

Direct Geometric Reconstruction of Flaws from Simulated Ultrasonic Inspection Data

Daniel Algernon

`daniel.algernon@svti.ch`

SVTI Swiss Association for Technical Inspections

Research Article

Keywords: ultrasonic testing, flaw characterization, inverse problem, geometric reconstruction, CIVA simulation, machine learning, deep learning, convolutional regression, nondestructive testing

Posted Date: October 27th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11219392/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Direct Geometric Reconstruction of Flaws from Simulated Ultrasonic Inspection Data

D. Algernon

SVTI Swiss Association for Technical Inspections, Nuclear Inspectorate, Laboratory for Nondestructive Evaluation (NDE), Switzerland

Corresponding author: Dr. Daniel Algernon

E-mail: daniel.algernon@svti.ch

Abstract

Quantitative ultrasonic flaw characterization requires inference of defect geometry from measurements that are only indirectly related to the underlying discontinuity. This study investigates whether, for a defined conventional single-element pulse-echo inspection, that inference can be formulated as a direct geometric inverse problem rather than as image reconstruction followed by interpretation. Physics-based CIVA simulations were used to generate paired ultrasonic responses and exact reference contours for inner-surface-breaking, faceted flaws in steel. A supervised convolutional regression model was trained to estimate an ordered flaw contour, from which engineering quantities such as depth and tilt can be derived. Model capacity was selected using validation performance in these physically interpretable quantities and assessed on held-out simulated cases. The results show that the principal geometric trends of the represented flaw family can be recovered directly from the ultrasonic response under the controlled conditions of the simulation. The study therefore provides a proof of concept for task-oriented geometric inversion in conventional ultrasonic testing and defines the principal requirements for subsequent experimental validation.

Keywords: ultrasonic testing; flaw characterization; inverse problem; geometric reconstruction; CIVA simulation; machine learning; deep learning; convolutional regression; nondestructive testing

1. Introduction

Engineering assets constitute fundamental elements of modern infrastructure, including energy-production facilities, aircraft, railway systems, petrochemical installations, and pipelines. Experience across safety-critical sectors has shown that material degradation can evolve over extended service periods before functional impairment becomes apparent. Reliable structural integrity management therefore depends on detecting and characterizing relevant degradation early enough to inform maintenance, continued operation, and safety decisions [1, 2, 3, 4].

Nondestructive testing (NDT), also commonly referred to as nondestructive evaluation (NDE), is central to this task because it enables the detection and characterization of material and structural discontinuities relevant to integrity assessment. Among established NDT methods, ultrasonic testing (UT) is particularly important as a volumetric inspection technique; elastic waves are introduced into the component and the received response reveals interactions with internal interfaces and discontinuities. Depending on flaw geometry and inspection configuration, these interactions include reflection, diffraction, mode conversion, and scattering [2, 3, 4].

Detection of an ultrasonic indication, however, is distinct from quantitative flaw characterization. Structural assessment requires information, such as position, depth, orientation, and extent. These quantities are not

measured directly; they must be inferred from a response that depends on transducer characteristics, material properties, beam angle, bandwidth, aperture, and the scattering behavior of the discontinuity. The resulting sizing problem is therefore an inverse problem, and the challenge of relating crack-tip, corner, and specular responses to flaw geometry is well established in ultrasonic practice [1, 5, 6, 7, 8].

In conventional single-element pulse-echo inspection, interpretation is commonly supported by spatially indexed A-scans and B-scan displays, and, when appropriate, synthetic-aperture reconstruction. The Synthetic Aperture Focusing Technique (SAFT) coherently combines measurements acquired at successive probe positions to improve spatial localization and provide a focused representation of the inspection region [10, 11, 12]. For mechanically scanned conventional UT, SAFT is the relevant reconstruction approach; it differs from FMC/TFM, which relies on array acquisition. The reconstructed intensity field remains an intermediate representation that must subsequently be interpreted in geometric terms.

The motivation for the present work arises from this separation between the measured ultrasonic response and the geometric quantities ultimately required for engineering assessment. For a fixed inspection configuration and a constrained family of flaws, it appears reasonable to ask whether the measurement-to-geometry relation can be estimated directly, while conventional displays and SAFT remain available for localization and expert review. Such a formulation shifts the computational target from an intermediate image toward a physically interpretable geometric representation.

Data-driven inverse modeling provides one route for examining this hypothesis. Supervised learning can approximate an inverse mapping from paired measurements and known geometries over the domain represented by the training data [13, 14, 15, 16, 17]. The principal practical limitation is the availability of reliable reference geometry. In experimental and especially in real-world field inspections, the number of encountered flaws is sparse and true subsurface contour is rarely known with sufficient spatial accuracy and volume for large-scale supervised learning. Physics-based simulation provides a controlled alternative by generating measurement-geometry pairs while making the validity domain of the resulting model explicit [18, 19].

Motivated by these considerations, the present study introduces a direct geometric reconstruction methodology for conventional single-element pulse-echo UT. CIVA simulations are used to generate ultrasonic inspection responses for randomly generated inner-surface-breaking flaw profiles, and a compact convolutional regressor is trained to estimate an ordered crack contour. The study assesses the feasibility of this formulation, referred to as UT2CRACK, using two engineering descriptors derived from the reconstructed contour, namely maximum depth and tilt, and examines the assumptions and limitations associated with the simulation-based results.

2. Background

2.1 Conventional ultrasonic imaging and geometric characterization

Conventional ultrasonic flaw sizing combines signal interpretation, reconstruction, and geometric inference. In mechanically scanned single-element inspection, SAFT synthesizes an aperture from successive probe positions and applies delay-and-sum processing to focus responses that are consistent with a specific spatial location [10, 11, 12]. This improves localization and can facilitate interpretation of crack-tip and corner responses. The acquisition architecture differs fundamentally from FMC/TFM, which uses an array and a full transmit-receive matrix [20, 21, 22].

In a time-domain formulation, SAFT may be viewed as a coherent back-projection of the recorded A-scans onto candidate points in the inspection region. For a candidate position r , the focused image value can be written schematically as shown in Equation (1) [10, 11]:

$$I(r) = |\sum_n w_n s_n(\tau_n(r))| \quad (1)$$

where $s_n(t)$ denotes the A-scan recorded at the n -th mechanical probe position, w_n is an optional aperture or amplitude weighting factor, and $\tau_n(r)$ is the theoretical pulse-echo time-of-flight between that probe position and the candidate point. In the simplest homogeneous direct pulse-echo case, the time-of-flight is given by Equation

(2), where $L_n(r)$ is the propagation distance and c is the relevant wave velocity. For an angle-beam inspection through a wedge, the time-of-flight calculation is instead obtained from the sum of the propagation times in the wedge and component, with the refracted path determined by the interface geometry and Snell's law. Samples that originate from a reflector at r are thereby shifted into temporal coincidence and add coherently, whereas unrelated contributions tend to add less constructively. [2, 4]

$$\tau_n(r) = 2L_n(r)/c \quad (2)$$

This formulation also clarifies what SAFT reconstructs. The quantity $I(r)$ is a focused measure of signal amplitude or coherence under an assumed propagation model; it is not itself the physical crack boundary. The quality of the reconstruction therefore depends on the correctness of the time-of-flight model, the effective aperture, bandwidth, beam geometry, and the strength of the relevant scattering features. For conventional mechanically scanned single-element UT, SAFT provides a physically transparent route from a set of spatially indexed A-scans to a spatial image, but quantitative crack dimensions still require a subsequent interpretation of that image or of individual diffracted and reflected responses.

Image formation does not remove the need for sizing. A focused pixel field represents signal coherence and amplitude under a chosen propagation model rather than a direct map of the physical crack boundary. Quantitative characterization therefore continues to rely on physically interpretable features such as time-of-flight, tip diffraction, back diffraction, amplitude behavior, and their relation to the inspection geometry [1, 5, 6, 7, 8]. These methods provide the principal conventional reference against which a learned geometric reconstruction should be interpreted.

2.2 Simulation-supported and data-driven inversion

Physics-based simulation is established in NDT for inspection design, virtual qualification, interpretation, and parametric studies. CIVA is a software platform for the simulation and processing of NDT data that combines semi-analytical propagation and scattering models with a CAD-oriented inspection environment [18]. Simulation is particularly useful for supervised learning because the measurement can be paired with an exact geometric reference, although the resulting model remains bounded by the assumptions and parameter space of the forward model.

The ultrasonic module of CIVA is designed for repeated forward calculations and therefore relies predominantly on semi-analytical approximations rather than full-domain time stepping. At the field level, the ultrasonic displacement field generated by the transducer at a point x and angular frequency ω can be represented schematically by an integral over the active transducer surface S_T , as summarized in Equation (3) [4, 18].

$$u(x, \omega) = \int_{S_T} G(x, \xi; \omega) q(\xi, \omega) dS_\xi \quad (3)$$

where q describes the source distribution on the transducer and G denotes the propagation kernel connecting a source point ξ to the field point x . In CIVA, this integral-field description is evaluated efficiently using semi-analytical beam-propagation procedures, including the pencil method, so that transmission and refraction through interfaces can be represented without explicitly discretizing the complete specimen volume. [18]

For flaw simulation, the forward problem can be expressed at an operator level as a sequence of propagation, interaction with the discontinuity, and reception. This sequence is summarized schematically in Equation (4) [16, 18].

$$U(\omega) = R\omega[SF(P\omega q)] \quad (4)$$

where $P\omega$ propagates the emitted field from the probe to the flaw, SF represents the scattering response associated with flaw geometry Γ , and $R\omega$ describes propagation back to and reception by the probe. This notation is schematic rather than a reproduction of the internal CIVA implementation, intended to make the role of the simulator in the present study explicit; the material, probe, wedge, and acquisition configuration define the forward operator, while the crack contour changes the scattering geometry. Repeated evaluation for different Γ therefore produces paired examples $\{U, \Gamma\}$, in which the reference geometry is known exactly. The

computational advantage of the semi-analytical formulation is important here because the training dataset requires many such evaluations, while the corresponding restriction is equally important: the learned inverse relation inherits the assumptions and validity range of the forward model. [18]

Recent machine learning-based ultrasonic NDT studies span image enhancement, detection and classification, synthetic-data generation, and inverse estimation [13, 23, 24, 14, 19, 25, 26, 27, 28]. The present study addresses the direct estimation of a compact flaw contour for a fixed conventional-UT configuration. In inverse-problem terms, the network acts as an amortized estimator over a constrained family of admissible geometries rather than as a general-purpose imaging system [16, 17, 29, 30, 31].

3. Physics-based simulation framework

3.1 Inspection model

The physics-based dataset was generated using the CIVA ultrasonic module [18] under a fixed conventional single-element pulse-echo inspection configuration. The simulated component was homogeneous steel with longitudinal-wave velocity 5900 m/s, shear-wave velocity 3230 m/s, and density approximately 7.8 g/cm³. A conventional single-element rectangular probe with a 20 × 22 mm crystal and 2 MHz center frequency was mounted on a 45° wedge. Signals were sampled at 70 MHz with 512 time samples. The flaw was represented as a faceted inner-surface-breaking profile described by 11 ordered contour points. Table 1 summarizes the simulation parameters.

The use of a fixed physical configuration isolates the geometric inverse problem. Material properties, probe characteristics, coupling geometry, sampling, and wave mode are held constant while flaw depth, tilt, and local profile vary. The resulting dataset therefore samples a constrained space of admissible flaw geometries rather than an unrestricted collection of image patterns. *Figure 1* shows the CIVA inspection geometry used in the simulations, including the wedge-mounted single-element probe, the steel component, and the inner-surface-breaking faceted flaw.

Holding the inspection parameters fixed allows the study to test the focused hypothesis that differences in the rendered ultrasonic response contain sufficient information to distinguish the represented contour geometries. It also defines the scope of the conclusions; the reported errors characterize interpolation within one virtual inspection procedure rather than universal crack reconstruction.

CIVA was selected as the means of conducting the simulations, as the implemented semi-analytical modeling offers a practical compromise between physical fidelity and computational cost for parametric studies [18]. While a fully resolved finite-element simulation could represent complex wave phenomena in detail, generating more than one thousand inspection cases at ultrasonic frequencies may be computationally prohibitive. Semi-analytical propagation and scattering models make systematic geometry variation feasible while preserving the principal relationships among probe, wedge, component, wave path, and flaw. The resulting data should nevertheless be regarded as outputs of a defined model, not as a complete substitute for experiments.

The model geometry further defines the interpretation of the 11-point target. The points describe a two-dimensional sectional profile of a faceted inner-surface-breaking flaw. Accordingly, the inverse task is sectional contour reconstruction under the assumptions of the simulation. Extension to finite-length three-dimensional cracks is intentionally excluded for the sake of keeping the study focused, as it would require either multiple sectional outputs, a surface representation, or a volumetric parameterization and would substantially increase both the forward-model parameter space and the amount of training data required.

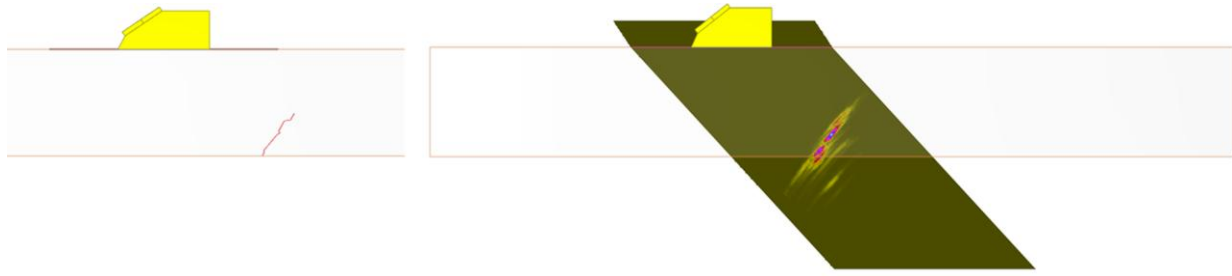


Figure 1. CIVA model of the simulated ultrasonic inspection configuration. A 2 MHz single-element probe is mounted on a 45° wedge and coupled to the outer surface of the steel component. The faceted, predominantly vertical flaw originates at the inner surface (back wall). Left: side view of the component and flaw. Right: simulated ultrasonic interaction with the flaw.

Table 1. Principal parameters of the simulated ultrasonic inspection.

Parameter	Value
Simulation environment	CIVA ultrasonic module
Analysis dimensionality	Two-dimensional
Component material	Steel
Longitudinal-wave velocity	5900 m/s
Shear-wave velocity	3230 m/s
Density	7.8 g/cm ³
Probe type	Single-element rectangular transducer
Active aperture	20 mm x 22 mm
Nominal center frequency	2 MHz
Relative bandwidth	50%
Wedge / refracted mode	45-degree shear wave
Sampling frequency	70 MHz
Samples per response	512
Model input representation	256 x 256 x 1 single-channel B-scan representation
Reference flaw representation	11 ordered (x, z) contour points
Number of paired cases	1,096
Target normalization	Coordinates divided by 50 mm wall thickness
Documented baseline partition	767 fitting cases / 329 held-out cases

3.2 Paired response and geometry data

Each simulated case produced a paired input-target record. The complete dataset contained 1,096 cases. The model input consisted of a single-channel B-scan representation of the simulated ultrasonic response, while the target was the exact polygon used to define the flaw in CIVA. The original B-scan data were resized to a common spatial dimension of 256 x 256 pixels before model input. This standardization provides a fixed tensor

geometry for all cases, avoids variation in input dimensionality across the dataset, and permits consistent convolution and pooling operations during training and inference. The 256 x 256 resolution was selected as a practical compromise between retention of the spatial signal structure relevant to flaw characterization and computational efficiency of the convolutional model.

The flaw contour is described by 11 ordered points. Point ordering is retained consistently across the dataset so that coordinate index i refers to the same progression along the profile for every sample, which is essential for coordinate-wise supervised regression. The use of simulation removes ambiguity in the target. Representative examples of the standardized B-scan inputs and corresponding reference flaw contours are shown in Figure 2.



Figure 2. Representative simulated inspection responses and associated reference flaw contours. Each contour is represented by 11 ordered points, corresponding to a 22-coordinate regression target.

The fixed number of points can be interpreted as a geometric sampling of the admissible profile. Eleven points provide more information than a scalar depth-and-angle description, because local changes in slope and curvature can be represented, while keeping the regression target small enough for a compact model. The choice is therefore a compromise between geometric detail and statistical efficiency. A denser point sequence would increase output dimensionality and could impose high-frequency point-to-point error without adding significant engineering value; a much smaller sequence would force distinct faceted profiles into the same coarse parameterization.

Coordinate regression assumes that the order of the predicted points corresponds to the order of the reference points. Without consistent correspondence, identical contours could be represented by multiple target vectors, making the supervised loss artificially large. This direct point regression appears appropriate for a constrained shape family. For arbitrary cracks, other representations may be more suitable.

The resulting $256 \times 256 \times 1$ tensor therefore represents one standardized single-channel B-scan per simulated inspection case.

4. Direct geometric inverse formulation

4.1 Forward and inverse mappings

Let Γ denote an admissible flaw contour and F the forward ultrasonic model defined by the inspection configuration. The simulated response U is related to the contour according to Equation (5) [16, 17].

$$U = F(\Gamma) \quad (5)$$

The characterization problem seeks an approximation of the inverse relation. A parameterized model $f\theta$ maps the inspection representation to an estimated contour according to Equation (6) [16, 17, 29].

$$\hat{\Gamma} = f\theta(U) \approx F^{-1}(U) \quad (6)$$

The inverse may not be globally unique for arbitrary inspections. In this study, however, the model is trained and evaluated only on contours belonging to the simulated family M and under a fixed inspection configuration. The practical problem is therefore to approximate the inverse mapping over this restricted domain.

Ill-posedness can arise in several ways. Different contours may produce responses that are similar within the finite bandwidth and aperture of the probe; small waveform perturbations may cause comparatively large changes in an inferred local boundary; and parts of a crack may be weakly observable because their scattering direction does not return energy to the probe. In classical inverse problems, these difficulties are handled through regularization, prior models, multiple views, or constraints on the admissible solution [16]. In the present approach, much of the regularization is implicit in the training distribution and in the low-dimensional ordered-contour representation.

The learned model can be viewed as an amortized inverse. A conventional optimization-based solution might search over candidate contours for each new measurement, repeatedly evaluating or approximating the forward model until a data-misfit objective is minimized. Amortized inversion moves computational effort offline; the network parameters are adjusted across many paired simulations so that a new input can be mapped to a contour in one forward pass. This can make online inference fast, but it also means that the prior encoded by the training set is inseparable from the solution. Geometries outside the represented family may be projected onto a plausible but incorrect in-domain contour.

This interpretation separates the scientific problem from the numerical solver. The convolutional network described in Section 5 is one approximator of F^{-1} over M ; it is not itself the definition of the methodology.

4.2 Geometric representation

The contour is represented as $\Gamma = \{(x_i, z_i)\}$ for $i = 1, \dots, 11$. The corresponding target vector is $y = [x_1, z_1, \dots, x_{11}, z_{11}]^T$. This ordered polygon was selected for the following reasons: it is directly available from the simulation geometry; it is compact; it preserves physically interpretable coordinates; it avoids the threshold dependence of pixel masks; and it can be converted directly into engineering quantities. The representation also regularizes the inverse problem by limiting the output to the geometric degrees of freedom represented in the dataset.

The ordered contour provides an intermediate representation between scalar sizing and dense segmentation: it retains local profile morphology while allowing engineering quantities to be derived directly.

Two derived quantities are used for model selection and test evaluation. Following established ultrasonic flaw-sizing practice, maximum flaw depth is used as a physically interpretable sizing quantity; in the present study it is operationally defined from the largest z-coordinate of the reconstructed contour as given in Equation (7) [1, 5].

$$d(\hat{\Gamma}) = \max_i(\hat{z}_i) \quad (7)$$

The tilt angle is calculated from the line joining the first and last ordered contour points. With x denoting the lateral coordinate and z the depth coordinate, the signed angle relative to the vertical axis is defined by Equation (8). The sign indicates the direction of inclination. The same definition is applied to the reference and reconstructed contours. Maximum depth and tilt summarize complementary aspects of reconstruction; the first is sensitive to a local extremum, whereas the second describes the global orientation of the profile. [1, 5]

$$\alpha = \arctan[(x_{11} - x_1)/(z_{11} - z_1)] \quad (8)$$

5. Numerical inverse model

5.1 Preprocessing

The original B-scan data were resized to 256 x 256 pixels before model input to provide a uniform spatial representation for all inspection cases. This standardization ensures identical tensor dimensions throughout the dataset and enables consistent convolution and pooling operations during training and inference. The selected resolution provides a practical balance between preservation of the spatial signal structure relevant to flaw characterization and computational efficiency of the convolutional model. The resized single-channel B-scans were then converted to floating-point values and normalized to the interval [0, 1]. The 22 target coordinates were divided by the 50 mm wall thickness for training and rescaled to millimeters for evaluation. The assembled dataset contained 1,096 paired cases. In the documented baseline implementation, the first 767 cases (70.0%) were assigned to model fitting, and the remaining 329 cases (30.0%) were held out. The reporting workflow subsequently used stored index sets so that architecture selection and final evaluation were performed on fixed subsets rather than newly sampled partitions.

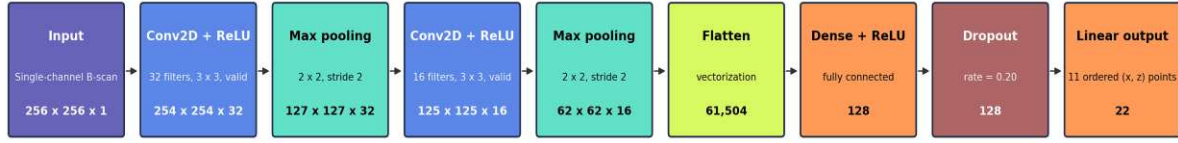
The implementation used a 70/30 partition for model fitting and held-out evaluation. The simulated flaw geometries were generated randomly rather than as an ordered parameter sweep, and file order therefore did not encode a progression of neighboring geometries.

5.2 Compact convolutional regressor

The convolutional regressor was implemented using TensorFlow with the Keras high-level API. The selected inverse model receives a 256 x 256 x 1 normalized single-channel B-scan and returns 22 linear coordinate values. It contains a first convolutional layer with 32 filters and 3 x 3 valid kernels, rectified-linear activation, and 2 x 2 max pooling; a second convolutional layer with 16 filters and the same kernel, activation, and pooling settings; a flattening operation; a fully connected layer with 128 rectified-linear units; a dropout layer with a rate of 0.20; and a linear 22-unit output layer. The output vector is separated into 11 lateral coordinates and 11 depth coordinates to reconstruct the ordered contour.

The architecture, displayed in Figure 3, is intentionally compact relative to common image-segmentation networks. This choice follows from the target representation: the model is not required to reconstruct a dense image, but to estimate a 22-dimensional geometric state within a constrained solution space.

The convolutional layers provide local receptive fields and shared weights, which are suitable for identifying spatially recurring patterns in the inspection representation. The first block extracts low-level gradients, echo boundaries, and local texture-like features; pooling reduces spatial resolution and introduces limited tolerance to small translations. The second block combines these local responses into more abstract patterns before the feature maps are flattened for global coordinate regression. Because the output describes a single contour rather than a dense image, an encoder-only structure is sufficient; a decoder such as that used in U-Net [33] would add parameters for spatial upsampling that are not required by the target representation.

Compact convolutional regressor for direct contour reconstruction

All convolution and pooling operations use valid padding. The 22 outputs encode 11 ordered contour points.

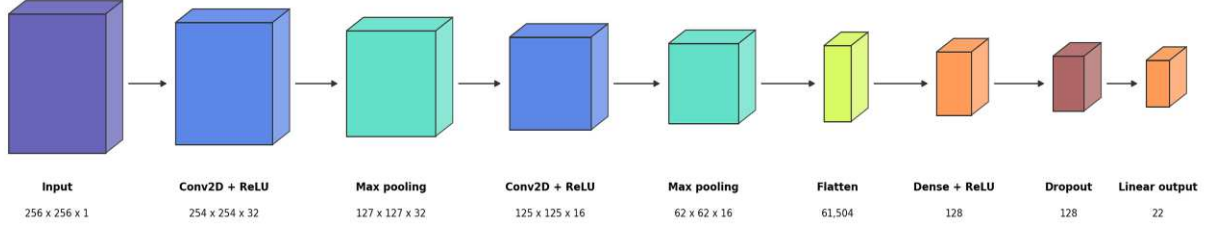
Layered network representation

Figure 3. Selected compact convolutional regression architecture mapping a $256 \times 256 \times 1$ single-channel ultrasonic B-scan representation to 11 ordered contour points.

5.3 Training and loss function

Model parameters were optimized with Adam using the learning rate of 1×10^{-3} . Training minimized mean squared error (MSE) over the normalized coordinate vector, while mean absolute error (MAE) was monitored. The documented training runs used a batch size of 32 and 100 epochs. No early-stopping callback was applied; instead, model checkpoints were written after each epoch while monitoring validation mean absolute error. All hyperparameters were optimized under the same numerical settings listed in Table 2 to preserve comparability.

Coordinate MSE assigns equal nominal importance to every normalized output coordinate. This encourages overall point-wise agreement but does not directly optimize the derived quantities used for engineering interpretation. The grid search partly addresses this mismatch by selecting capacity using depth and tilt validation errors, but the training objective itself remains purely coordinate based.

Coordinate-wise MSE is consistent with the continuous output representation, but it does not explicitly enforce contour smoothness, point ordering, or physical feasibility. These properties are learned from the training distribution rather than imposed by the loss function.

Table 2. Training and optimization settings documented in the implementation.

Training parameter	Documented setting
Optimizer	Adam
Initial learning rate	1×10^{-3} (Keras default)
Loss	Mean squared error on normalized coordinates
Monitored metric	Mean absolute error
Batch size	32
Epochs	100
Training-set shuffling	Enabled by Keras during model.fit
Early stopping	Not used
Checkpointing	Model written after every epoch; validation MAE monitored

6. Hyperparameter optimization

A structured grid search was performed over the number of filters in the two convolutional layers, convolutional kernel size, and dense-layer width. All configurations were evaluated on the same validation subset. Rather than selecting a model using coordinate loss alone, model capacity was assessed using the mean absolute errors of two derived physical quantities: maximum depth and tilt angle.

Model selection used Pareto dominance. A configuration was considered non-dominated if no alternative achieved an equal or lower validation error in both physical metrics and a strictly lower error in at least one. This criterion prevents an architecture that improves one engineering quantity at the expense of a substantial degradation in the other from being selected by an arbitrary scalar weighting.

Within the explored search space, the configuration with 32 and 16 filters in the first and second convolutional layers, respectively, a 3×3 kernel, and 128 dense units formed the unique non-dominated solution. The result indicates that tested larger models did not provide a simultaneous improvement in depth and tilt validation error under the applied training conditions. It should not be interpreted as proof of global architectural optimality outside the examined grid. The validation trade-off across all tested architectures is shown in Figure 4, while the corresponding aggregated depth and tilt error maps are shown in Figures 5 and 6, respectively.

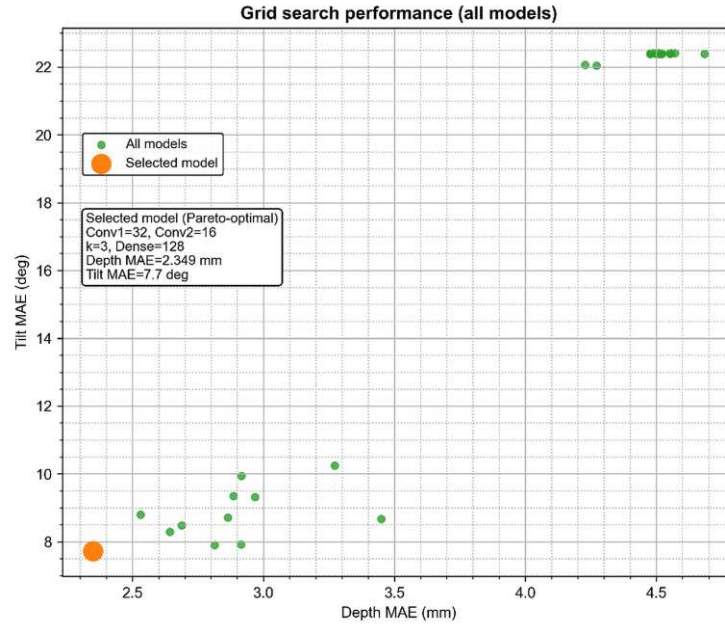


Figure 4. Validation depth MAE versus validation tilt MAE for all evaluated architectures. The selected model is the unique non-dominated configuration within the explored grid.

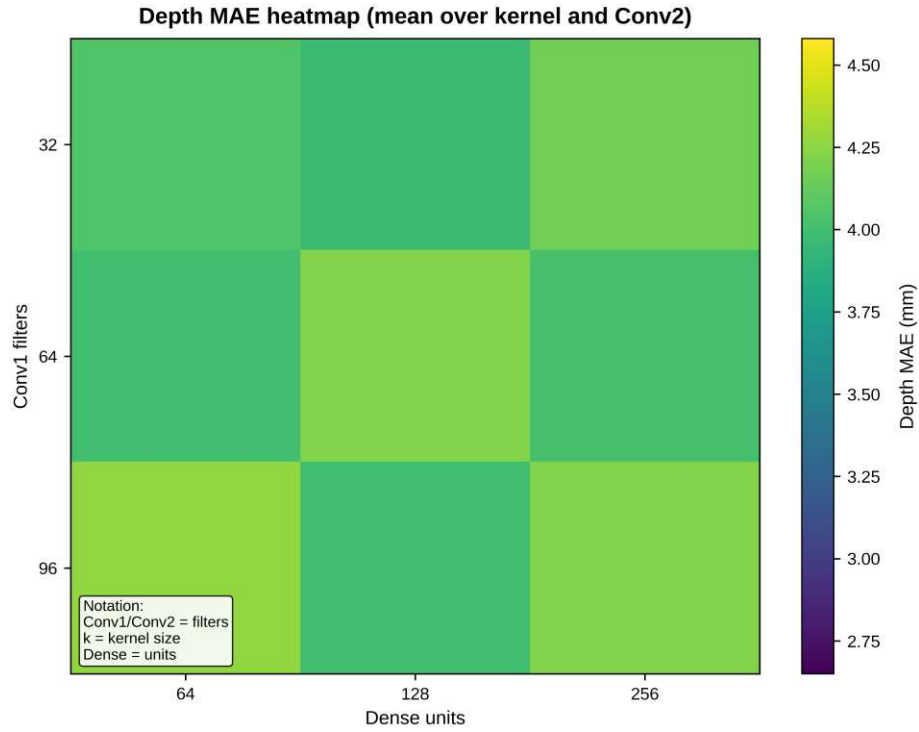


Figure 5. Mean validation depth MAE across selected architectural dimensions. Values are averaged over the remaining grid-search variables.

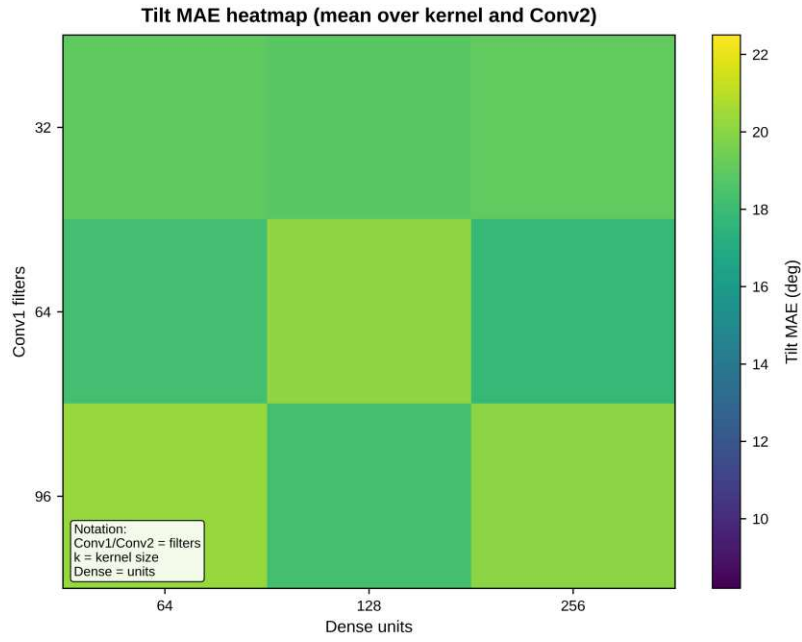


Figure 6. Mean validation tilt MAE across selected architectural dimensions. Values are averaged over the remaining grid-search variables.

7. Results

7.1 Test-set reconstruction performance

After architecture selection, the chosen model was evaluated on the fixed held-out test set. Maximum depth was reconstructed with a mean absolute error of approximately 2.10 mm and a coefficient of determination R^2 of approximately 0.716. Tilt angle was reconstructed with a mean absolute error of approximately 6.30 degrees and R^2 of approximately 0.886.

The higher coefficient of determination for tilt indicates that global profile orientation was recovered more consistently than the exact extremal depth. This difference is physically and numerically plausible. Tilt is informed by the coordinated positions of multiple contour points, whereas maximum depth depends on a single extremal location and is therefore more sensitive to localized vertical deviations.

Table 3. Performance on the fixed held-out test set.

Derived quantity	MAE	R^2
Maximum depth	≈ 2.10 mm	≈ 0.716
Tilt angle	$\approx 6.30^\circ$	≈ 0.886

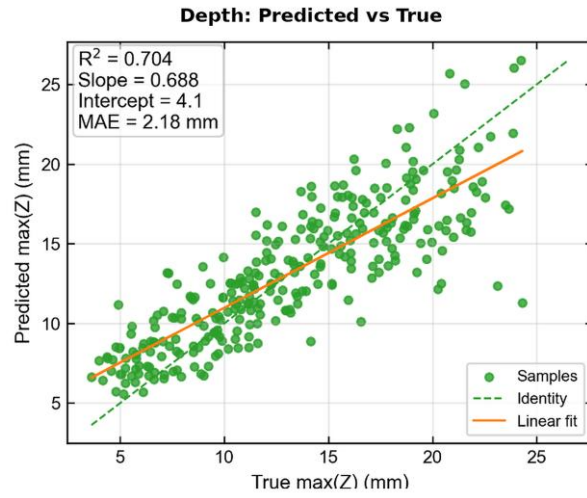


Figure 7. Reconstructed versus reference maximum flaw depth for the diagnostic evaluation subset.

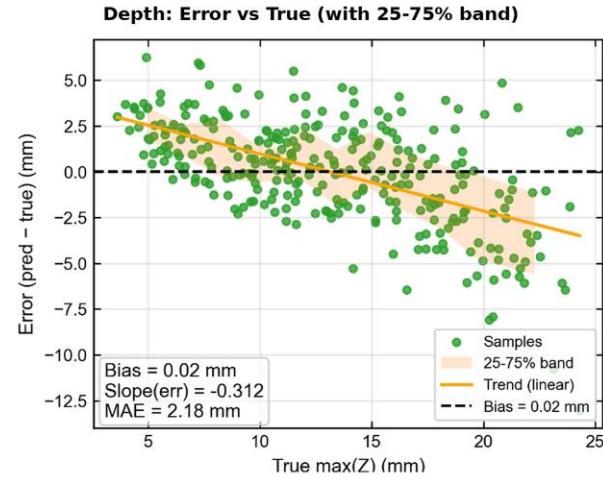


Figure 8. Signed maximum-depth error as a function of reference depth; the shaded region denotes the interquartile band.

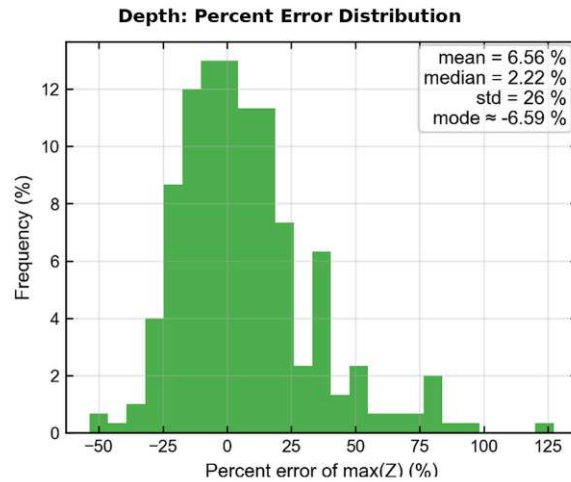


Figure 9. Distribution of relative maximum depth error for the diagnostic evaluation subset.

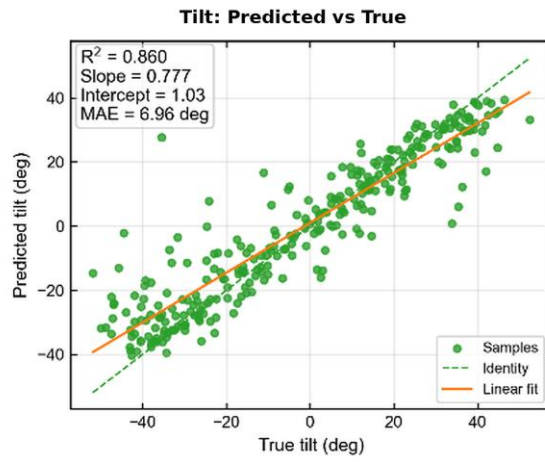


Figure 10. Reconstructed versus reference flaw tilt for the diagnostic evaluation subset.

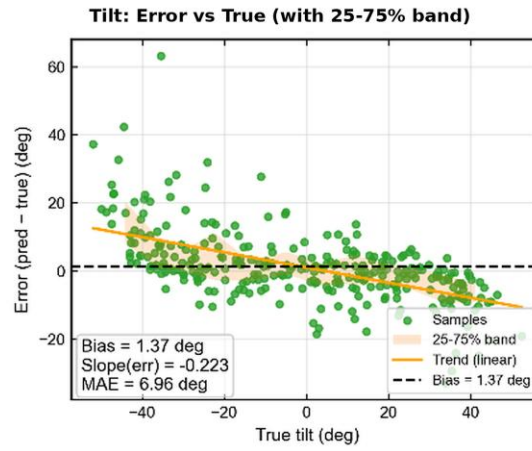


Figure 11. Signed tilt error as a function of reference tilt; the shaded region denotes the interquartile band.

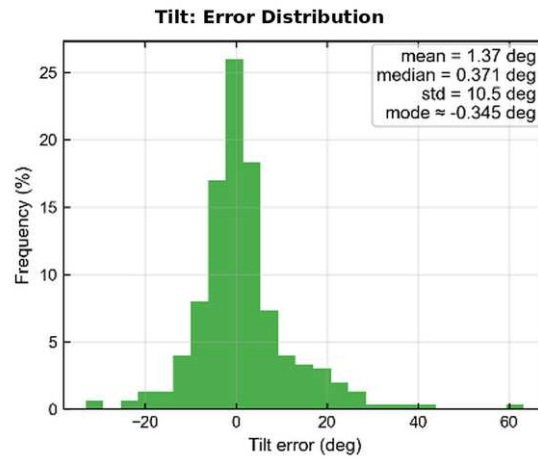


Figure 12. Distribution of signed tilt error for the diagnostic evaluation subset.

7.2 Error diagnostics

The diagnostic plots in Figures 7–12 complement the fixed held-out test-set summary in Table 3. Figures 7–9 show the depth behavior: Figure 7 presents reconstructed versus reference maximum depth, Figure 8 resolves the signed residual across the represented depth range, and Figure 9 summarizes the relative-error distribution. The plots confirm a clear depth-dependent signal while also showing greater dispersion than is observed for tilt.

Tilt diagnostics are shown in Figures 10–12. The parity relation in Figure 10 is comparatively tight, while Figure 11 shows the signed error across the angular range and Figure 12 the corresponding distribution. This error structure is consistent with the stronger held-out coefficient of determination for tilt reported in Table 3.

7.3 Qualitative contour reconstruction

Representative test cases show that the reconstructed ordered points generally preserve the position and overall direction of the reference contour. Deviations tend to appear as local vertical offsets or endpoint discrepancies rather than complete geometric misalignment. This visual behavior is consistent with the quantitative finding that tilt is more robustly recovered than maximum depth. Representative flaw reconstructions are shown in Figure 13.

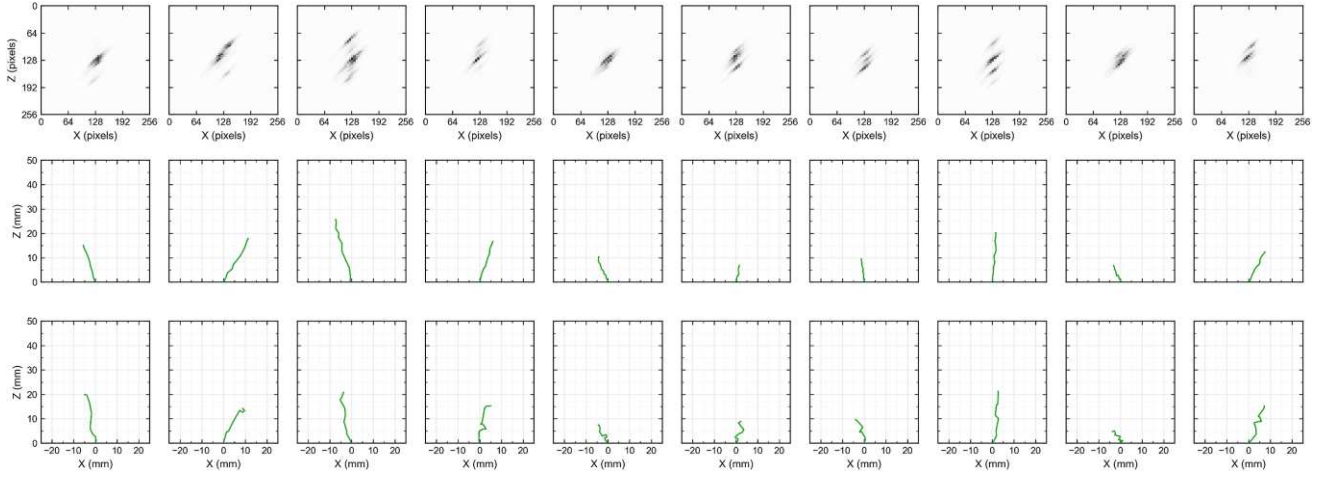


Figure 13. Representative test-set reconstructions showing the input inspection response, reconstructed contour, and reference contour.

8. Discussion

The results support the feasibility of estimating quantitative flaw geometry directly from a conventional ultrasonic inspection representation under controlled simulated conditions. The distinction from conventional reconstruction is primarily one of target definition. SAFT produces a spatial intensity representation that remains available for operator interpretation, whereas UT2CRACK estimates a compact contour from which geometric quantities are obtained directly. These approaches address different stages of the interpretation chain and are therefore complementary rather than mutually exclusive.

For practical implementation, the conventional ultrasonic data and, where required, a SAFT reconstruction would remain traceable inspection outputs. Direct inversion could provide an additional quantitative estimate. Agreement between the learned contour and conventional interpretation would provide complementary evidence, whereas substantial disagreement would identify cases requiring further examination. A similar relationship applies to model-based inversion: a learned contour may provide a rapid initial estimate, while forward-model consistency or iterative refinement can be introduced when additional physical verification is required.

The ordered contour establishes a consistent representation across simulation, supervision, and evaluation. This avoids intermediate conversion from image to mask and from mask to contour, while preserving a direct geometric interpretation. The representation is well-matched to the constrained family of faceted profiles considered here, but it relies on point correspondence: each predicted coordinate is expected to represent the

same relative location along the flaw profile as its reference counterpart. This assumption simplifies supervised regression but limits direct extension to branching, disconnected, or variable-complexity flaws.

The stronger recovery of tilt than of maximum depth is consistent with the geometry of the two descriptors. Tilt reflects the global orientation of the reconstructed profile and is informed by coordinated changes across several contour points. Maximum depth is governed by a local extremum and is therefore more sensitive to a localized coordinate error. This distinction also explains why full-contour evaluation should accompany scalar metrics in subsequent studies.

The compact network is best interpreted in relation to the low-dimensional output. The task is not dense image synthesis but estimation of 22 coordinates within a restricted solution space. The grid-search result indicates that additional capacity did not improve both validation quantities simultaneously within the explored range. The coordinate MSE used for training is intentionally simple; future work may examine additional geometric terms, for example penalties on endpoint orientation or local extrema, provided that improvements in a selected scalar quantity do not degrade the reconstructed contour as a whole.

The principal limitation is the use of simulated data for both training and evaluation. The CIVA model provides exact geometric reference data and controlled variation, but it does not reproduce the full variability of experimental inspection, including coupling changes, surface condition, probe positioning, electronic noise, attenuation uncertainty, and material microstructure. The reported performance should therefore be interpreted as in-domain reconstruction accuracy for the modeled inspection procedure rather than as evidence of field performance [14, 19].

The method is also bounded by the represented inspection configuration. Changes in probe frequency, refracted beam angle, wedge geometry, material properties, or component geometry alter the forward response and would require corresponding coverage during training or explicit conditioning of the inverse model. In addition, the fixed 11-point representation presumes a consistent contour topology and point ordering. These restrictions are appropriate for the present proof-of-concept study but define the conditions under which its results can be interpreted.

The current evaluation emphasizes maximum depth and tilt. These quantities are physically interpretable but do not fully characterize local contour agreement. Subsequent validation should include contour-level distance measures, endpoint and length errors, uncertainty intervals, and performance as a function of flaw geometry and signal quality. Because a deterministic estimator returns a contour for every input, uncertainty estimation or domain-validity checks will also be important for experimental use [35, 36, 37, 38].

As a next step, we suggest experimental validation using conventional single-element inspections acquired under a configuration corresponding as closely as possible to the simulated setup. Initial measurements should establish agreement in time base, bandwidth, propagation path, and background response before the inverse model is assessed on specimens with independently known flaw geometry. Simulation-to-experiment adaptation can then be introduced in a controlled manner, for example by varying physically meaningful simulation parameters or combining simulated flaw responses with measured backgrounds [19].

A direct comparison with a conventional SAFT-based sizing workflow should accompany this validation. Such a comparison should address accuracy, processing time, operator interaction, robustness to unsupported cases, and communication of uncertainty. The purpose is not to replace an established reconstruction method with a learned model, but to determine whether direct geometric inversion provides useful additional quantitative information for a defined inspection task.

9. Conclusions

This study examined whether quantitative flaw characterization in conventional single-element pulse-echo UT can be formulated as direct geometric inversion. Physics-based CIVA simulations supplied paired ultrasonic responses and exact ordered flaw contours, and a compact convolutional regressor was used to estimate the contour from the inspection representation. Model selection was based on validation errors in maximum depth and tilt rather than on coordinate loss alone.

Evaluation on held-out simulated cases showed meaningful agreement in both depth and orientation, with tilt recovered more consistently than the local depth extremum. These findings establish feasibility within the modeled inspection domain while leaving experimental transfer, uncertainty, and broader geometric generalization for subsequent validation.

The principal implication is methodological; for a defined conventional-UT task and a constrained flaw family, the reconstruction target can be aligned directly with flaw geometry rather than with an intermediate intensity image. This formulation provides a basis for evaluating learned geometric inversion alongside established signal interpretation and SAFT-based reconstruction in future experimental studies.

Declarations

CRedit authorship contribution statement: Daniel Algernon: Conceptualization, Methodology, Software, Investigation, Data curation, Formal analysis, Visualization, Writing - original draft, Writing - review & editing.

Funding: Parts of this work were conducted within the framework of the Program for Investigation of NDE by International Collaboration (PIONIC), with support from the Swiss Federal Nuclear Safety Inspectorate (ENSI).

References

- [1] M. Felice and Z. Fan, "Sizing of flaws using ultrasonic bulk wave testing: A review," *Ultrasonics*, vol. 88, pp. 26-42, 2018.
- [2] J. Krautkrämer and H. Krautkrämer, *Ultrasonic Testing of Materials*, 1990.
- [3] J. Blitz and G. Simpson, *Ultrasonic Methods of Non-Destructive Testing*, 1996.
- [4] L. Schmerr Jr. and S.-J. Song, *Ultrasonic Nondestructive Evaluation Systems: Models and Measurements*, 2007.
- [5] P. Doyle and C. Scala, "Crack depth measurement by ultrasonics: A review," *Ultrasonics*, vol. 16, pp. 164-170, 1978.
- [6] M. Silk, "The transfer of ultrasonic energy in the diffraction technique for crack sizing," *Ultrasonics*, vol. 17, pp. 113-121, 1979.
- [7] B. Shakibi, F. Honarvar, M. Moles, J. Caldwell and R. Sinclair, "Resolution enhancement of ultrasonic defect signals for crack sizing," *NDT & E International*, vol. 52, pp. 37-50, 2012.
- [8] P. Broberg, A. Runnemalm and M. Sjö Dahl, "Improved corner detection by ultrasonic testing using phase analysis," *Ultrasonics*, vol. 53, pp. 630-634, 2013.
- [9] DIN, "DIN EN ISO 5577:2026-02: Non-destructive testing - Ultrasonic testing - Vocabulary (ISO 5577:2025); German version EN ISO 5577:2025," DIN Media GmbH, Berlin, 2026.
- [10] J. Seydel, "SAFT - The evolution of a signal processing technology for ultrasonic testing," *NDT International*, vol. 19, pp. 163-167, 1986.
- [11] V. Schmitz, W. Müller and G. Schäfer, "Synthetic aperture focusing technique signal processing," *NDT International*, vol. 19, pp. 177-189, 1986.
- [12] V. Schmitz, S. Chakhlov and W. Müller, "Experiences with synthetic aperture focusing technique in the field," *Ultrasonics*, vol. 38, pp. 731-738, 2000.
- [13] S. Cantero-Chinchilla, P. Wilcox and A. Croxford, "Deep learning in automated ultrasonic NDE - Developments, axioms and opportunities," *NDT & E International*, vol. 131, p. 102703, 2022.
- [14] S. Uhlig, I. Alkhasli, F. Schubert, C. Tschöpe and M. Wolff, "A review of synthetic and augmented training data for machine learning in ultrasonic non-destructive evaluation," *Ultrasonics*, vol. 134, p. 107041, 2023.
- [15] "A neural network for crack sizing trained by finite element calculations," *NDT & E International*, vol. 29, pp. 147-155, 1996.
- [16] A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation*, 2005.
- [17] M. Bertero and P. Boccacci, *Introduction to Inverse Problems in Imaging*, 1998.
- [18] P. Calmon, S. Mahaut, S. Chatillon and R. Raillon, "CIVA: An expertise platform for simulation and processing NDT data," *Ultrasonics*, vol. 44, pp. e975-e979, 2006.
- [19] S. McKnight, S. Pierce, E. Mohseni, C. MacKinnon, C. MacLeod, T. O'Hare and C. Loukas, "A comparison of methods for generating synthetic training data for domain adaptation of deep learning models in ultrasonic non-destructive evaluation," *NDT & E International*, vol. 141, p. 102978, 2024.
- [20] B. Drinkwater and P. Wilcox, "Ultrasonic arrays for non-destructive evaluation: A review," *NDT & E International*, vol. 39, pp. 525-541, 2006.
- [21] C. Holmes, B. Drinkwater and P. Wilcox, "Post-processing of the full matrix of ultrasonic transmit-receive array data for non-destructive evaluation," *NDT & E International*, vol. 38, pp. 701-711, 2005.

- [22] J. Zhang, B. Drinkwater, P. Wilcox and A. Hunter, "Defect detection using ultrasonic arrays: The multi-mode total focusing method," *NDT & E International*, vol. 43, pp. 123-133, 2010.
- [23] S. Cantero-Chinchilla, P. Wilcox and A. Croxford, "A deep learning based methodology for artefact identification and suppression with application to ultrasonic images," *NDT & E International*, vol. 126, p. 102575, 2022.
- [24] G. Granados, R. Miorelli, F. Gatti, S. Robert and D. Clouteau, "Towards a multi-fidelity deep learning framework for a fast and realistic generation of ultrasonic multi-modal Total Focusing Method images in complex geometries," *NDT & E International*, vol. 139, p. 102906, 2023.
- [25] I. Virkkunen, T. Koskinen, O. Jessen-Juhler and J. Rinta-aho, "Augmented ultrasonic data for machine learning," *Journal of Nondestructive Evaluation*, vol. 40, p. 4, 2021.
- [26] T. Koskinen, I. Virkkunen, O. Siljama and O. Jessen-Juhler, "The effect of different flaw data to machine learning powered ultrasonic inspection," *Journal of Nondestructive Evaluation*, vol. 40, p. 24, 2021.
- [27] O. Siljama, T. Koskinen, O. Jessen-Juhler and I. Virkkunen, "Automated flaw detection in multi-channel phased array ultrasonic data using machine learning," *Journal of Nondestructive Evaluation*, vol. 40, p. 67, 2021.
- [28] L. Silva, E. Simas Filho, M. Albuquerque, I. Silva and C. Farias, "Segmented analysis of time-of-flight diffraction ultrasound for flaw detection in welded steel plates using extreme learning machines," *Ultrasonics*, vol. 102, p. 106057, 2020.
- [29] H. Engl, M. Hanke and A. Neubauer, *Regularization of Inverse Problems*, 1996.
- [30] P. Hansen, *Discrete Inverse Problems: Insight and Algorithms*, 2010.
- [31] J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*, 2005.
- [32] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mane, R. Monga, S. Moore, D. Murray and C. Olah, "TensorFlow: A System for Large-Scale Machine Learning," *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, p. 265–283., 2016.
- [33] O. Ronneberger, P. Fischer and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention*, 2015.
- [34] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in *3rd International Conference for Learning Representations*, San Diego, 2015.
- [35] A. Kendall and Y. Gal, "What uncertainties do we need in Bayesian deep learning for computer vision?," in *Advances in Neural Information Processing Systems 30*, 2017.
- [36] J. Gawlikowski, C. Tassi, M. Ali, J. Lee, M. Humt, J. Feng, A. Kruspe, R. Triebel, P. Jung and R. Roscher, "A survey of uncertainty in deep neural networks," *Artificial Intelligence Review*, vol. 56, pp. 1513-1589, 2023.
- [37] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in *Proceedings of the 33rd International Conference on Machine Learning*, 2016.
- [38] B. Lakshminarayanan, A. Pritzel and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," in *Advances in Neural Information Processing Systems 30*, 2017.