China Instrument and Control Society
Instrumentation
OPEN ACCESS
INSTRUMENTATION, Vol. 13, No. 3, September 2026 https://doi.org/10.15878/j.instr.202600402

Pipeline Inspection

Multi-View Information Fusion-Based Methods for Pipeline Inspection and Defect Quantification

Tianxi Wen[1], Quanwei Yu[3], Meiling Li[1], Bin Gao[1*], Gaige Ru[1], Chao Wang[2], Fei Luo[2], Zugao Chen[2]

[1] School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China

[2] Deyuan Technology Co., Ltd, Chengdu, Sichuan 610041, China

[3] Sichuan Natural Gas Sales Center, Natural Gas Branch of China Petroleum & Chemical Corporation (Sinopec Corp.), Chengdu, Sichuan 610011, China.

[*] Corresponding author email: bin_gao@uestc.edu.cn, ORCID ID: 0000-0001-9993-1013

Abstract: This paper addresses multiclass target detection and defect-depth quantification in magnetic flux leakage (MFL) in-line inspection. Structural responses and operating disturbances can mask weak defects and destabilize depth estimation. For detection, each mileage-aligned axial MFL segment is encoded as complementary time-intensity (TI) and grayscale (GR) views. Spatial-frequency enhancement strengthens weak-response cues, while distance-aware cross-view fusion limits interference from unrelated strong responses. Detected defect candidates are then mapped to the corresponding triaxial signals for depth quantification; engineering response priors and reliability-weighted probabilistic fusion combine complementary axis evidence and represent predictive uncertainty. On a 920-sample field test set from two in-service pipelines, the detector achieved 65.3% mAP50 and 48.1% mAP50:95, compared with 64.7% and 46.9% for an adapted Fusion-Mamba baseline. In ten-fold cross-validation on 604 artificial-defect samples, the depth model achieved a mean absolute error (MAE) of 0.240 mm and R^2 of 0.981, versus 0.285 mm MAE for the MLKG-NN comparator. Weak, small, and boundary-ambiguous targets remained challenging in field sections with low signal-to-noise ratio.

Keywords: magnetic flux leakage; defect detection; paired-view feature fusion; physics-informed modeling; defect quantification; uncertainty

Copyright: © 2026 by the authors. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Citation: Tianxi Wen, Quanwei Yu, Meiling Li, Bin Gao, Gaige Ru, Chao Wang, Fei Luo, Zugao Chen. "Multi-View Information Fusion-Based Methods for Pipeline Inspection and Defect Quantification." Instrumentation 13, no.3 (September 2026). https://doi.org/10.15878/j.instr.202600402

1 Introduction

Pipeline transportation plays a crucial role in national infrastructure networks owing to its large-scale continuous transmission capacity, high operational safety, and significant economic advantages. Regular inspection and maintenance are therefore essential to ensure reliable operation. Magnetic flux leakage (MFL) inspection has become one of the most widely used techniques for ferromagnetic pipelines because of its high detection accuracy, stable performance, and high degree of automation[1]. MFL in-line inspection signals are typical spatial time-series signals. By magnetizing the pipe wall and measuring the leakage magnetic field induced by local discontinuities, MFL inspection provides defect-related information from multiple channels and multiple magnetic-field components. Accordingly, defect identification in MFL in-line inspection can be regarded as an anomaly detection problem over high-resolution multichannel time-series data.

However, in practical engineering applications, MFL inspection still faces several critical challenges in data

preprocessing, defect identification, and quantitative analysis. First, MFL signals are highly susceptible to multisource noise, including sensor thermal noise, pipe surface roughness, weld structures, remanence effects, and background magnetic-field disturbances. In data acquired under complex or harsh operating conditions, such noise may cause signal distortion, blurred boundaries, waveform aliasing, and loss of local features, thereby degrading data availability and detection stability[2]. More importantly, defect quantification in MFL inspection, including length, width, depth, and shape reconstruction, is essentially an ill-posed inverse problem. The relationship between MFL signals and defect geometry is governed by a complex nonlinear mapping and is affected by coupled factors such as permeability variation, hysteresis effects, lift-off variation, and material magnetization states, which substantially increase the complexity of defect quantification[3]. In this context, data fusion and multi-view representation provide a feasible way to improve information completeness and reduce the limitations of single-source observations in pipeline inspection scenarios[4,5].

1.1 Defect Detection in MFL Signals

Early studies on MFL-related pipeline anomaly and defect detection mainly relied on one-dimensional channel signal analysis, where abnormal responses were identified using statistical, frequency-domain, or time-frequency features[6-8]. For example, Afzal and Udpa employed adaptive filtering and wavelet denoising to suppress structural noise in MFL signals[7]. Mukhopadhyay and Srivastava used discrete wavelet transform to extract features such as amplitude, span, and boundary variation rate for characterizing metal-loss defects[8]. These methods improve signal interpretability; however, they depend heavily on handcrafted features and single-channel local responses, making it difficult to capture spatial morphology and multi-channel correlations in high-resolution MFL data.

To overcome these limitations, recent studies have increasingly shifted toward image-based deep detection methods. Chen and Yang transformed MFL signals into two-dimensional images and used deep convolutional neural networks (CNNs) for corrosion defect identification[2]. Chen et al. further combined YOLOv5 and a Vision Transformer (ViT) to automatically locate pipeline defects in MFL data, thereby improving detection efficiency and spatial localization capability[9]. Nevertheless, most existing methods are based on a single image perspective or unimodal representation. Under complex operating conditions, they remain vulnerable to weld interference, strong responses from structural components, noise disturbances, and the loss of local features associated with weak defects, which may lead to missed detections and false alarms. In broader pipeline monitoring tasks, multisource multimodal feature fusion has also been shown to improve the robustness of small-leak detection, further indicating the value of complementary information integration[4]. Related cross-modality detection studies provide additional references for feature-level fusion. Dong et al. proposed Fusion-Mamba for cross-modality object detection; its Fusion-Mamba Block (FMB) uses State Space Channel Swapping (SSCS) for shallow feature exchange and Dual State Space Fusion (DSSF) for deep hidden-state fusion and suppression of pseudo-target information[10]. Zhang et al. proposed TFDet (Target-Aware Fusion for RGB-T Pedestrian Detection), where RGB-T denotes paired red-green-blue and thermal imagery; its target-aware fusion strategy combines a feature fusion module and a feature refinement module with a correlation-maximum loss to collect complementary features, distinguish target from background, and reduce false-positive responses[11]. For MFL-specific multi-view learning, Shen, Liu, and co-workers constructed multi-view images from axial and radial MFL signals and introduced a semisupervised circular teacher network (SSCT-Net) to exploit cross-view complementarity for defect detection with limited labeled samples[12]. Although these studies demonstrate the value of cross-modality and multi-view fusion, how to integrate complementary information across paired representations of the same axial MFL signal while limiting interference from unrelated strong responses remains a key challenge.

1.2 Defect Quantification in MFL Signals

MFL defect quantification aims to infer defect geometric parameters from inspection signals and is therefore an inverse problem. To avoid iterative solutions in complex physical inversion, data-driven methods have been widely applied to defect size estimation. Piao et al. used a standard CNN to classify and characterize coexisting defects from MFL data, demonstrating the feasibility of learning defect-related representations directly from inspection signals[13]. Lu et al. proposed a visual transformation convolutional neural network (VT-CNN) to estimate defect dimensions by spatially remapping MFL images[3]. Zhang et al. developed the Feature-Modeling-Decoupling Framework (FMD-Framework) for defect quantification in weld-affected regions, aiming to separate weld interference from defect-related information[14]. In recent years, physical priors and interpretable learning strategies have been further incorporated into deep-learning-based quantification models. Li et al. introduced a prior-knowledge-guided improved wavelet network (PGIWN) for interpretable defect quantification with permanent magnet perturbation[15]. Sun et al. proposed a physics-informed doubly fed cross-residual network (DfedResNet) to improve defect size estimation accuracy[16]. Wang et al. introduced MFL features and expert knowledge into a multilevel knowledge-guided neural network (MLKG-

NN), enhancing the stability of defect size quantification[17]. However, under complex noise, size coupling, and differences among triaxial responses, achieving defect quantification with high accuracy, robustness, and physical credibility still requires further investigation.

Viewed systematically, recent intelligent MFL studies can be organized along two complementary axes. The representation axis progresses from one-dimensional handcrafted descriptors, to single-view MFL images, and then to paired or multi-view representations. The inference axis progresses from target detection, to defect-size regression, and more recently to joint diagnosis. Representative developments since 2024 include cascaded YOLO-ViT detection from MFL images[9], attention-enhanced identification of small girth-weld defects[18], alternating-current magnetic flux leakage (AC-MFL) classification and quantification based on time-domain characteristics[19], a task-oriented physical collaborative network that jointly learns detection and quantification[20], and expertise-informed estimation of irregular-defect size[21]. These studies show that stronger image encoders, physical knowledge, and task interaction are becoming central themes. The present study differs by focusing on paired encodings of the same in-line axial signal, spatial-frequency complementarity, distance-constrained cross-view interaction, and reliability-aware triaxial depth inference under industrial noise.

To address the above challenges, this study investigates MFL defect detection and quantification under complex operating conditions. The main contributions are summarized as follows.

(i) A paired-view MFL defect detection framework is proposed. The homologous MFL signals are constructed into two complementary representations, namely time-intensity images and grayscale images. By integrating spatial-frequency collaborative enhancement and a distance-aware two-dimensional selective scan (SS2D) cross-view fusion mechanism, the proposed framework jointly models temporal evolution information, channel-wise structural information, and frequency-domain discriminative cues, thereby improving the detection stability of weak defects and multiclass targets.

(ii) A multi-axis probabilistic decision quantification method incorporating physical priors is proposed. Structured physical and engineering priors, together with complementary triaxial MFL information, are introduced into the defect depth inversion process. Uncertainty modeling is further employed to suppress high-risk decisions, improving the accuracy, robustness, and credibility of quantification results.

The remainder of this paper is organized as follows. Section 2 presents the proposed detection and quantification methodology. Section 3 reports the experimental results and analysis. Section 4 concludes the paper.

2 Methodology

In MFL in-line inspection of pipelines, multichannel magnetic-field signals contain not only defect-induced local disturbances but also responses from welds, tees, and other structural components, as well as non-stationary factors such as velocity fluctuation, attitude variation, and environmental noise. These factors make different targets highly overlapped in both temporal waveforms and spatial distributions, thereby increasing the risks of false alarms and missed detections. To address the limitations of a single temporal representation in characterizing multichannel spatiotemporal correlations, as well as the lack of collaborative modeling among different representational views, this study proposes a paired-view fusion detection framework based on spatial-frequency collaborative enhancement. The proposed method constructs two complementary representations, namely a time-intensity image and a grayscale image, from the same-source axial MFL signals. A cross-view SS2D fusion mechanism is further introduced at the feature level to jointly model temporal continuity, channel-wise spatial structures, and frequency-domain discriminative information.

2.1 Paired-View Fusion Detection Method Based on Spatial-Frequency Collaborative Enhancement

To improve the detection stability of multiple target categories under complex operating conditions, a paired-view fusion detection framework based on spatial-frequency collaborative enhancement is constructed, as illustrated in Figure 1. First, attitude correction and axial alignment are performed on the multichannel axial MFL signals using inertial measurement unit (IMU) information and odometer data. The same-source signals are then transformed into two complementary views: a time-intensity image (TI) and a grayscale image (GR). Subsequently, the two views are fed into a dual-branch backbone network to extract multiscale features. A spatial-frequency collaborative enhancement (SFCE) module is embedded in the middle- and high-level feature layers to enhance the spatial structural representation and frequency-domain discriminative capability of each single-view feature. On this basis, a distance-aware SS2D fusion block (DASF-Block) is introduced to perform constrained cross-view interaction between TI and GR features at the same scale. Finally, the fused features are sent to the neck for multiscale integration, and the detection head outputs target categories and bounding-box locations.

2.1.1 Paired-View Imaging of MFL Signals

To reduce the influence of detector attitude variation and axial misalignment of the sensor array on imaging consistency, the multichannel signals are first corrected

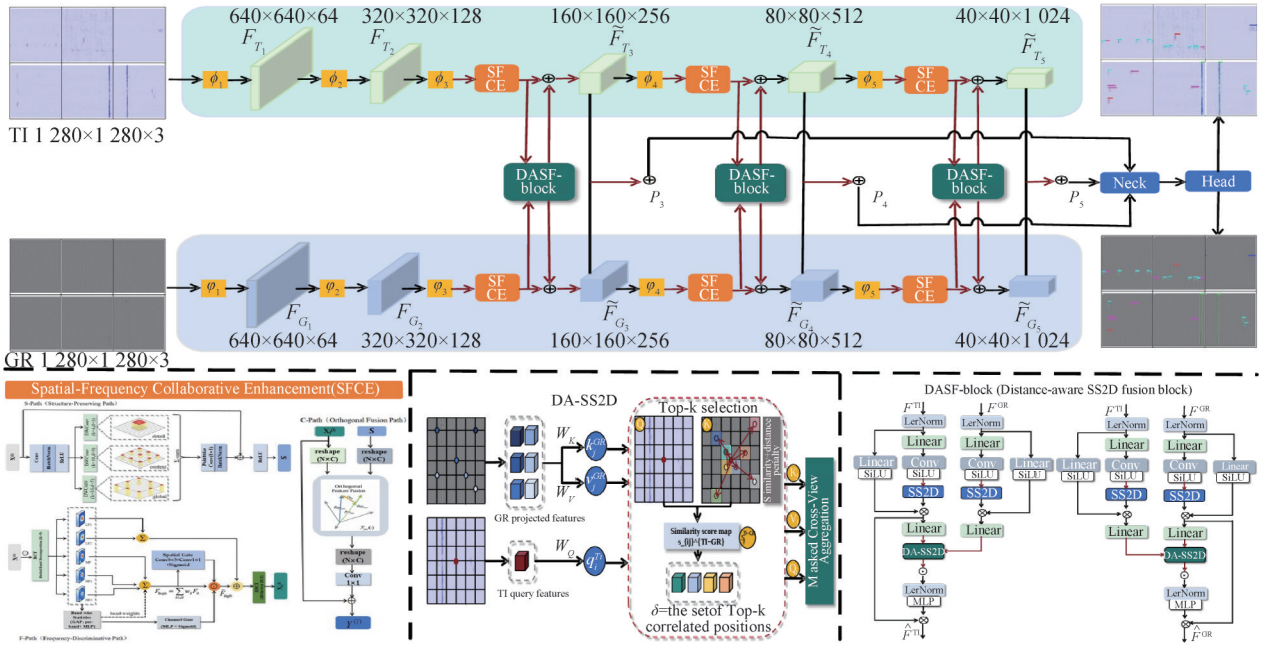


Fig.1 Paired-view fusion detection framework based on spatial-frequency collaborative enhancement. TI and GR denote the time-intensity image and grayscale image, respectively. SFCE is used for spatial-frequency feature enhancement, and DASF-Block is used for distance-aware cross-view SS2D fusion.

using IMU attitude information and odometer information. Let the original axial MFL signal of the c-th channel be $s_c(x)$, where x denotes the mileage coordinate along the pipeline axis, and Δ_c denotes the axial offset of this channel caused by its physical installation position. The corrected signal can be expressed as

$$\tilde{s}_c(x) = s_c(x + \Delta_c) \tag{1}$$

where $\tilde{s}_c(x)$ denotes the aligned channel signal. After correction, the signals from all channels are mapped into the two-dimensional image space according to the order of the sensor array, forming the TI and GR representations. The TI representation emphasizes the temporal evolution characteristics of MFL responses along the mileage direction, whereas the GR representation focuses on spatial textures, edges, and structural distributions of multichannel responses. The two views are derived from the same MFL signal segment and are strictly aligned with the same target annotations, thereby providing consistent inputs for subsequent feature-level fusion.

The terms TI and GR therefore denote two views (two encodings) of one axial MFL measurement, not independent physical modalities. For each approximately 3 m segment, mileage samples are arranged along the horizontal axis and the circumferential sensor channels retain their physical order along the vertical axis after IMU- and odometer-based alignment. A common sample identifier, consisting of the mileage interval and sensor-channel range, links the two 1280 × 1280 images to the same bounding-box annotations. The TI view retains the continuous mileage-wise trajectories and local energy perturbations of individual channels, whereas the GR view renders the aligned channel-by-mileage response field in grayscale to emphasize texture, edge continuity, and structural boundaries. Thus, the views preserve different aspects of the same source matrix while remaining spatially registered.

Reproducibility boundary: the archived project package retains the paired images, labels, and the verifiable alignment procedure, but not the original pixel-level image-generation script. Consequently, the exact numerical normalization, clipping, interpolation, dynamic-range conversion, and missing/saturated-value handling cannot be reconstructed reliably from the retained source. These undocumented choices are not presented as methodological contributions, and exact preprocessing replication remains a limitation of the current study.

2.1.2 Spatial-Frequency Collaborative Feature Enhancement Module

Multiclass MFL targets exhibit substantial overlap in scale distribution. Therefore, relying only on spatial morphology may make the model vulnerable to strong responses from structural components. In contrast, defects and abnormal signals usually contain local mutations, boundary discontinuities, and fine-grained texture variations, which provide useful frequency-domain discriminative cues. Accordingly, the SFCE module models spatial structural information and frequency response information in parallel at the same-scale feature layer.

Let the input feature at the l-th scale be

$$X_l \in \mathbb{R}^{C_l \times H_l \times W_l} \tag{2}$$

where C_l, H_l, and W_l denote the number of channels, height, and width, respectively. The spatial enhancement

path adopts multi-receptive-field depthwise convolutions to extract structural context, and its output is formulated as

$$X_s = X_l + \alpha\phi_s\left(\left[\mathrm{DWConv}_{k_i,d_i}(X_l)\right]_{i=1}^{N}\right) \tag{3}$$

where $\mathrm{DWConv}_{k_i,d_i}$ denotes a depthwise convolution with kernel size k_i and dilation rate d_i; N is the number of parallel branches; $[\cdot]$ denotes channel concatenation; $\phi_s(\cdot)$ is the channel-mixing mapping; α is the residual scaling coefficient; and X_s denotes the spatially enhanced feature.

The frequency enhancement path first maps the feature into the frequency domain through the two-dimensional discrete cosine transform (DCT), while spatial reconstruction after modulation uses the inverse discrete cosine transform (IDCT). For multichannel feature maps, the transform is applied channel-wise:

$$F(u,v) = \alpha_u\alpha_v\sum_{i=0}^{H_l-1}\sum_{j=0}^{W_l-1}X_l(i,j)\cos\frac{\pi(2i+1)u}{2H_l}\cos\frac{\pi(2j+1)v}{2W_l} \tag{4}$$

where $F(u,v)$ denotes the frequency-domain coefficient at the frequency coordinate (u,v), and α_u and α_v are normalization coefficients. According to frequency-band decomposition, the low-frequency component F_L and the high-frequency component F_H are obtained. The high-frequency response is then denoised and modulated through channel gating G_c and spatial gating G_s:

$$X_f = \mathrm{IDCT}(F_L + G_c \odot G_s \odot F_H) \tag{5}$$

where $\mathrm{IDCT}(\cdot)$ denotes the two-dimensional inverse discrete cosine transform, $\odot$ denotes element-wise multiplication, and X_f is the spatial feature enhanced by frequency-domain information. To avoid redundant amplification caused by direct superposition of spatial- and frequency-domain information, the module extracts an orthogonal complementary residual from the frequency-domain feature using the spatial feature as the reference:

$$R_f = X_f - \frac{\langle X_f, X_s\rangle}{\|X_s\|_2^2+\varepsilon}X_s \tag{6}$$

where $\langle\cdot,\cdot\rangle$ denotes the inner product after feature flattening, $\|\cdot\|_2$ denotes the Euclidean norm, ε is a numerical stability term, and R_f denotes the frequency-domain residual complementary to the spatial structure. The final enhanced feature is given by

$$Y_l = X_l + \phi\left(\left[X_s, R_f\right]\right) \tag{7}$$

where $\phi(\cdot)$ is a lightweight channel-fusion mapping, and Y_l denotes the output feature of the SFCE module.

In the retained implementation description, the spatial path uses three depthwise-convolution branches (N = 3) with kernel sizes ki ∈ {4, 10, 16} and dilation rates di ∈ {1, 2, 3}, followed by pointwise channel mixing. In the frequency path, K masked DCT bands provide bandwise channel statistics; global channel gating and local spatial gating jointly modulate the high-frequency component before inverse transformation. The collaboration is therefore not a simple sum of mature spatial and frequency operators. The orthogonal projection in Eq. (6) removes from the frequency response the component collinear with the spatial feature, so that only a complementary residual is fused with the structure-preserving path.

This design is specific to the MFL discrimination problem. Welds and other strong structural components generally exhibit continuous geometry and spatially coherent high-amplitude responses, whereas weak metal loss, corrosion spots, dents, and spike-type anomalies are characterized more strongly by local discontinuities, boundary changes, and fine texture. Directly adding spatial and frequency responses can duplicate the dominant weld response and amplify noise. In contrast, the orthogonal residual suppresses redundant structure-aligned energy while retaining frequency evidence that is not explained by the spatial branch. The single-view ablations in Table 2 support this interpretation at the aggregate level; they do not, however, establish class-specific causal gains without a class-wise module ablation or feature-attribution study.

2.1.3 Distance-Aware SS2D Paired-View Fusion Module

TI and GR differ in salient regions, noise patterns, and discriminative advantages. To prevent strong-response diffusion and false association caused by unconstrained fusion, this study adopts the DASF-Block to perform selective cross-view interaction within relevant candidate regions. Let the paired-view features at the l-th scale be $X_l^{\mathrm{TI}}, X_l^{\mathrm{GR}} \in \mathbb{R}^{C_l\times H_l\times W_l}$.

First, the two features are processed by SS2D selective scanning to propagate long-range contextual information:

$$H_l^m = X_l^m + \mathrm{SS2D}\left(\mathrm{Norm}\left(X_l^m\right)\right), \quad m \in \{\mathrm{TI}, \mathrm{GR}\} \tag{8}$$

where $\mathrm{Norm}(\cdot)$ denotes the normalization operation, and H_l^m denotes the context-enhanced feature of view m. Then, query, key, and value projections are performed on the two features, and cross-view correlation is constructed by combining content similarity with a spatial distance penalty. Taking information transfer from GR to TI as an example, the correlation between the i-th position and the j-th position is defined as

$$a_{ij} = \frac{\left(q_i^{\mathrm{TI}}\right)^T k_j^{\mathrm{GR}}}{\sqrt{d}} - \lambda\left\|p_i - p_j\right\|_2 \tag{9}$$

where q_i^{TI} is the query vector at the i-th position of the TI branch, k_j^{GR} is the key vector at the j-th position of the GR branch, d is the feature dimension, p_i and p_j are spatial coordinates, and λ is the spatial distance penalty coefficient. Based on a_{ij}, the top-k candidate set Ω_i is selected, and information from the GR branch is aggregated only within this set:

$$M_i^{\mathrm{GR}\to\mathrm{TI}} = \sum_{j\in\Omega_i}\mathrm{Softmax}_{j\in\Omega_i}\left(a_{ij}\right)V_j^{\mathrm{GR}} \tag{10}$$

where V_j^{GR} denotes the value vector at the j-th position of the GR branch, and $M_i^{\mathrm{GR}\to\mathrm{TI}}$ denotes the cross-view message written into the TI branch. Finally, a gating mechanism is used to control the writing strength of complementary information:

$$Y_i^{\mathrm{TI}} = H_i^{\mathrm{TI}} + \sigma\left(W_g\left[H_i^{\mathrm{TI}}; M_i^{\mathrm{GR}\to\mathrm{TI}}\right]\right)\odot M_i^{\mathrm{GR}\to\mathrm{TI}} \quad (11)$$

where $[\cdot;\cdot]$ denotes feature concatenation, W_g is the gating mapping, $\sigma(\cdot)$ is the sigmoid function, and Y_i^{TI} denotes the fused TI feature. The GR branch receives information from the TI branch in a symmetric manner. This mechanism constrains cross-view fusion to candidate regions that are both content-related and spatially adjacent. Therefore, it can exploit complementary cues from TI and GR while suppressing the propagation of irrelevant responses, thereby improving the unified detection stability of defects, structural components, and abnormal signals in complex pipeline scenarios.

The original SS2D component provides within-view multi-directional scan expansion, selective state-space recurrence, scan merging, and residual context propagation. DASF-Block adds three operations that are absent from a standalone SS2D layer: symmetric cross-view query/key/value projection, distance-penalized top-k candidate restriction, and gated message write-back to each view. Its essential difference is therefore not the distance term alone; it converts an intra-view state-space operator into a selective and symmetric cross-view interaction mechanism.

Distance is important because a physical MFL target produces a localized response that is continuous over neighboring mileage samples and adjacent sensor channels after alignment. Similar high-amplitude patterns at distant grid positions are more likely to belong to different welds, structures, or noise events. Penalizing such long-range matches and admitting only valid in-grid top-k candidates reduces false cross-view association while SS2D still supplies long-range context within each branch. The archived source defines the top-k and distance-penalty terms symbolically, but it does not retain the numerical top-k value, state width, projection width, or a parameter-sensitivity sweep; no numerical values or robustness claims are introduced beyond the available ablation evidence.

2.2 Multi-Axis Decision-Based Defect Quantification Method Incorporating Physical Prior Knowledge

After MFL defect detection and localization, the defect candidate region can be further mapped to the corresponding mileage coordinates and sensor-channel range, thereby obtaining the axial, circumferential, and radial components of the MFL signal. Unlike detection, defect quantification aims to infer continuous geometric parameters from observed signals and is essentially an ill-posed inverse problem. Because defect length, width, and depth are mutually coupled, and because signal amplitude and morphology are easily affected by magnetization state, lift-off variation, and operational disturbances, it is difficult to ensure the stability and credibility of quantification results by relying only on single-axis amplitude or end-to-end regression. Therefore, this study constructs a multi-axis decision-based defect quantification model incorporating physical prior knowledge, named M²P-DecNet (Macro-Micro Prior Decision Network). The overall workflow is shown in Figure 2. The model consists of three stages. First, triaxial defect signals are cropped according to the detection results and subjected to robust preprocessing. Second, macro-level engineering priors are used to modulate axis-level features and enhance defect-related responses. Finally, micro-level physical prior labels are introduced to drive multi-axis probabilistic decision-making, producing a global predictive distribution of defect depth. In this framework, length and width information is mainly used as geometric prior information, while the probabilistic regression target is the defect depth.

2.2.1 Defect Sample Construction and Robust Preprocessing

Let the signal of the a-th observed axis cropped from the detected candidate region be

$$X^{(a)} \in \mathbb{R}^{T\times S}, \quad a \in \{x, y, z\}, \quad (12)$$

where x, y, and z denote the axial, circumferential, and radial MFL components, respectively; T is the sampling length in the mileage direction; and S is the number of sensor channels. In-service inspection signals usually contain low-frequency background drift and local high-frequency disturbances. Therefore, a sliding-window baseline estimation and detrending operation are first applied to the single-channel discrete sequence $s[k]$:

$$\bar{s}[k] = \frac{1}{W}\sum_{i=k-\lfloor W/2\rfloor}^{k+\lfloor W/2\rfloor} s[i], \quad \tilde{s}[k] = s[k] - \bar{s}[k] \quad (13)$$

where W is the sliding-window length, $\bar{s}[k]$ is the local baseline, and $\tilde{s}[k]$ is the detrended relative response. Boundary handling is implemented according to the available sampling range. A zero-phase finite-impulse-response (FIR) low-pass filter is then used to suppress high-frequency noise:

$$\hat{s}[k] = (\tilde{s} * h)[k] = \sum_{m=0}^{M-1} h[m]\,\tilde{s}[k-m] \quad (14)$$

where $h[m]$ denotes the coefficient of the FIR filter with length M, and * denotes convolution. The processed triaxial matrices $\hat{X}^{(a)}$ are used as unified inputs for subsequent macro-prior construction and axis-level feature encoding.

2.2.2 Representation Modulation Driven by Macro-Level Engineering Priors

To improve the stability of axis-level evidence in depth inversion, this study introduces an expertise-

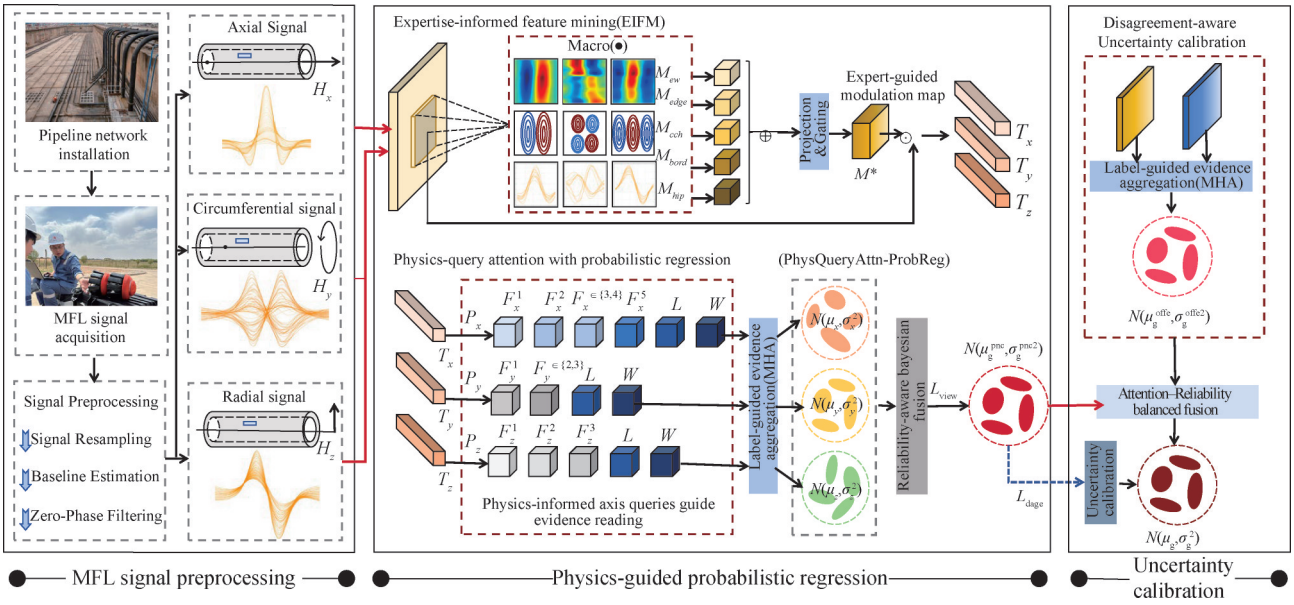


Fig.2 M²P-DecNet architecture; EIFM denotes expertise-informed feature mining.

informed feature mining (EIFM) module at the encoding stage. Based on typical engineering responses of MFL signals, this module constructs a macro-level empirical response-field library:

$$\mathcal{M}\left(X^{(a)}\right)=\left\{M^{(1)},\cdots,M^{(K)}\right\},$$
$$M^{(k)}=G_k\left(X^{(a)}\right)\in\mathbb{R}^{T\times S} \tag{15}$$

where $G_k(\cdot)$ denotes the k-th prior-generation operator, and K is the number of response fields. The response fields adopted in the original framework include the energy-envelope field, boundary-variation field, cross-channel consistency field, multiscale morphology-disturbance decomposition field, and enhanced response field designed for bipolar responses. To avoid redundancy caused by equal-weight superposition of different response fields, the model assigns a weight α_k to each response field through a sparse adaptive selection mechanism, thereby obtaining a combined macro-level response field:

$$M^{*}=\sum_{k=1}^{K}\alpha_k M^{(k)} \tag{16}$$

The operators have the following engineering definitions. Let X denote the preprocessed response matrix for one axis, with t as the mileage index and s as the sensor-channel index, and let Smooth(·) denote light two-dimensional smoothing (S elsewhere denotes the number of channels). The envelope and boundary fields smooth, respectively, the response magnitude and its absolute first difference along mileage. The coherence field uses the median absolute deviation MAD(t) across channels at each t and assigns $M_{coh}(t,s)=\exp[-\mathrm{MAD}(t)/\tau]$ to every channel, where $\tau>0$ controls the decay. For the morphology-disturbance field, LP_L and LP_H are mileage-direction low-pass filters with lower and higher cutoff frequencies, respectively. Thus $X_{low}=LP_L(X)$ retains the broad response, whereas $X_{high}=X-LP_H(X)$ retains the fine-scale residual. Their weighted field is $M_{band}=\mathrm{Smooth}(\alpha|X_{low}|+(1-\alpha)|X_{high}|)$, where $0\leqslant\alpha\leqslant1$ is a mixing coefficient distinct from the selection weights α_k in Eq. (16). For an axis with polarity reversal, u(t) is the median of X(t,s) over all channels s. For $t\geqslant2$, set $b(t)=|u(t)-u(t-1)|$ when $u(t)u(t-1)<0$ and $b(t)=0$ otherwise; set $b(1)=0$. Replicating b(t) across channels and smoothing gives the bipolar response field. Sparse adaptive selection combines the fields before the soft-mask mapping in Eqs. (17)-(18).

The combined response field M^{*} is then downsampled to the encoding feature scale and mapped into a soft mask:

$$A=\sigma\left(f\left(\mathrm{Norm}\left(\mathrm{pool}\left(M^{*}\right)\right)\right)-0.5+\beta\right),$$
$$A\leftarrow\alpha_{\min}+(1-\alpha_{\min})A, \tag{17}$$

where $\sigma(\cdot)$ is the sigmoid function, $f(\cdot)$ denotes the mapping function, $\mathrm{pool}(\cdot)$ denotes the scale-matching operation, β is the bias term, and $\alpha_{\min}$ is the lower bound of the mask. Let the axis-level encoded feature be $F^{(a)}$. The modulated feature is formulated as

$$\tilde{F}^{(a)}=F^{(a)}\odot A^{(a)},\ T_a=\mathrm{Flatten}\left(\tilde{F}^{(a)}\right)\in\mathbb{R}^{N\times d} \tag{18}$$

where $\odot$ denotes element-wise multiplication, N is the number of flattened local features, and d is the feature dimension. Through this mechanism, the model can highlight stable regions consistent with defect responses while suppressing structural interference and random noise.

2.2.3 Intra-Axis Evidence Reading Guided by Micro-Level Physical Prior Labels

Macro-level priors mainly function at the representation level, whereas depth quantification further requires physical constraints at the decision level. In this study, typical observables in the triaxial signals are organized into micro-level physical prior labels:

$$P \to \left(P_x, P_y, P_z\right) \tag{19}$$

where P_a denotes the physical label vector corresponding to the a-th observed axis. According to the original setting, the axis-specific labels consist of peak-valley amplitude, peak-valley spacing, skewness, and geometric conditions such as length and width:

$$P_x = \left[F_x^1, F_x^2, F_x^3, F_x^4, F_x^5\right], \quad P_y = \left[F_y^1, F_y^2, F_y^3\right],$$
$$P_z = \left[F_z^1, F_z^2, F_z^3\right]. \tag{20}$$

Here, F_a^i denotes the i-th type of depth-related physical observable on the a-th axis. The geometric conditions in P_a are used to constrain depth inference rather than to redefine the regression target. To make the physical labels participate in intra-axis evidence reading, P_a is first mapped into an axis-level query vector:

$$q_a = f_a(P_a) \in \mathbb{R}^d \tag{21}$$

where $f_a(\cdot)$ denotes the axis-specific label mapping function. Then, a multi-head attention (MHA) mechanism is used to read depth-related evidence consistent with the physical conditions from the axis-level feature sequence T_a:

$$e_a = \mathrm{MHA}\left(q_a, T_a, T_a\right) \in \mathbb{R}^d, \quad a \in \{x, y, z\} \tag{22}$$

where $\mathrm{MHA}(\cdot)$ denotes multi-head attention, and e_a is the evidence representation of the a-th axis. This process does not impose hard cropping on the original signals. Instead, it adaptively selects effective response regions through a physical-label-conditioned attention mechanism.

2.2.4 Multi-Axis Probabilistic Fusion and Uncertainty Calibration

After obtaining the triaxial evidence representations, they are organized as

$$E = \left[e_x; e_y; e_z\right] \in \mathbb{R}^{3\times d} \tag{23}$$

where $[\cdot;\cdot]$ denotes stacking or concatenation along the evidence dimension. The global physical prior P is mapped as a conditional query and interacts with the triaxial evidence to generate the Gaussian parameters of the attention branch:

$$\left(\mu_g^{\mathrm{attn}}, \log\left(\left(\sigma_g^{\mathrm{attn}}\right)^2\right)\right) = h_g\left(\mathrm{MHA}\left(f_g(P), E, E\right)\right) \tag{24}$$

where μ_g^{attn} and $\left(\sigma_g^{\mathrm{attn}}\right)^2$ denote the global mean and variance predicted by the attention branch, respectively; $f_g(\cdot)$ and $h_g(\cdot)$ are mapping functions; and the log-variance output is used to ensure that the variance remains positive after exponentiation.

Meanwhile, the output of each axis is modeled as a Gaussian distribution $N\left(\mu_a, \sigma_a^2\right)$, and the axis-level precision is defined as the inverse of the variance $\pi_a = \sigma_a^{-2}$. Under the assumption of conditionally independent Gaussian axis-level observations, the reliability-based fusion of triaxial evidence can be written as:

$$\mu_g^{\mathrm{prec}} = \sum_{a\in\{x,y,z\}} w_a \mu_a,$$
$$\left(\sigma_g^{\mathrm{prec}}\right)^2 = \left[\sum\nolimits_{a\in\{x,y,z\}} \pi_a\right]^{-1}, w_a = \frac{\pi_a}{\sum_{k\in\{x,y,z\}} \pi_k} \tag{25}$$

where w_a denotes the reliability weight obtained by normalizing the axis-level precision. The final global predictive distribution is obtained by interpolating the attention branch and the precision branch:

$$\mu_g = \alpha\mu_g^{\mathrm{attn}} + (1-\alpha)\mu_g^{\mathrm{prec}},$$
$$\sigma_g^2 = \alpha\left(\sigma_g^{\mathrm{attn}}\right)^2 + (1-\alpha)\left(\sigma_g^{\mathrm{prec}}\right)^2 \tag{26}$$

where $\alpha \in [0, 1]$ is the balancing coefficient between the two fusion mechanisms. To prevent the model from becoming overconfident when there is disagreement among multi-axis predictions, the inter-axis mean discrepancy is further defined as

$$\mathrm{dis} = \frac{\left(\mu_x - \mu_y\right)^2 + \left(\mu_x - \mu_z\right)^2 + \left(\mu_y - \mu_z\right)^2}{3} \tag{27}$$

and conservative calibration is applied to the global variance:

$$\sigma_g^2 \leftarrow \sigma_g^2 + \lambda_{\mathrm{dis}}\mathrm{dis} \tag{28}$$

where λ_{dis} is the discrepancy calibration coefficient. The final output of the model is the defect-depth predictive distribution $N\left(\mu_g, \sigma_g^2\right)$, where μ_g is the estimated depth, and σ_g^2 characterizes the prediction uncertainty.

2.2.5 Training Objective

Given the true defect depth d, the model uses the Gaussian negative log-likelihood to constrain the global predictive distribution:

$$L_g = \frac{1}{2}\left[\frac{\left(d-\mu_g\right)^2}{\sigma_g^2+\varepsilon} + \log\left(\sigma_g^2+\varepsilon\right)\right] \tag{29}$$

where ε is a numerical stability term. To prevent degradation of the axis-level branches, weak supervision is introduced for the triaxial predictive distributions:

$$L_{\mathrm{view}} = \frac{1}{3}\sum\nolimits_{a\in\{x,y,z\}} \frac{1}{2}\left[\frac{\left(d-\mu_a\right)^2}{\sigma_a^2+\varepsilon} + \log\left(\sigma_a^2+\varepsilon\right)\right] \tag{30}$$

Furthermore, to constrain the consistency between the axis-level distributions and the global distribution, a discrepancy-aware distribution alignment term is introduced:

$$L_{\mathrm{align}} = \mathrm{gate}\cdot\frac{1}{3}\sum\nolimits_{a\in\{x,y,z\}}\left[\left(\mu_a-\mu_g\right)^2 + \left(\sigma_a-\sigma_g\right)^2\right],$$
$$\mathrm{gate} = \max\left\{\eta, \exp\left(-k\,\mathrm{dis}\right)\right\} \tag{31}$$

where gate denotes the alignment-strength gate, $\eta \in (0, 1)$ is the lower bound of the gate, and $k > 0$ controls the sensitivity of the gate to inter-axis discrepancy. The total loss is defined as:

$$L = L_g + \lambda_{\mathrm{view}}L_{\mathrm{view}} + \lambda_{\mathrm{align}}L_{\mathrm{align}}, \tag{32}$$

where λ_{view} and λ_{align} denote the weights of the axis-level

supervision term and the distribution alignment term, respectively. This objective function simultaneously constrains global regression accuracy, axis-level output usability, and cross-axis prediction consistency, enabling the model to achieve more stable and credible defect-depth quantification under complex operating conditions and uncertain annotation scenarios.

3 Results

3.1 Experimental Platform and Dataset

The experiments were conducted to evaluate the proposed method in both multiclass MFL target detection and defect-depth quantification. The overall experimental data consist of field MFL in-line inspection data collected from in-service pipelines and artificial-defect pipeline data used for quantitative evaluation. The field data were mainly used to verify the detection capability of the proposed paired-view fusion framework under practical operating conditions, while the artificial-defect data were used to evaluate the accuracy and stability of defect-depth inversion.

The archived software environment was Ubuntu 22.04.3, Python 3.10.8, PyTorch 2.1.2+cu121, and CUDA runtime 12.1 on an NVIDIA GeForce RTX 4090 D (24 GB), with an Intel Xeon Platinum 8474C host. All reported ablations and model comparisons used a common retained data assignment, output scale, training strategy, and evaluation definition. The source package does not contain the complete optimizer schedule, batch size, epoch count, augmentation recipe, all loss weights, or parameter-matched compute profiles for every comparison model; these values are therefore not reconstructed or inferred.

The MFL in-line detector used in this study consists of a hardware platform, a multi-sensor acquisition unit, and a mechanical adaptation structure. The hardware platform is mainly responsible for magnetization and sampling timing control, multichannel data acquisition, synchronous calibration, data storage, and system power supply. The multi-sensor acquisition unit contains two rows of circumferentially arranged magnetic sensor array, an IMU, and three odometer wheels. The magnetic sensor array is used to acquire the leakage magnetic-field responses induced by pipe-wall discontinuities, while the IMU and odometer wheel provide attitude and mileage information for motion-state perception, spatial alignment, and defect localization. The mechanical structure integrates the above modules and adopts a centering and protective design, including guide cups, support wheels, odometer wheels, and anti-collision components, so that the detector can pass through complex pipeline structures such as welds, bends, and local geometric variations while maintaining stable lift-off and acquisition consistency. The arrangement of these detector components is shown in Figure 3.

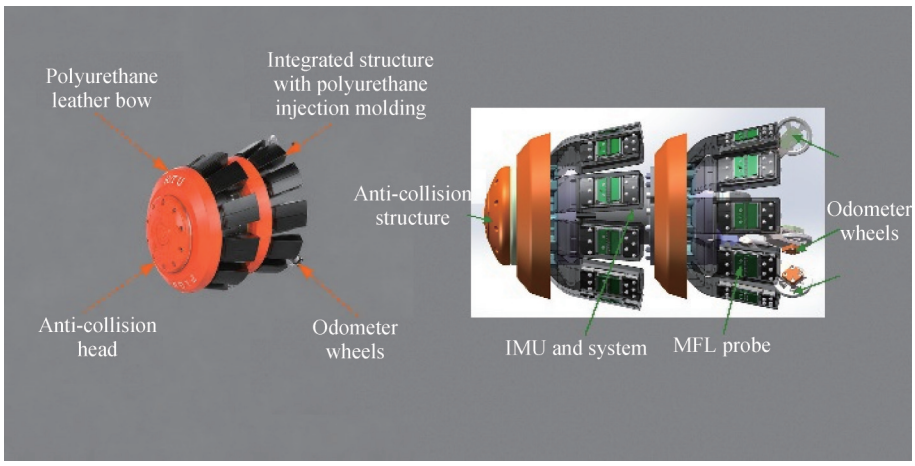


Fig.3 Structure of MFL in-line inspection detector

During field inspection, the detector runs inside the pipeline and continuously records multichannel MFL signals together with IMU and mileage information. Because in-service buried pipelines usually span long distances and contain complex operating conditions, all acquired data were exported after inspection and processed offline. The offline processing workflow includes data synchronization, mileage correction, sensor-channel alignment, signal normalization, image construction, and target annotation. Based on the same-source MFL signal segments, two aligned image representations were generated for detection: the time-intensity image (TI), which emphasizes the temporal evolution of multichannel responses along the mileage direction, and the grayscale image (GR), which highlights spatial texture, channel distribution, and structural boundaries. These two views share the same defect annotations and are used as paired inputs for the proposed paired-view detection model.

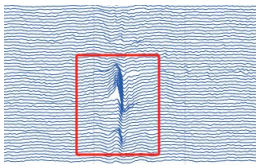
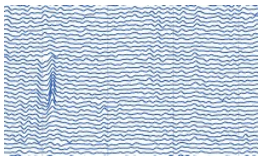
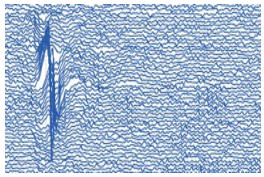
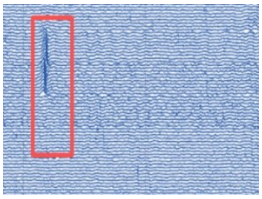
The field detection data were collected from two in-service gas pipelines. Gas pipeline A covers approximately 13 km, with an outer diameter of 219 mm, a wall thickness of 7 mm, and 200 probes. Its relatively high signal-to-noise ratio (SNR) records contain welds,

dents, metal loss, corrosion spots, spike-type abnormal responses, and review-required regions. Gas pipeline B covers approximately 21 km with the same outer diameter, wall thickness, and probe count, but contains longer low-SNR sections in which background noise and structural texture more strongly mask weak MFL responses. These pipelines provide contrasting signal-quality conditions, but not a cross-diameter field evaluation.

The detection dataset was constructed from the field MFL inspection data. Each sample covers approximately 3 m of pipeline mileage and is converted into a paired TI-GR image sample with a resolution of 1280 × 1280. A total of 9,200 paired samples were retained and assigned to training, validation, and testing sets with 6,900, 1,380, and 920 samples, respectively. The classes are grouped as structures, defects, and abnormal/review responses. To replace ambiguous legacy names, this paper uses Weld, Tee (legacy label: 'Three links'), Branch connection ('Branch pipe'), Metal loss ('Flaw'), Dent ('Sunken'), Corrosion spot, Spike-type abnormal response ('Harsh'), and Review-required region ('Key focus areas'). The last class denotes an atypical or ambiguous response retained for engineering review rather than being forced into a confirmed defect category.

Field labels were prepared under a common annotation specification by the authors and industrial technicians using the raw multichannel traces, mileage information, paired images, and inspection records. Signal-region candidates were initially screened and then manually reviewed; representative geometric labels were aligned with excavation or external-measurement records when available. For atypical responses without sufficient physical confirmation, the review-required class was used. The retained record does not contain an inter-rater agreement statistic, so the manuscript does not claim a quantified annotator-consistency level. Representative field verification cases are summarized in Table 1.

Table 1 Representative field verification cases.

Pipeline	Field Photograph	Defect Category	Signal Visualization Result	Size (mm)
Gas pipeline A		Dent		67 × 70 × 4.60
		Metal loss		65 × 20 × 1.20
Gas pipeline B		Dent		50 × 40 × 5.49
		Metal loss		65 × 70 × 0.70

For defect-depth quantification, artificial-defect pipeline data were used to construct a supervised evaluation dataset. The straight-pipe platforms contained 23, 87, 61, 59, and 71 physical defects at diameters of 168, 219, 273, 323, and 406 mm, with wall thicknesses of 5, 7, 7, 8, and 8 mm, respectively. The 219 mm loop platform contained 303 defects in a 7 mm wall and included bends, tees, and flanges. The resulting 604 samples cover depths of approximately 0.43-6.06 mm, lengths of 1-250 mm, and widths of 1-120 mm. Each record includes raw triaxial MFL signals and labels for mileage, channel range, length, width, and depth. Ten-fold cross-validation was used in the original experiment; however, the retained fold records do not establish grouping by physical defect identity or acquisition run and do not provide fold-level diameter/platform

distributions. Accordingly, the reported cross-validation measures within-dataset stability and is not presented as leave-one-diameter-out, leave-one-platform-out, or independent field-transfer validation.

For detection, we used precision, recall, mean average precision at an intersection-over-union (IoU) threshold of 0.50 (mAP50), and mean average precision averaged over IoU thresholds from 0.50 to 0.95 (mAP50:95). For quantification, mean absolute error (MAE), the coefficient of determination (R^2), and the proportion of samples whose relative error is within 10% (Acc@10%) were used to evaluate regression error, statistical agreement, and engineering-tolerance satisfaction, respectively.

3.2 Detection Result Analysis

Table 2 presents the ablation results of the paired-view detection model. Compared with the single-view baselines, the introduction of SFCE significantly improves the localization quality for both TI and GR inputs. Specifically, $mAP_{50:95}$ increases from 29.2% to 39.4% with TI input and from 31.6% to 40.5% with GR input. This indicates that spatial-frequency collaborative enhancement can strengthen weak-defect boundaries and fine-grained discriminative features. In contrast, naive feature concatenation of TI and GR does not bring performance improvement. The corresponding mAP_{50} and $mAP_{50:95}$ decrease to 53.1% and 28.9%, respectively, suggesting that unconstrained fusion may introduce cross-view noise accumulation and interference from strong structural responses. After further introducing the DASF-Block, $mAP_{50:95}$ increases to 45.4%, validating the importance of distance-aware constraints and SS2D-based cross-view interaction for stable fusion. The complete model, which incorporates both SFCE and DASF-Block, achieves the best overall performance, reaching 72.5%, 65.3%, and 48.1% in precision, mAP_{50}, and $mAP_{50:95}$, respectively.

Table 2 Ablation results of the detection model

Method	Input	SFCE	DASF	Precision (%)	Recall (%)	$\boldsymbol{mAP}_{50}$ (%)	$\boldsymbol{mAP}_{50:95}$ (%)
Baseline	TI	-	-	58.1	49.3	55.5	29.2
Baseline	GR	-	-	59.3	54.1	57.2	31.6
+SFCE	TI	√	-	65.2	56.4	59.1	39.4
+SFCE	GR	√	-	64.7	55.9	62.4	40.5
Naive merge	TI+GR	-	-	56.2	52.2	53.1	28.9
+DASF	TI+GR	-	√	66.8	62.4	64.7	45.4
Ours	TI+GR	√	√	72.5	61.8	65.3	48.1

For the comparative evaluation, the FMB, TFDet, and SSCT-Net baselines were implemented within our project pipeline, and their input interfaces were adapted to accept the aligned TI+GR image pair. The archived experiment records document a common 6,900/1,380/920 training/validation/test split, a unified training strategy, and identical evaluation definitions for all methods. Accordingly, the reported comparison values were generated on our MFL dataset under the same experimental setting and were not copied from the original publications. Under this common protocol, the proposed method achieves mAP50 and mAP50:95 values of 65.3% and 48.1%, respectively, compared with 64.7% and 46.9% for FMB, 62.3% and 44.7% for TFDet, and 59.3% and 34.5% for SSCT-Net. These results indicate that the proposed spatial-frequency feature enhancement and constrained cross-view fusion improve feature discrimination and paired-view fusion stability under complex MFL backgrounds.

For the higher-SNR field pipeline, Weld and Spike-type abnormal response achieved recalls of 93.98% and 93.05%, whereas Metal loss and Dent achieved 80.99% and 81.95%. Metal loss and Dent still showed mutual confusion of approximately 0.11-0.12, Dent precision was approximately 68.3%, and Corrosion spot exhibited high precision (approximately 96.2%) but lower recall (approximately 73.0%). Tee and Branch connection had only two and one ground-truth instances, respectively, so their class statistics are not stable. For the lower-SNR pipeline, background noise and structural texture increasingly masked weak, small, and boundary-ambiguous responses; strong Weld responses remained reliable, but defect and review-required classes exhibited more missed detections and false alarms. These results show engineering utility within the studied pipelines while also identifying the principal recall bottleneck.

3.3 Defect Quantification Result Analysis

Table 3 presents the ablation results of the M²P-DecNet quantification model. After removing EIFM, the MAE increases to 0.701 mm, indicating that macro-level engineering priors play an important role in enhancing defect-related responses. When probabilistic regression is replaced with deterministic regression, model performance also decreases, suggesting that uncertainty modeling helps mitigate the effects of noise and multi-axis disagreement in depth inversion. After further removing reliability modeling, the MAE is 0.398

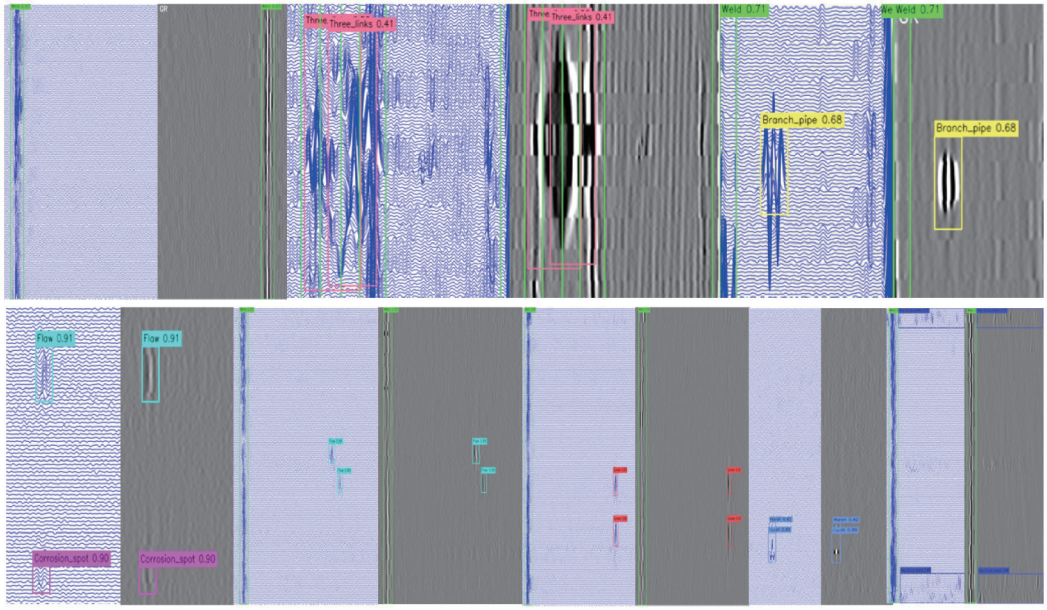


Fig.4 Detection result visualization examples.

mm, which is higher than that of the complete model. This demonstrates that reliability fusion based on axis-level variance can effectively improve the utilization efficiency of multi-axis evidence. The complete model achieves the best results, with an MAE of 0.240 mm, an R^2 of 0.981, and an Acc@10% of 0.905. Meanwhile, the smaller standard deviations indicate that the model has better stability across different validation folds.

Table 3 Ablation results of the quantification model

Method	Description	MAE (mm)	R^2	Acc@10%
A0	Without EIFM	0.701±0.119	0.935±0.081	0.764±0.116
A1	Deterministic regression	0.591±0.104	0.947±0.075	0.778±0.105
A2	Without reliability modeling	0.398±0.089	0.960±0.041	0.894±0.063
A3	Complete model	0.240±0.052	0.981±0.032	0.905±0.048

Table 4 compares the proposed method with existing defect quantification methods. Compared with CNN, VT-CNN, DfedResNet, PGIWN, FMD-Framework, and MLKG-NN, the proposed method achieves the lowest MAE and the highest R^2 and Acc@10%. Compared with MLKG-NN, the MAE is reduced from 0.285 mm to 0.240 mm, while Acc@10% increases from 0.884 to 0.905. These results indicate that the joint modeling of macro-level engineering priors, micro-level physical-prior queries, and multi-axis probabilistic decision-making can effectively improve the accuracy and robustness of defect depth estimation.

Table 4 Performance comparison of different quantification methods

Method	MAE (mm)	R^2	Acc@10%
CNN	0.844±0.241	0.891±0.117	0.652±0.127
VT-CNN	0.787±0.178	0.943±0.082	0.741±0.107
DfedResNet	0.667±0.211	0.941±0.072	0.764±0.113
PGIWN	0.535±0.151	0.967±0.021	0.821±0.081
FMD-Framework	0.402±0.096	0.956±0.038	0.851±0.061
MLKG-NN	0.285±0.061	0.976±0.051	0.884±0.055
Ours	0.240±0.052	0.981±0.032	0.905±0.048

The retained cross-validation analysis also showed that absolute depth error increased with the model's predicted uncertainty. Low-uncertainty samples had more concentrated errors, whereas cases with stronger triaxial

disagreement or disturbance received larger uncertainty and more conservative estimates. This supports uncertainty as a useful warning signal, but the retained records do not provide a complete error stratification by depth interval, noise level, pipe diameter, or platform. Therefore, the present results do not establish the exact operating range in which M²P-DecNet improves most, and individual large-error cases cannot be assigned to a single physical cause.

Finally, the detection and quantification results were compared with excavation measurements. In typical cases, the predicted size of an external metal-loss defect was 85 mm×90 mm×3.00 mm, while the measured size was 92 mm×90 mm×2.85 mm. For a dent defect, the predicted size was 100 mm×77 mm×3.74 mm, while the measured size was 105 mm×75 mm×3.75 mm. These results show that the proposed method can achieve stable defect localization and credible depth quantification in practical MFL inspection scenarios.

3.4 Engineering Scope, Failure Modes, and Limitations

Engineering workflow and efficiency. The detection and quantification networks are algorithmically separate but operationally cascaded: the paired-view detector first locates candidate regions, after which M²P-DecNet is invoked only for selected candidates. This structure is compatible with the offline processing workflow used for the long in-service inspection runs and avoids dense quantification over every mileage position. Nevertheless, the retained study did not record parameter counts, floating-point operations (FLOPs), latency, or frames per second (FPS) for the complete cascade and comparison models. The manuscript therefore makes no real-time claim, and computational fairness cannot be assessed beyond the common dataset and metric definitions.

Partitioning and generalization. The two field pipelines both have an outer diameter of 219 mm and a wall thickness of 7 mm. The archived 6,900/1,380/920 assignment does not specify whether 3 m windows were non-overlapping or grouped by mileage block, physical target, or pipeline, and it does not retain training-only normalization statistics. Spatial leakage between nearby windows cannot therefore be excluded. Likewise, the 604-sample ten-fold record does not verify grouping by physical defect and acquisition run. Results should be interpreted as within-dataset evidence, not as an unseen-pipeline, leave-one-diameter-out, leave-one-platform-out, or artificial-to-field generalization claim.

Detection failure modes and visual evidence. The aggregate ablation shows that SFCE raises single-view mAP50: 95 by 10.2 percentage points for TI and 8.9 points for GR, while DASF-Block raises naive paired-view mAP50: 95 from 28.9% to 45.4%. These results support enhanced localization and constrained interaction, but they do not isolate class-wise module gains. Missed detections are concentrated in weak, small, low-SNR, and boundary-ambiguous responses, where each view may contain insufficient evidence and the current fusion block may still propagate residual structural interference. Figure 4 provides output-level examples; feature maps and gradient-weighted class activation mapping (Grad-CAM) visualizations were not retained and are not reconstructed post hoc.

Reproducibility and sensitivity. The available thesis supports the module equations, the three-branch SFCE setting, the EIFM response fields, the dataset ranges, and the reported ablations. It does not contain executable code or verified numerical records for all architectural dimensions, top-k, distance penalty, residual and fusion coefficients, discrepancy coefficient, alignment weights, or sensitivity sweeps. The paper therefore avoids inventing these values. Releasing a versioned preprocessing and training pipeline and conducting grouped, cross-domain, sensitivity, and efficiency experiments are necessary before deployment claims can be generalized.

4 Conclusion

This study tested a linked detection and depth-estimation workflow for MFL inspection. On the 920-sample field test set, paired-view detection reached 65.3% mAP50 and 48.1% mAP50: 95, 0.6 and 1.2 percentage points above adapted Fusion-Mamba. Constrained cross-view fusion raised mAP50: 95 from 28.9% with naive merging to 45.4%, supporting selective interaction under strong structural interference. Across 604 artificial-defect samples, ten-fold depth validation achieved 0.240 ± 0.052 mm mean absolute error (MAE) and R^2 of 0.981 ± 0.032, versus 0.285 ± 0.061 mm MAE for MLKG-NN. MAE rose to 0.701 mm without engineering-prior modulation and 0.398 mm without reliability weighting. These findings support improved multiclass localization and depth estimation within the evaluated datasets.

The findings remain within-dataset evidence: both field gas pipelines have the same diameter, and depth regression was assessed on artificial defects. The archived records do not verify separation by mileage block or physical defect across partitions; leakage-free grouped tests and independent pipeline, diameter, and platform evaluations are needed before transfer claims. Future work should document the full image-generation, preprocessing, and training settings and measure runtime, parameter sensitivity, and predictive-interval coverage. These checks will establish how broadly the measured detection and depth-estimation gains carry beyond the evaluated offline setting.

Author Contributions:

Tianxi Wen: Methodology, investigation, data curation, software, validation, visualization, writing—original draft, and writing—review and editing; Quanwei

Yu: resources, domain expertise in pipeline management and technical support; Meiling Li: conceptualization, methodology, investigation, data curation, formal analysis, validation, and visualization; Bin Gao: conceptualization, methodology, project administration, funding acquisition, resources, supervision, and writing—review and editing; Gaige Ru: formal analysis, validation, and writing—review and editing; Chao Wang: resources, funding acquisition, data curation, and validation; Fei Luo: resources, data curation, and validation; Zugao Chen: resources, data curation, and validation.

Funding Information:

The work was supported by the "Enterprises Seeking Technology" Two-Way "JieBangGuaShuai" Project, Research Project on Intelligent In-Line Inspection Robot Technology for Unpiggable Pipelines and the National Natural Science Foundation of China (No. 62401107).

Data Availability:

The in-service pipeline inspection data and industrial project records supporting this study are not publicly available because of confidentiality agreements. Aggregated experimental results and non-confidential methodological details are reported in the manuscript.

Conflicts of Interest:

The authors declare no competing interests.

Ethics Statement

Not applicable.

Use of Generative AI and AI-Assisted Technologies

No generative AI was used.

References

[1] Feng B., Wu J., Tu H., Tang J., & Kang Y. (**2022**). A review of magnetic flux leakage nondestructive testing. *Materials,* 15 (20), 7362. doi:10.3390/ma15207362

[2] Chen T., & Yang X. (**2025**). Research on magnetic flux leakage detection data analysis and corrosion defect identification model for boiler water walls based on deep learning. 2025 International Conference on Electrical Drives, Power Electronics & Engineering (EDPEE), Athens, *Greece,* 267-272. doi:10.1109/EDPEE65754.2025.00051

[3] Lu S., Feng J., Zhang H., Liu J., & Wu Z. (**2019**). An estimation method of defect size from MFL image using visual transformation convolutional neural network. *IEEE Transactions on Industrial Informatics,* 15(1), 213-224. doi: 10.1109/TII.2018.2828811

[4] Yan W., Liu W., Zhang Q., Bi H., Jiang C., Liu H., Wang T., Dong T., & Ye X. (**2024**). Multisource multimodal feature fusion for small leak detection in gas pipelines. *IEEE Sensors Journal,* 24(2), 1857-1865. doi:10.1109/JSEN.2023.3337228

[5] Baroudi U., Al-Roubaiey A. A., & Devendiran A. (**2019**). Pipeline leak detection systems and data fusion: A survey. *IEEE Access*,7,97426-97439. doi:10.1109/ACCESS.2019.2928487

[6] Wang F., Lin W., Liu Z., Wu S., & Qiu X. (**2017**). Pipeline leak detection by using time-domain statistical features. *IEEE Sensors Journal,* 17(19), 6431-6442. doi: 10.1109/JSEN.2017.2740220

[7] Afzal M., & Udpa S. (**2002**). Advanced signal processing of magnetic flux leakage data obtained from seamless gas pipeline. *NDT & E International,* 35(7), 449-457. doi: 10.1016/S0963-8695(02)00024-5

[8] Mukhopadhyay S., & Srivastava G. P. (**2000**). Characterisation of metal loss defects from magnetic flux leakage signals with discrete wavelet transform. *NDT & E International,* 33(1), 57-65. doi:10.1016/S0963-8695(99)00011-0

[9] Chen P., Li R., Fu K., Zhong Z., Xie J., Wang J., & Zhu J. (**2024**). A cascaded deep learning approach for detecting pipeline defects via pretrained YOLOv5 and ViT models based on MFL data. *Mechanical Systems and Signal Processing,* 206, 110919. doi:10.1016/j.ymssp.2023.110919

[10] Dong W., Zhu H., Lin S., Luo X., Shen Y., Guo G., & Zhang B. (**2025**). Fusion-Mamba for cross-modality object detection. *IEEE Transactions on Multimedia,* 27, 7392-7406. doi: 10.1109/TMM.2025.3599020

[11] Zhang X., Zhang X., Wang J., Ying J., Sheng Z., Yu H., Li C., & Shen H.-L. (**2025**). TFDet: Target-aware fusion for RGB-T pedestrian detection. *IEEE Transactions on Neural Networks and Learning Systems,* 36(7), 13276-13290. doi: 10.1109/TNNLS.2024.3443455

[12] Shen X., Liu J., Sun J., Jiang L., Zhao H., & Zhang H. (**2023**). SSCT-Net: A semisupervised circular teacher network for defect detection with limited labeled multiview MFL samples. *IEEE Transactions on Industrial Informatics,* 19(10), 10114-10124. doi:10.1109/TII.2022.3232764

[13] Piao G., Ling J., & Li J. (**2023**). Classification and characterization of coexisting defects from magnetic flux leakage data using deep learning method. *AIP Advances,* 13 (1), 015029. doi:10.1063/9.0000451

[14] Zhang X., Feng J., Lu S., & Zhang S. (**2023**). *FMD-Framework: A size estimation method for pipeline defects in weld-affected zones*. IEEE Transactions on Instrumentation and Measurement, 72, 3508711 (1-11). doi: 10.1109/TIM.2023.3250239

[15] Li J., Liu S., Han F., Zhang B., Cao C., & Wang H. (**2026**). A defect quantification method based on non-destructive testing of permanent magnet perturbation and interpretable wavelet networks. *Measurement,* 257, 118776. doi: 10.1016/j.measurement.2025.118776

[16] Sun H., Peng L., Huang S., Li S., Long Y., Wang S., & Zhao W. (**2022**). Development of a physics-informed doubly fed cross-residual deep neural network for high-precision magnetic flux leakage defect size estimation. *IEEE Transactions on Industrial Informatics,* 18(3), 1629-1640. doi: 10.1109/TII.2021.3089333

[17] Wang L., Zhang H., Liu J., Qu F., & Zuo F. (**2023**). Defect

size quantification for pipeline magnetic flux leakage detection system via multilevel knowledge-guided neural network. *IEEE Transactions on Industrial Electronics,* 70(9), 9550-9560. doi:10.1109/TIE.2022.3210557

[18] Xu L., Dong S., Wei H., Peng D., Qian W., Ren Q., Wang L., & Ma Y. (**2024**). Intelligent identification of girth welds defects in pipelines using neural networks with attention modules. *Engineering Applications of Artificial Intelligence,* 127, 107295. doi:10.1016/j.engappai.2023.107295

[19] Wang R., Chen Y., Yu H., Xu Z., Tang J., Feng B., Kang Y., & Song K. (**2025**). Defect classification and quantification method based on AC magnetic flux leakage time domain signal characteristics. *NDT & E International,* 149, 103250. doi:10.1016/j.ndteint.2024.103250

[20] Shen X., Liu J., Ren Y., Jiang L., Wang L., Zhao H., & Li R. (**2025**). A task-oriented physical collaborative network for pipeline defect diagnosis in a magnetic flux leakage detection system. *Computers in Industry,* 169, 104290. doi: 10.1016/j.compind.2025.104290

[21] Wang L., Zhang H., Liu J., Shen X., & Zuo F. (**2024**). Irregular defect size estimation for the magnetic flux leakage detection system via expertise-informed collaborative network. *IEEE Transactions on Industrial Electronics,* 71 (10), 13189-13200. doi:10.1109/TIE.2024.3357845