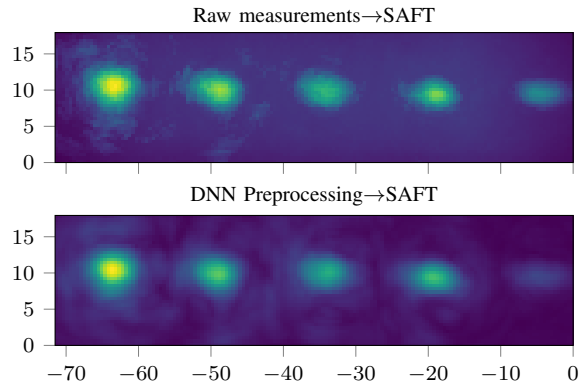

Fig. 7: Comparison of top view of preprocessed measurement and raw measurement reconstructions of the flat bottom holes at $50\,\mathrm{mm}$ depth.

to make it more robust to scaling and dynamic range issues. It is defined as

$$\mathrm{GCNR} = 1 - \int \min\left(p_+(x), p_-(x)\right) \mathrm{d}x,$$

where $p_+(x)$ and $p_-(x)$ are the probability density functions of the true positive and true negative regions, respectively. It yields a value in $[0, 1]$, where 1 means that there is no overlap between $p_+(x)$ and $p_-(x)$. The GCNR values for the reconstructions given in the title of Fig. 5 give a quantitative classification of the improvement achieved by using the DNN. The results show that the DNN preprocessing improves the GCNR from 0.8758 to 0.9935, which corresponds to an improvement by a factor of around 20.[1]

As a next step, we acquired freehand measurements from the aluminium specimen depicted in Fig. 6 using a conventional piezo-electric circular transducer with center frequency of $4\,\mathrm{MHz}$ as well as the 3DSmartInspect assistance system [16, Sec. 5] to test how well our model generalizes to real measurements. Subregions of size 10×10 grid points covering $25\,\mathrm{mm}^2$ were used for the preprocessing. The preprocessing was done offline after the full measurement was collected. Fig. 7 represents a comparison between the reconstruction of the raw freehand measurements of the five flat bottom holes at a depth of $50\,\mathrm{mm}$ and the measurements preprocessed using the DNN. The size of the defects ranges from $5\,\mathrm{mm}$ to $1\,\mathrm{mm}$ in diameter from left to right. To ensure a fair comparison of the reconstruction of the different defects in the specimen, in each subregion we randomly selected $\mu = 16$ scans as input to the DNN and discarded the remaining scans, if a subregion contained more scans than the threshold. Table III conveys a better chance of recognizing the defect by preprocessing the

[1] Since the GCNR is in $[0, 1]$ where 1 is the optimum, the GCNR improvement from g_1 to g_2 is measured as $(1 - g_1)/(1 - g_2)$.

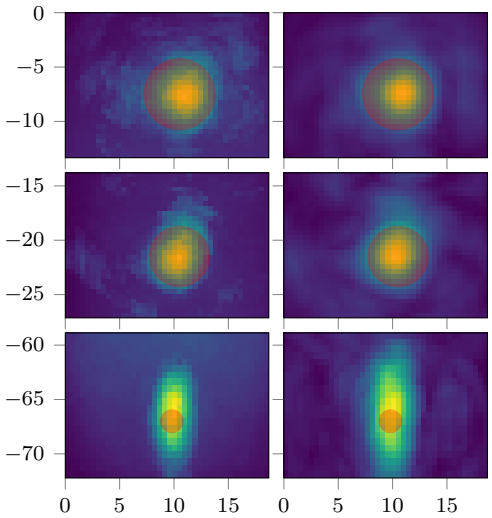

Fig. 8: Zoomed in comparison of the first, second and fifth defect (from left to right) in Fig. 7. Left: no preprocessing, right: with preprocessing. The circular overlay marks the true size of the defect. The reconstruction from raw measurements has more artifacts compared to reconstruction from preprocessed measurements using the proposed DNN.

measurements (e.g., with a factor 3 improvement in GCNR for the second defect). In all cases except for the smallest defect, the threshold was met and we witness an improvement in GCNR. Fig. 8 depicts a zoomed comparison for the defects of size $5\,\mathrm{mm}$, $4\,\mathrm{mm}$ and $1\,\mathrm{mm}$. The first and second column show the reconstruction from raw and preprocessed measurements, respectively. The subregion around the $1\,\mathrm{mm}$ hole in the third row did not meet the threshold, which resulted in a worse estimation of the defect dimension compared to the unprocessed reconstruction.

Data	w/o preprocessing	preprocessing
Defect 1	0.957	0.974
Defect 2	0.961	0.987
Defect 3	0.983	0.990
Defect 4	0.972	0.989
Defect 5	0.957	0.936

TABLE III: GCNR comparison for the five investigated defects

VI. Conclusion

The results clearly show improvements in the preprocessed reconstructed images compared to the state of the art SAFT results. The GCNR improves by a factor of 20 and 3 for the synthetic and measured data sets, respectively. As it is very difficult to obtain a huge labeled dataset, one of the important contributions of this work is the training of the model on synthetic data and its ability to generalize well on measurement data. Although the improvement in GCNR on the real measurements is lower as compared to the improvements on the synthetic measurements, it should be considered that the model is trained only for point sources, but tested on flat bottom holes of varying dimensions. Finally, the proposed preprocessing not only improves imaging by minimizing artifacts outside the true region, but adds an additional local uncertainty measure to the assistance system by comparing the local ratio of measured and predicted scans.

References

[1] B. Raj, T. Jayakumar, and M. Thavasimuthu. *Practical Non-Destructive Testing*. Woodhead Publishing Limited, 2002.

[2] M. Spies, H. Rieder, A. Dillhöfer, V. Schmitz, and W. Müller. Synthetic aperture focusing and time-of-flight diffraction ultrasonic imaging—past and present. *Journ. of NDE*, 31:310–323, 2012.

[3] F. Krieg, J. Kirchhof, S. Kodera, S. Lugin, A. Ihlow, T. Schwender, G. Del Galdo, F. Römer, and A. Osman. SAFT processing for manually acquired ultrasonic measurement data with 3D smartInspect. *Insight - The Journal of The British Institute of Non-Destructive Testing*, 61, November 2019.

[4] F. Krieg, S. Kodera, J. Kirchhof, F. Römer, A. Ihlow, S. Lugin, A. Osman, and G. D. Galdo. 3D reconstruction of handheld data by SAFT and the influence of measurement inaccuracies. In *2019 IEEE International Ultrasonics Symposium (IUS)*, pages 2095–2098, Oct 2019.

[5] M. H. Mozaffari and W.-S. Lee. Freehand 3-D ultrasound imaging: A systematic review. *Ultrasound in Medicine & Biology*, 43(10):2099–2124, 2017.

[6] T. Wen, Q. Zhu, W. Qin, L. Li, F. Yang, Y. Xie, and J. Gu. An accurate and effective FMM-based approach for freehand 3D ultrasound reconstruction. *Biomedical Signal Processing and Control*, 8(6):645–656, 2013.

[7] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[8] Y. H. Yoon, S. Khan, J. Huh, and J. C. Ye. Efficient B-mode ultrasound image reconstruction from sub-sampled RF data using deep learning. *IEEE Transactions on Medical Imaging*, 38(2):325–336, 2019.

[9] B. Luijten, R. Cohen, F. J. De Bruijn, H. A. W. Schmeitz, M. Mischi, Y. C. Eldar, and R. J. G. Van Sloun. Adaptive ultrasound beamforming using deep learning. *IEEE Transactions on Medical Imaging*, pages 1–1, 2020.

[10] R. Prevost, M. Salehi, S. Jagoda, N. Kumar, J. Sprung, A. Ladikos, R. Bauer, O. Zettinig, and W. Wein. 3D freehand ultrasound without external tracking using deep learning. *Medical image analysis*, 48:187–202, 2018.

[11] X. Chai, H. Gu, F. Li, H. Duan, X. Hu, and K. Lin. Deep learning for irregularly and regularly missing data reconstruction. *Scientific Reports*, 10(1):1–18, 2020.

[12] X. Chai, G. Tang, S. Wang, K. Lin, and R. Peng. Deep learning for irregularly and regularly missing 3-D data reconstruction. *IEEE Transactions on Geoscience and Remote Sensing*, pages 1–22, 2020.

[13] N. Budyn, R. L. T. Bevan, J. Zhang, A. J. Croxford, and P. D. Wilcox. A model for multiview ultrasonic array inspection of small two-dimensional defects. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 66(6):1129–1139, 2019.

[14] J. Kirchhof, S. Semper, and F. Römer. GPU-accelerated matrix-free 3D ultrasound reconstruction for nondestructive testing. In *Proceedings of the 2018 IEEE International Ultrasonics Symposium*, Kobe, Japan, October 2018.

[15] A. Rodriguez-Molares, O. Rindal, J. D'hooge, S. Måsøy, A. Austeng, M. Bell, and H. Torp. The generalized contrast-to-noise ratio: a formal definition for lesion detectability. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 67(4):745–759, 2019.

[16] B. Valeske, A. Osman, F. Römer, and R. Tschunky. Next generation NDE systems as IIoT elements of industry 4.0. *Research in Nondestructive Evaluation*, 2020.