Interpretable Feature Decomposition and Semantic Representation in Complex Magnetic Field Imaging of Multi-Phase Permanent Magnet Motors

Ran Zhou[1,2], Keqilao Meng[2,3*], Dajiang Jia[4]

[1] College of Energy and Power Engineering, Inner Mongolia University of Technology, Hohhot 010051, China
[2] Key Laboratory of Wind Energy and Solar Energy, Ministry of Education, Hohhot 010051, China
[3] School of Renewable Energy, Inner Mongolia University of Technology, Ordos 017010, China
[4] Shanghai Ghrepower Green Energy Co., Ltd., Shanghai 201600, China

Corresponding Author Email: mengkeqilao23@163.com

ABSTRACT

Magnetic field imaging of multi-phase permanent magnet motors serves as a fundamental tool for condition monitoring and fault diagnosis. Conventional image processing methods often result in feature representations that lack interpretability and remain disconnected from underlying physical mechanisms. Furthermore, substantial deficiencies in capturing multi-scale features and long-range dependencies within magnetic field image sequences cannot meet practical engineering requirements. To address these challenges, an interpretable feature decomposition and semantic representation method was proposed for complex magnetic field imaging in multi-phase permanent magnet motors. The core deep, collaborative dual-path architecture integrates shared encoding, dual-path decoding, physical constraints, and structural priors, enabling interpretable decomposition of physical features and precise representation of structured semantics in magnetic field images. A shared spatiotemporal encoder is designed, employing a hybrid structure of three-dimensional convolution and axial self-attention to capture multi-scale spatiotemporal features and long-range harmonic coupling effects within magnetic field sequences. An interpretable physical feature decomposition network is constructed, incorporating a physical-aware channel modulation module and spectral orthogonality regularization to explicitly separate magnetic field components with distinct spatial frequencies. A structure-guided semantic segmentation and representation network is developed, integrating a geometry-modulated attention module that fuses motor structural priors with magnetic field features to achieve precise segmentation of motor components and quantification of regional magnetic field features. Bidirectional feature collaboration is implemented to tightly couple the two decoding branches, with end-to-end optimization facilitated through multi-task joint loss functions. Experimental validation was conducted on a magnetic field image dataset comprising both healthy and faulty multi-phase permanent magnet motors. Comparative analysis against conventional and similar deep learning methods demonstrates significant advantages in interpretability of feature decomposition, semantic segmentation accuracy, and magnetic field reconstruction fidelity, thereby confirming the effectiveness and superiority of the proposed approach.

1. INTRODUCTION

Multi-phase permanent magnet motors have been widely adopted in critical sectors such as renewable energy generation, railway transportation, and high-end equipment manufacturing, owing to their advantages of high efficiency, enhanced reliability, and strong fault tolerance [1-3]. Magnetic field imaging, as a fundamental technology for characterizing the internal field distribution and reflecting the operational status of motor components [4, 5], plays a pivotal role in industrial image processing. The quality and analytical precision of these images directly determine the reliability of condition monitoring and fault diagnosis [6, 7]. Existing magnetic field image processing methods are typically designed for singular tasks, focusing either on magnetic field image reconstruction [8] or on the semantic segmentation of motor components [9]. Consequently, the simultaneous achievement of effective physical feature decoupling and precise structured semantic understanding remains elusive. Features extracted through such methods lack explicit physical significance and suffer from poor interpretability, precluding the establishment of a deep correlation between feature representations and the underlying physical mechanisms of motor magnetic fields. This limitation hinders the ability to meet practical engineering demands for accurate motor condition assessment and early fault warning under complex operating conditions.

Current research on magnetic field image processing,

feature extraction from motor magnetic fields, and interpretable deep learning within industrial image processing has witnessed certain advancements [10-13]. Conventional approaches predominantly rely on classical algorithms such as Fourier transforms and wavelet analysis for feature extraction and segmentation [10]. Deep learning-based methods have progressively enhanced processing accuracy, primarily through single-path convolutional neural network architectures [14]. However, four fundamental limitations persist, constraining the performance and practicality of magnetic field image processing. First, the spatiotemporal correlations inherent in magnetic field images are inadequately considered during feature extraction. This leads to an insufficient capacity to capture multi-scale variations across two-dimensional spatial dimensions and one-dimensional temporal or phase-sequence dimensions, particularly the long-range harmonic coupling effects spanning extensive spatial regions within the motor [15]. Second, the absence of effective physical constraints results in a "black-box" feature during magnetic field feature decomposition. The decomposed components lack explicit correspondence to physical quantities, exhibiting poor interpretability and failing to align with the intrinsic physical laws governing motor magnetic fields [16]. Third, the prior information pertaining to motor topological structure is not integrated. This decoupling between the semantic segmentation task and the physical features of the magnetic field compromises segmentation accuracy, especially at component boundaries, thereby hindering the precise discrimination of magnetic field patterns dominated by different motor parts [17, 18]. Fourth, multi-task collaboration mechanisms remain underdeveloped. The two core tasks—magnetic field feature decomposition and semantic representation—are often performed independently with separate training regimens, precluding bidirectional enhancement and limiting further improvements in overall image processing performance [19].

To address the aforementioned challenges, an interpretable magnetic field feature decomposition and semantic representation method was proposed from the perspective of image processing. This method aims to achieve a deep integration of physical mechanisms, feature representation, and semantic understanding for magnetic field images. It establishes a novel paradigm for magnetic field image processing in multi-phase permanent magnet motors and promotes the development of interpretable deep learning within industrial imaging applications. Furthermore, it provides high-precision and interpretable image analysis support for motor fault diagnosis and performance optimization, demonstrating significant theoretical value and engineering potential.

The core research objective is to design an interpretable feature decomposition and semantic representation method tailored for complex magnetic field imaging in multi-phase permanent magnet motors. The primary focus lies in the development of a deep, collaborative dual-path architecture, the construction of a fusion strategy incorporating physical constraints and structural priors, the establishment of a multi-task collaborative optimization mechanism, and experimental validation with performance analysis based on real-world datasets. Through these efforts, interpretable decomposition of physical magnetic field features and precise representation of structured semantics are ultimately realized.

Based on the aforementioned research content, the core innovations of this work are summarized below. First, a hybrid shared spatiotemporal encoder structure is proposed, combining the strengths of three-dimensional convolution and axial self-attention mechanisms. This facilitates the collaborative extraction of multi-scale spatiotemporal features and long-range dependencies from magnetic field image sequences, effectively addressing the insufficient capture of long-range harmonic coupling effects. Second, a physical-aware channel modulation module and spectral orthogonality regularization mechanism are designed. Explicit decomposition of magnetic field physical features is achieved through dynamic convolution guided by frequency-domain basis functions, with orthogonality regularization ensuring the physical interpretability of the decomposed components. This overcomes the "black-box" limitation inherent in traditional feature decomposition approaches. Third, a structure-guided dual-stream semantic segmentation architecture is constructed. By introducing a geometry-modulated attention module, deep fusion between motor topological structure priors and magnetic field features is realized. This significantly enhances the accuracy of semantic segmentation for motor components, particularly optimizing segmentation performance at component boundaries. Fourth, a bidirectional feature collaboration and multi-task joint optimization strategy is established. Tight coupling between the two decoding branches—physical feature decomposition and semantic representation—is achieved through this mechanism. End-to-end collaborative optimization of the network is driven by a multi-task loss function, enabling bidirectional enhancement between the two tasks and comprehensively improving overall image processing performance.

The remainder of this study is organized below. Section 2 elaborates on the proposed interpretable feature decomposition and semantic representation method, including the overall architectural design, technical details of each core module, the bidirectional feature collaboration mechanism, and the multi-task joint optimization strategy. In Section 3, ablation studies, comparative experiments, and robustness and generalization evaluations are conducted to validate the effectiveness and superiority of the proposed method, followed by in-depth analysis and discussion of the experimental results. Section 4 provides further discussion, analyzing the core advantages, engineering application feasibility, and existing limitations of the proposed method, with future research directions outlined. Section 5 summarizes the research work, core innovations, and experimental conclusions. Among these, Section 2 (methodology) and Section 3 (experimental validation) constitute the main focus of this study.

2. PROPOSED INTERPRETABLE FEATURE DECOMPOSITION AND SEMANTIC REPRESENTATION METHOD

2.1 Overall framework of the method

An interpretable image processing method is proposed for complex magnetic field imaging in multi-phase permanent magnet motors. The core objective is to achieve interpretable decomposition of physical features and precise representation of structured semantics from magnetic field images. The overall architecture is constructed as a deep, collaborative dual-path structure, with the primary innovation lying in the deep integration of shared encoding, dual-path decoding,

physical constraints, and structural priors. This design overcomes the limitations of conventional magnetic field image processing, where physical features and semantic representations are treated separately and lack mechanistic support, thereby enabling an integrated optimization of "feature extraction-physical decomposition-semantic representation." Distinct from traditional single-task, single-path processing models, this architecture, through its dual-branch collaborative design, ensures both the interpretability and accuracy of physical magnetic field feature decomposition while enhancing the precision of semantic segmentation for motor components, aligning with the core requirements of industrial magnetic field image processing.

Multi-channel magnetic field image sequences are taken as input. Unified extraction of multi-scale spatiotemporal features is first performed through a shared spatiotemporal encoder, generating feature tensors that capture both local details and long-range dependencies. These tensors serve as the common input for two subsequent decoding branches. The two decoding branches are an interpretable physical feature decomposition network and a structure-guided semantic segmentation and representation network. The former is dedicated to the explicit decomposition of magnetic field physical components, while the latter focuses on the semantic segmentation of motor components and the quantification of regional magnetic field features. These two branches do not operate in isolation; tight coupling is achieved through a bidirectional feature collaboration mechanism, where auxiliary information is mutually provided to enhance the processing performance of each branch. End-to-end training of the entire network is adopted, driven by a multi-task joint loss function for the collaborative optimization of global parameters. The final outputs include interpretable physical

component feature maps, pixel-level semantic label maps, and regional semantic embedding vector fields. These outputs are interrelated and mutually corroborative, conforming to the superposition principle of motor magnetic field physics while meeting the precision requirements of image processing. Consequently, high-precision and interpretable image analysis foundations are provided for subsequent motor condition monitoring and fault diagnosis.

2.2 Shared spatiotemporal encoder

The shared spatiotemporal encoder serves as the core feature extraction component of the proposed method. Its design objective is to address the critical limitations of conventional magnetic field feature extraction, namely the insufficient capture of local spatiotemporal variations and the absence of long-range harmonic coupling effect modeling. A hybrid structure combining three-dimensional convolution and axial self-attention is innovatively adopted, leveraging the complementary advantages of both mechanisms to achieve collaborative extraction of multi-scale spatiotemporal features and long-range dependencies from magnetic field image sequences. High-quality and versatile feature support is thereby provided for subsequent physical feature decomposition and semantic representation. The core innovation of this encoder lies in the deeply integrated design of axial self-attention with three-dimensional convolution and the introduction of a dual-axis separated attention mechanism, specifically targeting the modeling challenges of long-range harmonic coupling in motor magnetic field images. A schematic diagram of the three-dimensional convolution-axial self-attention hybrid structure of the shared spatiotemporal encoder is presented in Figure 1.
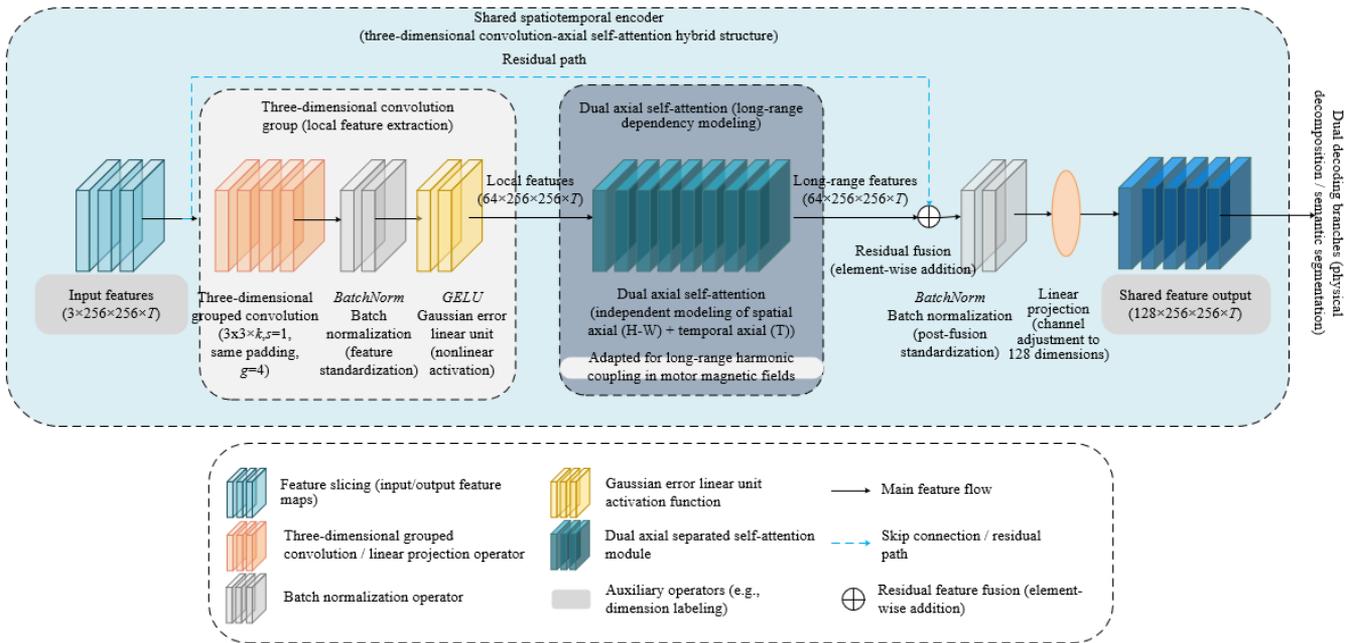
Figure 1. Architecture of the three-dimensional convolution-axial self-attention hybrid structure in the shared spatiotemporal encoder

A grouped convolution design is adopted for the three-dimensional convolutional layers of the encoder. The convolutional kernel size is set to $3 \times 3 \times k$, where k corresponds to the temporal or phase-sequence dimension. A stride of 1 is used for all dimensions, and same padding is

employed to ensure that the feature map dimensions remain consistent with those of the input magnetic field image sequences. Local variation patterns of the magnetic field across two-dimensional spatial dimensions and one-dimensional temporal or phase-sequence dimensions are thus

accurately captured, adapting to subtle features such as local magnetic field gradients and harmonic distributions. The Gaussian error linear unit activation function is selected to mitigate the vanishing gradient problem and enhance the nonlinear representation capacity of feature extraction. Following the three-dimensional convolutional layers, an axial self-attention module is embedded. A spatial-temporal dual-axis separated design is implemented to effectively reduce computational complexity while improving the accuracy of long-range dependency modeling. Spatial axial self-attention focuses on the two spatial dimensions, modeling magnetic field coupling relationships across different spatial positions. Temporal or phase-sequence axial self-attention focuses on the one-dimensional temporal or phase-sequence dimension, modeling the correlations of magnetic field variations across different time steps or phase sequences. The core computational formula is expressed as:

$$\text{Att}(X) = \text{Softmax}\left(\frac{(XW_q)(XW_k)^T}{\sqrt{d_k}}\right)(XW_v) \qquad (1)$$

where, X represents the input feature tensor, and W_q, W_k, and W_v denote the linear projection matrices for queries, keys, and values, respectively. d_k is the projection dimension, and the Softmax function is applied to normalize the attention weights. Precise modeling of long-range dependencies is achieved through this mechanism, with particular emphasis on capturing harmonic coupling effects spanning extensive spatial regions within the motor magnetic field. During the feature fusion stage, residual connections are utilized to integrate the local features extracted by three-dimensional convolution with the long-range features extracted by axial self-attention. Following batch normalization to eliminate feature distribution discrepancies and accelerate training convergence, linear projection is applied to adjust the feature channel dimensions. The final output is a multi-scale spatiotemporal feature tensor of dimensions $C \times H \times W \times T$, where C denotes the number of feature channels, H and W represent the image height and width, respectively, and T indicates the temporal or phase-sequence length. This feature tensor encompasses both local details and long-range correlations, providing a reliable feature foundation for the efficient operation of the subsequent dual decoding branches.

2.3 Interpretable physical feature decomposition network

As one of the two decoding branches, the interpretable physical feature decomposition network is designed with the core objective of achieving explicit and interpretable decomposition of physical features from magnetic field images, ensuring that the decomposed components remain consistent with the inherent physical mechanisms of the motor magnetic field. A gated multi-scale feature pyramid decoding structure is innovatively adopted in this network. Through the collaborative design of upsampling, cross-layer feature fusion, and gating mechanisms, the significant feature loss and prominent noise interference commonly encountered in traditional decoding processes are effectively mitigated, thereby establishing a foundation for precise physical feature decomposition.

A four-level gated multi-scale feature pyramid is constructed, where each decoding layer follows a sequential process of upsampling, feature fusion, and gating regulation. Upsampling is implemented using transposed convolution with a stride of 2, enabling accurate mapping of low-resolution feature maps to high-resolution feature maps and ensuring the spatial precision of the decomposition results. Feature fusion is accomplished through cross-layer skip connections, where features from corresponding levels of the shared spatiotemporal encoder are integrated with features from the current decoding layer. Spatiotemporal details from different scales are thus preserved, compensating for feature loss during upsampling. The gating mechanism employs Sigmoid gating units to dynamically adjust the weight coefficients of features at different scales. Effective information pertinent to magnetic field physical features is preferentially retained, while interference from inherent noise in magnetic field images is suppressed, accommodating the practical requirements of industrial scenarios characterized by significant noise interference in magnetic field images. A schematic diagram illustrating the internal structure and operational workflow of the physical-aware channel modulation module is presented in Figure 2.

The physical-aware channel modulation module represents the core innovation enabling interpretable decomposition of magnetic field physical features. Embedded after feature fusion in each decoding layer, this module consists of two components: a frequency-domain basis function generator and a dynamic convolutional layer. Explicit guidance for the network to separate magnetic field physical components with distinct spatial frequencies is provided through this design. The frequency-domain basis function generator is responsible for generating K sets of trainable frequency-domain basis functions. Each set of basis functions corresponds to the frequency band features of a specific magnetic field physical component: fundamental magnetic fields are associated with low-frequency basis functions, space harmonics with mid-frequency basis functions, and cogging harmonics with high-frequency basis functions. Gaussian frequency-domain functions are selected as the basis function form, with their parameters dynamically optimized through network training to ensure precise alignment with the inherent frequency-domain features of each physical component of the motor magnetic field. The convolutional kernel weights of the dynamic convolutional layer are generated through a linear combination of the K sets of frequency-domain basis functions, with the combination coefficients serving as trainable parameters. The core computational formula is expressed as:

$$W = \sum_{i=1}^{K} \alpha_i \times \varphi_i \qquad (2)$$

where, W represents the convolutional kernel weights of the dynamic convolutional layer, α_i denotes the combination coefficient for the i-th set of frequency-domain basis functions, and φ_i represents the i-th set of frequency-domain basis functions. A convolutional kernel size of 3×3 with a stride of 1 is employed in the dynamic convolutional layer, and same padding is adopted to maintain consistent feature map dimensions. Through this design, a one-to-one correspondence between channels and physical components is established, enabling each set of feature channels to correspond to a specific magnetic field physical component. The interpretability of magnetic field feature decomposition is thereby significantly enhanced, overcoming the black-box limitation inherent in traditional feature decomposition approaches.
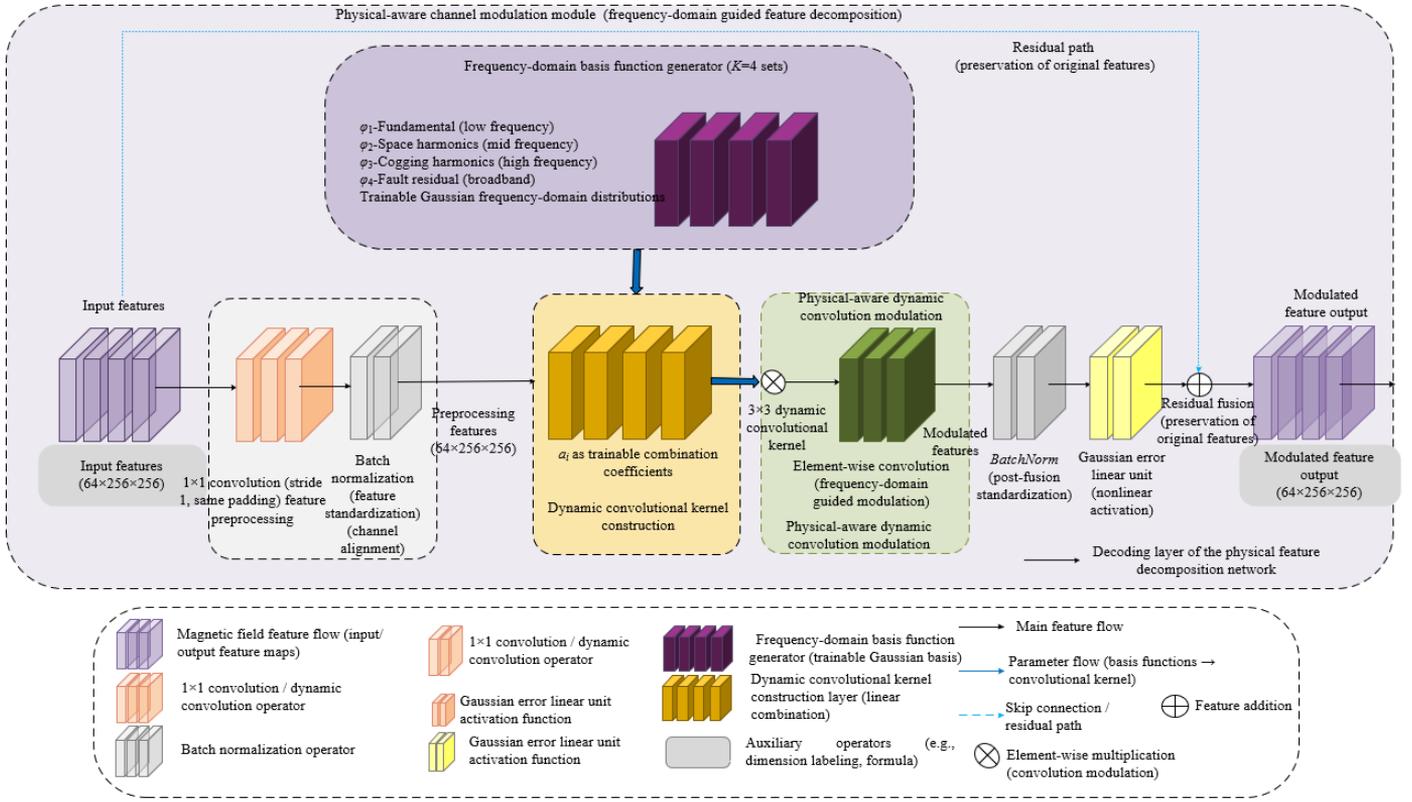
Figure 2. Schematic diagram of the internal structure and workflow of the physical-aware channel modulation module

A parallel design is adopted for the network output layer, where K independent lightweight subnetworks are established. Each subnetwork is responsible for generating a feature map corresponding to a specific magnetic field physical component, specifically including the fundamental magnetic field feature map, the v-th space harmonic feature map, the cogging harmonic feature map, and the fault residual field feature map. The dimensions of the output feature maps from all subnetworks are maintained consistent with those of the input magnetic field images, ensuring spatial completeness of the decomposition results and facilitating subsequent applications. A simple structure comprising convolution, batch normalization, and Gaussian error linear unit activation functions is employed in the lightweight subnetworks. Through parameter-lightweight design, the computational complexity of the network is effectively reduced while overfitting is prevented, ensuring extraction accuracy for each physical component feature map. Targeted capture of different types of physical components within the motor magnetic field is enabled by this parallel output design, allowing independent extraction and precise characterization of each component. Explicit physical feature support is thereby provided for subsequent motor condition monitoring and fault diagnosis.

To ensure the interpretability and physical plausibility of the magnetic field feature decomposition results, two types of physical constraint regularization terms are introduced and integrated throughout the network training process, forming a collaborative optimization mechanism between physical constraints and network learning. The first type is the spectral orthogonality regularization term. Its primary function is to enforce orthogonality, to the extent possible, between the Fourier-domain feature spectra of different physical component feature maps, thereby preventing feature confusion among distinct magnetic field physical components. The computational formula is expressed as:

$$L_{orth} = \sum_{i \neq j} |\mathrm{Cov}(F_i, F_j)| \tag{3}$$

where, L_{orth} represents the spectral orthogonality regularization loss, F_i and F_j denote the Fourier-domain feature spectra of the i-th and j-th physical component feature maps, respectively, and $Cov(\ )$ represents the covariance calculation function. By minimizing this loss term, the correlation between the feature spectra of different physical components is driven toward zero. The second type is the physical field reconstruction loss term. Designed based on the superposition principle of magnetic physical fields, an L1 loss function is employed to enforce minimization of the error between the sum of the K physical component feature maps and the original magnetic field image. The computational formula is expressed as:

$$L_{rec} = \left\| \sum_{i=1}^{K} F_i - I_{ori} \right\|_1 \tag{4}$$

where, L_{rec} represents the physical field reconstruction loss, F_i is the feature map of the i-th physical component, I_{ori} denotes the original magnetic field image, and $\| \ \|_1$ represents the L1 norm. Through the synergistic effect of these two regularization terms, conformity of the decomposition results with the physical laws of the motor magnetic field is ensured, while high-fidelity reconstruction of the original magnetic field image from the decomposed components is guaranteed. The interpretability and accuracy of magnetic field feature decomposition are thereby further enhanced.

2.4 Structure-guided semantic segmentation and representation network

As the other core decoding branch, the structure-guided

semantic segmentation and representation network is designed with the primary objective of integrating prior information on motor topological structure to achieve precise semantic segmentation of motor components and quantitative characterization of regional magnetic field features. Its core innovation lies in the design of a dual-stream feature interaction decoder. Through the collaborative processing of two parallel feature streams, deep integration of structural priors and magnetic field features is realized, overcoming the limitation of traditional approaches where semantic segmentation is decoupled from magnetic field physical features. A dual-stream input design is adopted in this network, where two parallel feature streams are constructed to process different types of inputs, forming complementary feature representations. The magnetic field feature stream takes the multi-scale spatiotemporal feature tensors from the shared spatiotemporal encoder as input, focusing on extracting key features such as grayscale, texture, and frequency-domain features from magnetic field images to accurately depict the magnetic field distribution properties of different motor components. The structural feature stream takes preprocessed motor topological structure maps as input. These topological structure maps comprise stator-rotor mask images and their corresponding signed distance fields. Structural feature extraction is accomplished through simple two-dimensional convolutional layers, focusing on capturing structural prior information, including the geometry, boundary locations, and spatial distributions of motor components. The geometric contours of core components such as permanent magnets, air gaps, and stator teeth are precisely delineated, providing explicit structural constraints for subsequent semantic segmentation. After independent feature extraction by the two streams, they are fed into the dual-stream feature interaction decoder for deep fusion, ensuring the synergistic effect of structural priors and magnetic field features.

The layered structure of the dual-stream feature interaction decoder is consistent with that of the interpretable physical feature decomposition network, both comprising four levels. Its core innovation lies in the geometry-modulated attention module embedded within each fusion level. This module is the key to achieving structure-guided optimization of magnetic field features and enhancing segmentation accuracy. The geometry-modulated attention module utilizes the structural feature output from the structural feature stream as guidance to dynamically generate spatially adaptive attention weight maps. The core computational procedure and formulas are described below. First, global average pooling and linear projection are applied to the structural features, yielding a structure guidance vector with unified dimensions. Subsequently, this structure guidance vector is multiplied channel-wise with the magnetic field features output from the magnetic field feature stream, generating an attention weight map with values controlled between 0 and 1. Finally, the attention weight map is multiplied pixel-wise with the magnetic field features, achieving dynamic modulation of the magnetic field features. The magnetic field feature responses at component boundaries and core regions are thereby selectively enhanced, while feature interference from background and irrelevant regions is effectively suppressed. The core computational formula is expressed as:

$$A=\sigma(Linear(GAP(S))\odot M), \quad M'=A\odot M \qquad (5)$$

where, A represents the attention weight map, σ denotes the

Sigmoid activation function, $Linear$ represents the linear projection operation, GAP denotes the global average pooling operation, S represents the structural feature output from the structural feature stream, M represents the magnetic field features output from the magnetic field feature stream, $\odot$ denotes the element-wise multiplication operation, and M' represents the modulated and optimized magnetic field features. Distinguished from traditional attention mechanisms that rely solely on image features, this module guides attention allocation through structural prior information. The segmentation ambiguity caused by significant variations in magnetic field features within the same component type and feature overlap between different component types in those images is effectively addressed. Notably, segmentation accuracy at component boundaries is significantly improved, meeting the core requirements of boundary segmentation in magnetic field image processing.

A dual-output design is adopted for the decoder output layer, accommodating both the requirements of semantic segmentation accuracy and the quantification of regional magnetic field features, thereby addressing the dual demands of image processing and electrical motor engineering applications. The first output type is the pixel-level semantic label map. Classification of the final fused features is performed using the Softmax activation function. Label categories are defined according to the actual topological structure of the motor, including permanent magnet N-pole, permanent magnet S-pole, air gap, stator tooth, stator yoke, rotor tooth, and background. Precise semantic segmentation of core motor components is thereby achieved, providing a spatial localization foundation for subsequent regional magnetic field feature analysis.

The second output type is the regional semantic embedding vector field. The innovation lies in achieving quantitative characterization of regional magnetic field features through the following process: for each region corresponding to a specific semantic label, a regional average pooling operator is applied to aggregate the optimized magnetic field features within that region. Global information of the magnetic field features within the region is extracted, and a semantic embedding vector of fixed dimension is output. The vector dimension is set to 128 to balance representational capacity and computational efficiency. This semantic embedding vector quantitatively describes the core features of the magnetic field in the corresponding region, including statistical properties and spectral properties. Statistical properties encompass the mean and variance of the magnetic field strength, while spectral properties include the main frequency band distribution and harmonic amplitudes of the magnetic field. Precise quantification and standardized representation of regional magnetic field features are thus realized, providing quantifiable and comparable feature support for subsequent motor condition assessment and fault diagnosis. A bridge between image processing and electrical motor engineering applications is thereby established. The feature fusion and weight generation mechanism of the geometry-modulated attention module is illustrated in Figure 3.

2.5 Bidirectional feature collaboration mechanism and multi-task joint optimization

The bidirectional feature collaboration mechanism serves as the core component for achieving tight coupling between the

interpretable physical feature decomposition network and the structure-guided semantic segmentation and representation network. Its innovation lies in establishing a collaborative mode of forward and reverse bidirectional interaction, overcoming the limitation of independently operating dual decoding branches and realizing bidirectional enhancement between physical feature decomposition and semantic representation. Overall image processing performance is thereby significantly improved. Through bidirectional feature transmission and constraint, a deep correlation between physical features and structural semantics is established, enabling the two branches to assist each other and undergo collaborative optimization.
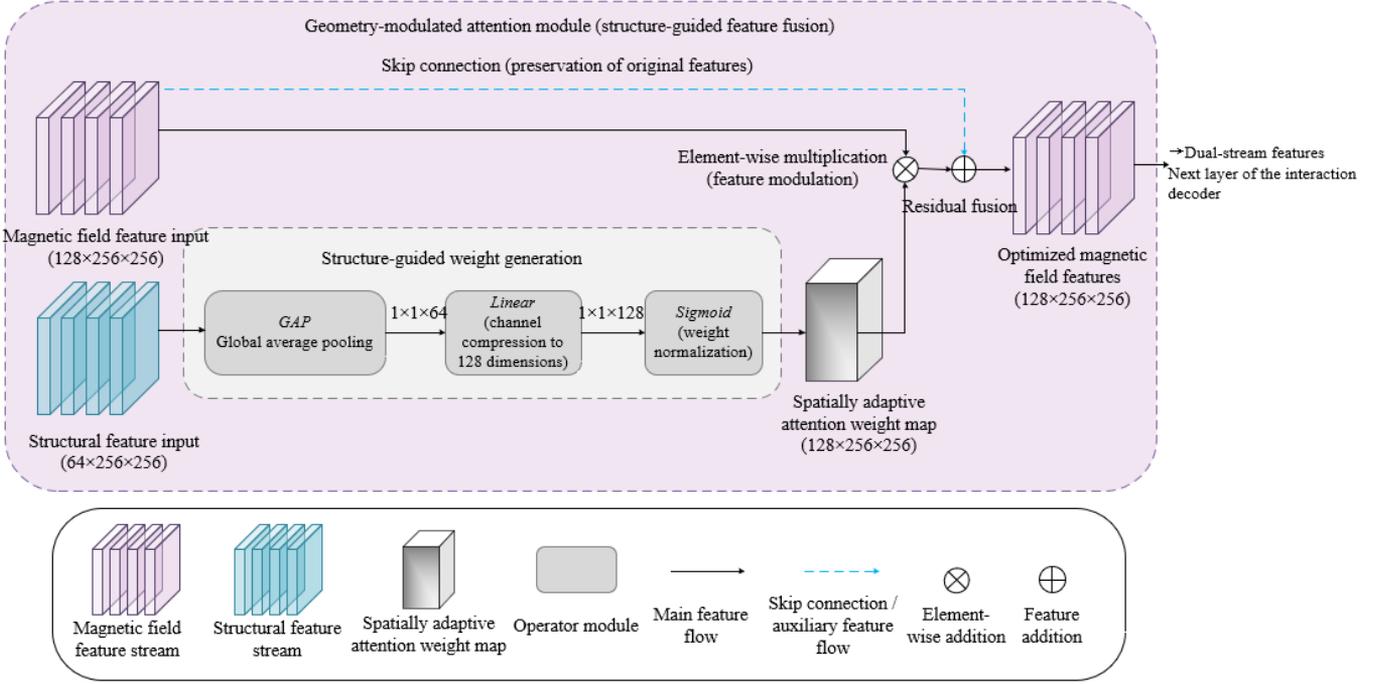
Figure 3. Feature fusion and weight generation mechanism of the geometry-modulated attention module

In the forward collaboration process, the preliminary harmonic feature maps output by the physical feature decomposition branch are projected through linear projection to adjust the channel dimensions before being precisely injected into the second and third decoding levels of the semantic segmentation branch. Explicit support from magnetic field physical features is thereby provided for the semantic segmentation task, assisting the network in rapidly identifying magnetic field patterns dominated by different motor components. The semantic segmentation results are consequently made more consistent with the physical laws of the motor magnetic field, enhancing the plausibility and accuracy of segmentation. In the reverse collaboration process, the semantic label maps output by the semantic segmentation branch are processed through the Softmax activation function to generate soft masks. Each pixel value in these masks corresponds to the probability that the pixel belongs to a specific motor component. These soft masks are fed back to each decoding level of the physical feature decomposition branch, where pixel-wise multiplication is performed with the magnetic field features of the current decoding level. A spatially adaptive constraint mechanism is thereby formed, guiding the network to optimize the extraction accuracy of corresponding physical components specifically within regions of particular motor components. Precise decomposition of magnetic field physical components across different component regions is thus achieved.

A unified optimization objective for end-to-end network training is provided by the multi-task joint loss function. Its innovation lies in the integration of four types of loss terms,

through which a reasonable balance is achieved among the training weights of the physical feature decomposition, semantic segmentation, and regional semantic representation tasks. Domination of the training process by any single task is prevented, ensuring synchronized optimization of performance across all tasks. The expression for the total loss function is given as:

$$L_{total}=\lambda_1 \times L_{rec}+\lambda_2 \times L_{orth}+\lambda_3 \times L_{seg}+\lambda_4 \times L_{emb} \qquad (6)$$

where, λ_1 through λ_4 represent the weight coefficients for each loss term, determined through 5-fold cross-validation to ensure synergistic progression of all tasks. L_{rec} denotes the physical field reconstruction loss, implemented using the L1 loss function to constrain the error between the sum of the physical component feature maps and the original magnetic field image, thereby ensuring high fidelity of physical feature decomposition. λ_1 is set to 1.0. L_{orth} denotes the spectral orthogonality regularization loss, implemented using the mean squared error loss to enforce orthogonality among the feature spectra of different physical components, ensuring the interpretability of the decomposition results. λ_2 is set to 0.5. L_{seg} denotes the semantic segmentation loss, combining cross-entropy loss and Dice loss to effectively address the issue of semantic label imbalance in magnetic field images and enhance the accuracy of motor component semantic segmentation. λ_3 is set to 1.5. L_{emb} denotes the semantic embedding contrastive loss, computed by measuring the similarity of semantic embedding vectors within regions of the same component and across regions of different components.

Consistency of magnetic field features within the same component and distinctiveness of features across different components are enforced, ensuring the effectiveness of regional semantic representation. λ_4 is set to 0.3. Through the synergistic effect of these four loss terms, the network simultaneously prioritizes the interpretability and fidelity of physical feature decomposition alongside the accuracy of semantic segmentation and feature representation during training, ultimately achieving optimal overall performance. A schematic diagram of the forward and reverse feature transmission and constraint within the bidirectional feature collaboration mechanism is presented in Figure 4.
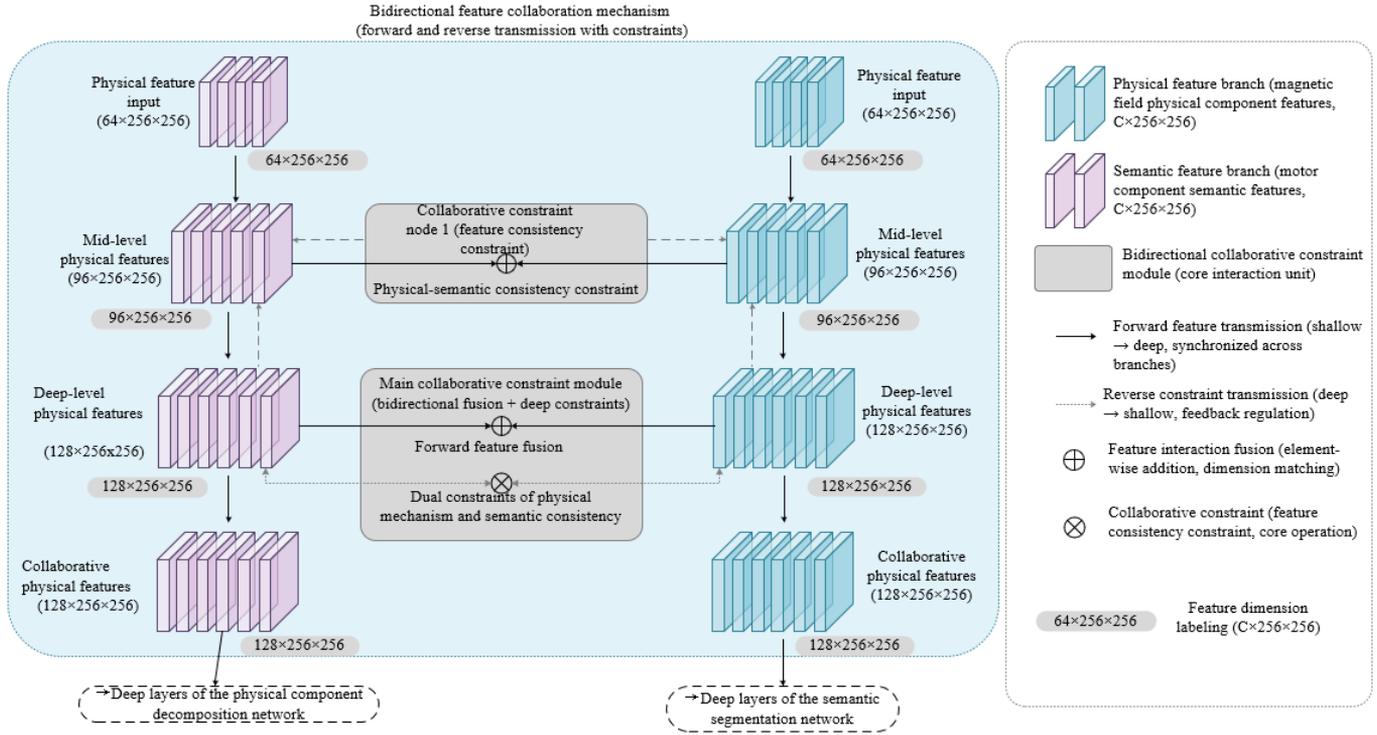
Figure 4. Schematic diagram of forward and reverse feature transmission and constraint in the bidirectional feature collaboration mechanism

2.6 Network training details

To ensure stable training and optimal performance of the proposed deep, collaborative dual-path network, a targeted training strategy is formulated, considering the features of magnetic field image processing and the requirements of multi-task collaborative optimization. All training processes are systematically conducted within a unified hardware and software environment, with detailed design aspects closely adapted to the structural features and task requirements of the proposed method, as specified below. In the dataset preprocessing stage, the issues of scale inconsistency and insufficient data volume in magnetic field image sequences are primarily addressed. First, normalization is applied to all magnetic field image sequences, mapping pixel values to the [0, 1] interval to eliminate dimensional influences and accelerate network convergence. The normalization formula is expressed as:

$$I_{norm} = \frac{I - I_{min}}{I_{max} - I_{min}} \qquad (7)$$

where, I represents the original pixel value of the magnetic field image, I_{min} and I_{max} denote the minimum and maximum pixel values of a single image, respectively, and I_{norm} represents the normalized pixel value. Subsequently, a data augmentation strategy is implemented, including random horizontal flipping, random rotation within $\pm 15°$, and the addition of Gaussian noise with a variance of 0.01-0.03. The generalization capability of the network is thereby effectively enhanced, adapting to the diversity and noise interference features of magnetic field images in practical industrial scenarios. The complexity of multi-task collaborative training is fully considered in the training parameter settings. The AdamW optimizer is selected to mitigate gradient vanishing and overfitting problems. The initial learning rate is set to 1e-4, and the weight decay coefficient is set to 1e-5 to suppress network parameter redundancy. A cosine annealing strategy is adopted for learning rate scheduling, where periodic adjustments to the learning rate prevent the training process from becoming trapped in local optima. The batch size is reasonably set to 8 based on graphics processing unit memory capacity, and the total number of training epochs is set to 100. An early stopping strategy is introduced, where training is terminated when the total loss on the validation set fails to decrease for 10 consecutive epochs, further preventing overfitting and ensuring that the network learns feature representations with stronger generalization capability.

The training hardware environment utilizes an NVIDIA RTX 3090 graphics processing unit (24GB memory) equipped with 64GB of random access memory, efficiently supporting parallel computation of multi-channel magnetic field image sequences and network parameter updates. The software environment is built on Python 3.8, with the PyTorch 1.12.0 deep learning framework employed for network construction and training. OpenCV 4.6.0 is utilized for image preprocessing and data reading, while NumPy 1.24.3 is used for numerical computations. The efficiency and stability of the training

process are thereby ensured, providing reliable support for the multi-task collaborative optimization of the network and the functional realization of each innovative module.

3. EXPERIMENTS AND RESULTS ANALYSIS

3.1 Experimental dataset and setup

To comprehensively validate the effectiveness, superiority, and generalization capability of the proposed method, experiments were conducted using a custom-built magnetic field image dataset for multi-phase permanent magnet motors. This dataset closely aligns with practical industrial application scenarios, covering different operating conditions and fault modes, thereby exhibiting strong representativeness and credibility. The dataset comprised multi-channel magnetic field image sequences encompassing various load conditions, different rotational speeds, and typical fault states such as permanent magnet demagnetization and stator tooth wear. Corresponding motor topological structure maps, magnetic field physical component annotation maps, and motor component semantic label annotation maps were also provided, offering precise annotation support for performance evaluation of physical feature decomposition and semantic segmentation tasks. The dataset was partitioned into training, validation, and test sets in a ratio of 70%, 15%, and 15%, respectively. The training set was used for network parameter training, the validation set was utilized for parameter adjustment and early stopping determination during training, and the test set was employed for final performance evaluation. Detailed statistical information of the dataset is as follows: image dimensions are uniformly 256 × 256 pixels, the number of channels is 3, the total number of samples is 1200 sets, and the number of semantic label categories is 7, including permanent magnet *N*-pole, permanent magnet *S*-pole, air gap, stator tooth, stator yoke, rotor tooth, and background, thereby ensuring comprehensiveness and objectivity of experimental evaluation.

Mainstream methods in the fields of image processing and motor magnetic field processing were selected as comparative methods, categorized into three types to ensure comprehensive and comparable evaluation, adhering to the assessment standards of image processing-related science citation index journals. The first category comprised traditional magnetic field feature extraction and segmentation methods, including wavelet transform combined with support vector machine segmentation, and Fourier decomposition combined with U-Net segmentation. The second category encompassed deep learning-based image decomposition methods, including traditional U-Net, residual UNet, and masked autoencoder decomposition models. The third category included deep learning methods incorporating physical constraints or structural priors, such as physics-constrained U-Net and structure-guided semantic segmentation models. Evaluation metrics were established based on the core objectives of the proposed method and assessment standards in the image processing field, divided into two categories of core metrics and one category of auxiliary metrics.

Physical feature decomposition evaluation metrics included reconstruction accuracy metrics, i.e., peak signal-to-noise ratio and structural similarity, for assessing the fidelity of magnetic field decomposition and reconstruction, as well as feature spectrum orthogonality and physical component error for

evaluating the interpretability of decomposition results. Semantic segmentation and representation evaluation metrics included segmentation accuracy metrics, i.e., mean intersection over union, pixel accuracy, and F1-score for assessing the precision of motor component segmentation, along with intra-component semantic embedding similarity and inter-component semantic embedding similarity for evaluating the effectiveness of regional semantic representation. The auxiliary metric was inference speed, i.e., frames per second, for validating the real-time capability of the method, adapting to the requirements of practical industrial application scenarios. All evaluation metrics were computed using domain-standard calculation methods to ensure comparability and rigor of experimental results.

3.2 Ablation studies

The primary objective of the ablation studies is to validate the necessity and effectiveness of each innovative module within the proposed deep, collaborative dual-path architecture. Five ablation models were constructed by sequentially removing core innovative modules from the complete model. Performance comparisons between these ablation models and the complete model were conducted under identical experimental settings. All models were trained with the same parameters and evaluated using the same metrics, with analysis focused on the contribution of each module to magnetic field feature decomposition accuracy, semantic segmentation accuracy, and interpretability from an image processing perspective. The experimental results are presented in Table 1.

Based on the data presented in Table 1, along with the loss curves, peak signal-to-noise ratio/structural similarity curves, and mean intersection over union curves during the training process, an analysis of the performance differences among the ablation models and the contributions of the innovative modules is provided below. Optimal performance across all evaluation metrics is achieved by the complete model, validating the effectiveness of the synergistic interaction among all innovative modules. In Model 1, where the axial self-attention module is removed, peak signal-to-noise ratio and structural similarity decrease by 2.55 dB and 0.037, respectively, and mean intersection over union decreases by 4.5 percentage points. This indicates that the axial self-attention module effectively captures long-range dependencies in magnetic field image sequences, compensating for the limitations of three-dimensional convolution in local feature extraction. Its role is particularly significant in capturing long-range harmonic coupling effects within the motor magnetic field, thereby enhancing the completeness and accuracy of magnetic field feature extraction.

In Model 2, where the physical-aware channel modulation module is removed, feature spectrum orthogonality decreases substantially by 13.7 percentage points, physical component error increases by 76.2%, and mean intersection over union decreases by 3.3 percentage points. This demonstrates that the module effectively guides the network in separating magnetic field physical components with distinct spatial frequencies, establishing a one-to-one correspondence between channels and physical components. It constitutes the core component for enhancing the interpretability of magnetic field feature decomposition while also indirectly contributing to improved semantic segmentation accuracy. In Model 3, where spectral

orthogonality regularization is removed, feature spectrum orthogonality decreases by 21.9 percentage points and physical component error increases by 100%, while peak signal-to-noise ratio and structural similarity exhibit only minor degradation. This indicates that the primary role of spectral orthogonality regularization is to ensure the interpretability of physical feature decomposition by enforcing orthogonality among the feature spectra of different physical components, thereby preventing feature confusion. Its contribution is critical to the physical plausibility of the decomposition results.

Table 1. Performance comparison results of ablation studies

Model Type	Peak Signal-to-Noise Ratio (dB)	Structural Similarity	Feature Spectrum Orthogonality	Physical Component Error
Complete model	32.67	0.958	0.923	0.021
Model 1 (without the axial self-attention)	30.12	0.921	0.918	0.023
Model 2 (without the physical-aware channel modulation)	31.05	0.935	0.786	0.038
Model 3 (without the spectral orthogonality regularization)	32.43	0.955	0.721	0.042
Model 4 (without the geometry-modulated attention)	32.21	0.952	0.920	0.022
Model 5 (without the bidirectional feature collaboration)	30.89	0.931	0.897	0.030
Model Type	**Mean Intersection over Union**	**Pixel Accuracy**	**F1-Score**	**Frames per Second**
Complete model	0.912	0.945	0.928	45.3
Model 1 (without the axial self-attention)	0.867	0.903	0.885	48.7
Model 2 (without the physical-aware channel modulation)	0.879	0.915	0.897	47.2
Model 3 (without the spectral orthogonality regularization)	0.883	0.918	0.901	46.1
Model 4 (without the geometry-modulated attention)	0.854	0.896	0.876	46.8
Model 5 (without the bidirectional feature collaboration)	0.848	0.889	0.870	49.5

In Model 4, where the geometry-modulated attention module is removed, mean intersection over union decreases by 5.8 percentage points, while pixel accuracy and F1-score decrease by 4.9 and 5.2 percentage points, respectively. Changes in metrics related to physical feature decomposition are relatively minor. This suggests that the module effectively integrates motor structural priors with magnetic field features, optimizing attention allocation and significantly enhancing the accuracy of semantic segmentation for motor components. Segmentation performance at component boundaries is particularly improved, addressing the challenges posed by large intra-class feature variations and inter-class feature overlap in magnetic field images. In Model 5, where the bidirectional feature collaboration mechanism is removed, all core metrics exhibit notable degradation, with peak signal-to-noise ratio decreasing by 1.78 dB and mean intersection over union decreasing by 6.4 percentage points. This indicates that the bidirectional feature collaboration mechanism achieves tight coupling between the two decoding branches, enabling bidirectional enhancement between physical feature decomposition and semantic representation. Performance bottlenecks resulting from independent training of the two branches are thereby avoided, making this mechanism the key to improving overall image processing performance. In conclusion, each innovative module proposed in this work is demonstrated to be indispensable. Their synergistic interaction significantly enhances the comprehensive performance of the method.

3.3 Comparative experiments and results analysis

For the comparative experiments, seven mainstream methods from the fields of image processing and motor magnetic field processing were selected and evaluated against the proposed method on the same test set. These methods were categorized into three types to ensure comprehensive and comparable evaluation. The advanced nature and superiority of the proposed method were validated through a combination of quantitative and qualitative analysis.

Quantitative analysis was conducted using all evaluation metrics defined in this study to compare the performance differences among the methods. The advantages of the proposed method in terms of magnetic field feature decomposition fidelity, semantic segmentation accuracy, and interpretability were analyzed in focus. The experimental data are presented in Figure 5 and Table 2. The methods in the

figure correspond sequentially to Method 1 (wavelet transform + support vector machine segmentation), Method 2 (Fourier decomposition + U-Net segmentation), Method 3 (the traditional U-Net), Method 4 (residual UNet), Method 5 (the masked autoencoder decomposition model), Method 6 (physics-constrained U-Net), and Method 7 (the structure-guided segmentation model).

As evidenced by the data in Figure 5 and Table 2, the proposed method achieves significantly superior performance across all evaluation metrics compared to various categories of competing methods, fully demonstrating its advanced nature. Regarding metrics related to magnetic field feature decomposition, a peak signal-to-noise ratio of 32.67 dB is attained by the proposed method, representing an improvement of 1.11 dB over the best competing method (structure-guided segmentation model) and 7.33 dB over the traditional method (wavelet transform + support vector machine segmentation). A structural similarity of 0.958 is achieved, which is 0.013 higher than the best competing method and 0.091 higher than traditional methods. These results indicate a clear advantage of the proposed method in the fidelity of magnetic field feature decomposition and reconstruction. A feature spectrum orthogonality of 0.923 is obtained, surpassing the best competing method by 7.0 percentage points and traditional methods by 31.1 percentage points. The physical component error is merely 0.021, representing a 30.0% reduction compared to the best competing method and a 75.9% reduction compared to traditional methods. This fully validates the core advantage of the proposed method in achieving interpretable decomposition of magnetic field physical features, enabling effective decoupling of different physical components with results conforming to the physical laws of motor magnetic fields.
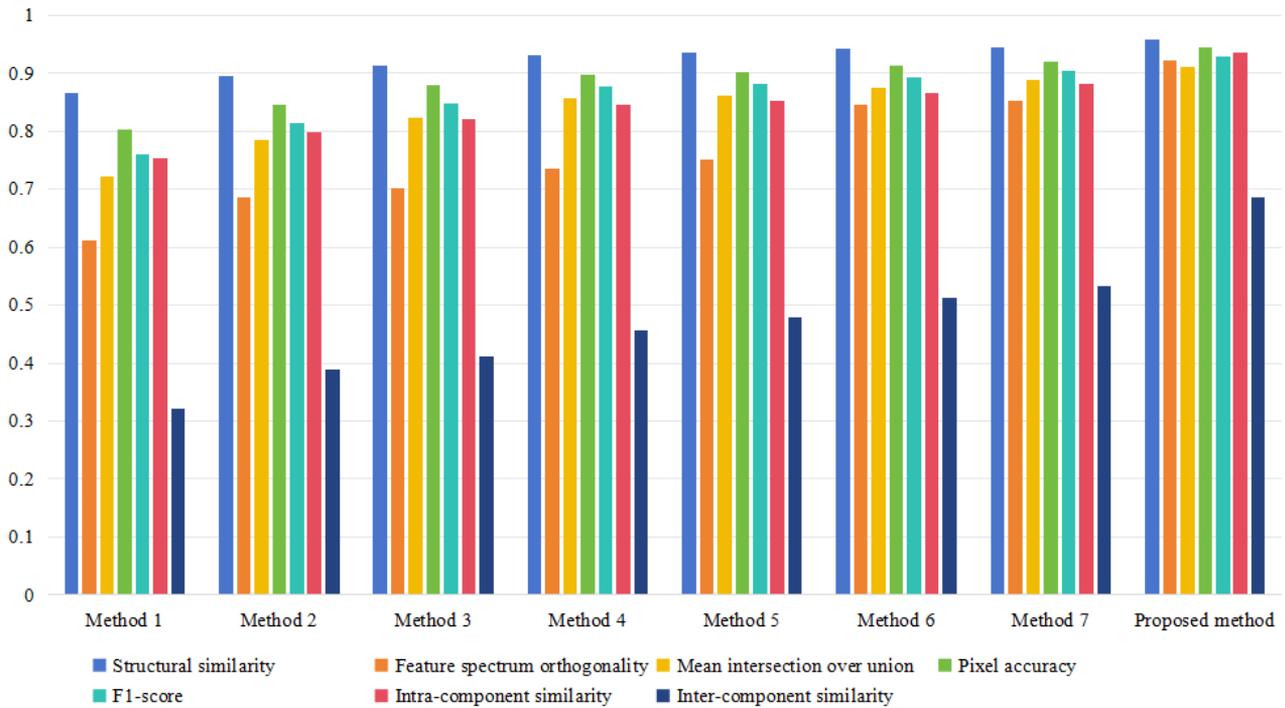

Figure 5. Performance comparison results of comparative experiments

Table 2. Comparison results of the peak signal-to-noise ratio, physical component error, and frames per second

Metric	Wavelet Transform + Support Vector Machine Segmentation	Fourier Decomposition + U-Net Segmentation	Traditional U-Net	Residual UNet
Peak signal-to-noise ratio (dB)	25.34	27.89	29.15	30.52
Physical component error	0.087	0.065	0.058	0.043
Frames per second	52.6	49.8	47.5	44.2
Metric	Masked Autoencoder Decomposition Model	Physics-Constrained U-Net	Structure-Guided Segmentation Model	Proposed Method
Peak signal-to-noise ratio (dB)	30.87	31.23	31.56	32.67
Physical component error	0.040	0.032	0.030	0.021
Frames per second	38.9	41.7	40.3	45.3

In terms of semantic segmentation metrics, a mean intersection over union of 0.912 is achieved by the proposed method, which is 2.3 percentage points higher than the best competing method (structure-guided segmentation model) and 19.1 percentage points higher than traditional methods. Pixel accuracy and F1-score reach 0.945 and 0.928, respectively, both significantly exceeding those of all comparative methods. This demonstrates that precise semantic segmentation of motor components is realized by the proposed method, with particular advantages in segmentation accuracy at component boundaries. Concerning semantic embedding representation metrics, an intra-component similarity of 0.935 and an inter-component similarity of 0.687 are attained, showing substantial improvements over all comparative methods. This indicates that accurate quantitative characterization of regional magnetic field features is achieved, ensuring consistency of magnetic field features within the same component and distinctiveness across different components. Regarding inference speed, 45.3 frames per second are achieved by the proposed method, which is higher than residual UNet, the masked autoencoder decomposition model, physics-constrained U-Net, and the structure-guided segmentation model, and only lower than traditional methods. A balance between performance and real-time capability is thus maintained, meeting the requirements of practical industrial application scenarios. In conclusion, through the synergistic interaction of its innovative modules, the proposed method effectively breaks through the performance bottlenecks of traditional approaches and existing deep learning methods, demonstrating significant superiority across all core tasks of magnetic field image processing.

To further visually substantiate the advantages of the proposed method, magnetic field images from three typical scenarios within the test set—normal operating condition, permanent magnet demagnetization fault, and stator tooth wear fault—were selected. A comparative analysis of the output results from the proposed method and various competing methods was conducted, with a focus on examining differences in the physical component decomposition maps and semantic segmentation maps. Performance comparisons of magnetic field component decomposition and semantic segmentation are presented in Figure 6.

Under normal operating conditions, the proposed method clearly separates the fundamental magnetic field, *v*-th order space harmonics, cogging harmonics, and the background field. Each physical component exhibits clear boundaries without confusion. In the semantic segmentation maps, the contours of components such as permanent magnets, air gaps, and stator teeth are complete with precise boundaries, enabling accurate differentiation between the *N*-pole and *S*-pole of the permanent magnets. In contrast, the traditional method (Fourier decomposition + U-Net segmentation) only achieves a rough separation of fundamental and harmonic components, failing to effectively distinguish cogging harmonics from the background field. Significant omissions and misclassifications are observed in the semantic segmentation results. Residual UNet achieves relatively good magnetic field reconstruction; however, the physical component decomposition lacks interpretability, with blurred boundaries between components. Semantic segmentation exhibits blurring issues at the boundaries between the air gap and stator teeth. The structure-guided segmentation model demonstrates relatively good semantic segmentation performance, but the accuracy of physical component decomposition is insufficient, preventing

precise separation of different harmonic components.

(a)

(b)

(c)
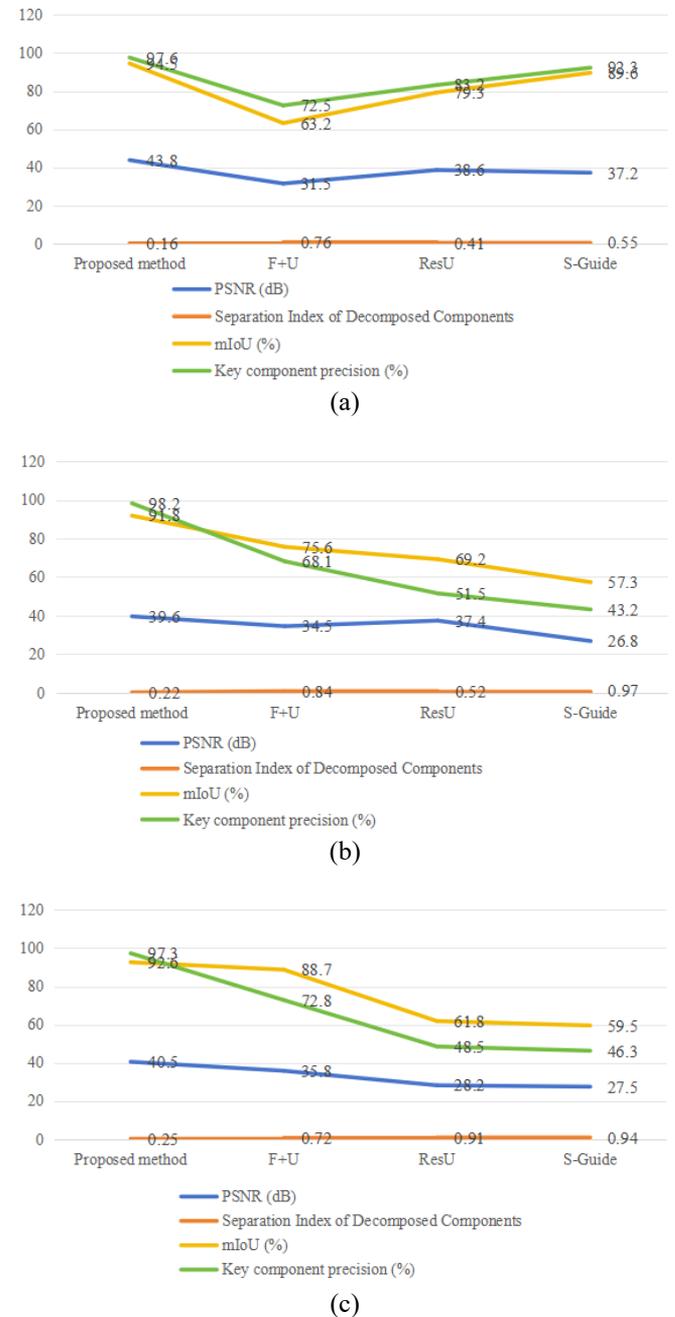
Figure 6. Quantitative performance comparison of magnetic field component decomposition and semantic segmentation

Under the permanent magnet demagnetization fault condition, the proposed method accurately separates the fault residual field, clearly presenting the abnormal magnetic field distribution in the demagnetized region. Simultaneously, the demagnetized permanent magnet components are correctly segmented, exhibiting high correspondence between semantic labels and the fault region. Among the comparative methods, physics-constrained U-Net captures the abnormal magnetic field distribution but fails to effectively separate the fault residual field from other harmonic components, making fault type identification difficult. The masked autoencoder decomposition model achieves relatively good magnetic field reconstruction; however, semantic segmentation accuracy in the fault region is low, preventing precise localization of the demagnetized area. Traditional methods are unable to effectively capture the subtle magnetic field variations induced

by the fault, resulting in significant errors in both physical component decomposition and semantic segmentation.

Under the stator tooth wear fault condition, the proposed method accurately separates the anomalous components of the cogging harmonics, clearly presenting the magnetic field distortion corresponding to the worn stator tooth region. In the semantic segmentation maps, the worn stator tooth is correctly identified with precise boundary segmentation. Among the various comparative methods, the structure-guided segmentation model approximately locates the stator tooth region but fails to capture the anomalous variations in the cogging harmonics. Both residual UNet and traditional methods are unable to effectively distinguish the magnetic field differences between normal and worn regions, rendering the physical component decomposition and semantic

segmentation inadequate for fault diagnosis requirements.

### 3.4 Robustness experiments and generalization experiments

Robustness experiments were conducted to verify the adaptability of the proposed method under varying noise interference and different image resolutions, a performance indicator of significant interest to image processing journals. The peak signal-to-noise ratio, structural similarity, mean intersection over union, and frames per second were selected as core evaluation metrics to analyze performance variations under different interference conditions. The experimental data are presented in Table 3.

Table 3. Robustness experiment (noise interference) performance comparison results

Noise Type	Interference Intensity	Peak Signal-to-Noise Ratio (dB)	Structural Similarity	Mean Intersection over Union	Frames per Second
No noise	—	32.67	0.958	0.912	45.3
Gaussian noise	Variance=0.01	31.89	0.947	0.898	45.1
	Variance=0.02	30.76	0.932	0.885	45.0
	Variance=0.03	29.53	0.915	0.871	44.8
Salt-and-pepper noise	Density=0.05	31.72	0.945	0.895	44.9
	Density=0.10	30.25	0.926	0.878	44.7
	Density=0.15	28.96	0.903	0.862	44.6
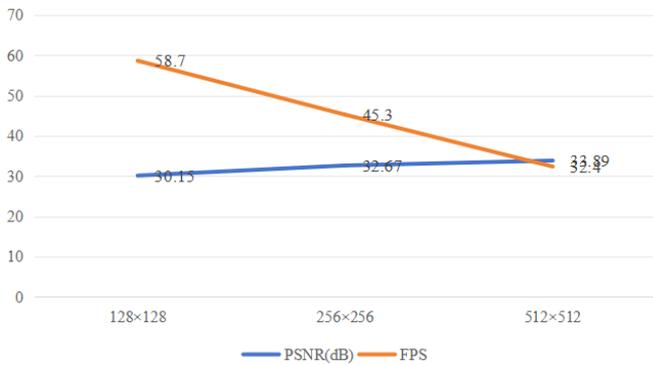
(a) Comparison results of the peak signal-to-noise ratio and frames per second
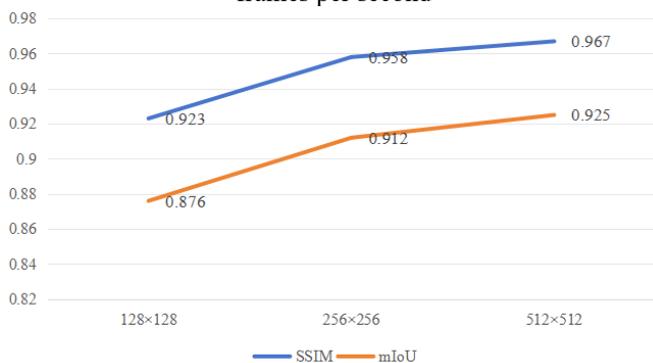
(b) Comparison results of structural similarity and the mean intersection over union

Figure 7. Robustness experiment (resolution variation) performance comparison results

From the noise interference robustness experimental data presented in Table 3, it is observed that as the Gaussian noise variance and salt-and-pepper noise density increase, all performance metrics of the proposed method exhibit a gradual declining trend, albeit with relatively small decreases,

demonstrating strong noise resilience. When the Gaussian noise variance reaches 0.03, the peak signal-to-noise ratio, structural similarity, and mean intersection over union are 29.53 dB, 0.915, and 0.871, respectively, representing declines of only 3.14 dB, 0.043, and 4.1 percentage points compared to the no-noise scenario. When the salt-and-pepper noise density reaches 0.15, the peak signal-to-noise ratio, structural similarity, and mean intersection over union are 28.96 dB, 0.903, and 0.862, respectively, representing declines of 3.71 dB, 0.055, and 5.0 percentage points compared to the no-noise scenario, with the frames per second remaining largely stable. These results indicate that the gating mechanism and physical constraint regularization within the proposed method effectively suppress noise interference, mitigating the impact of noise on magnetic field feature extraction, physical component decomposition, and semantic segmentation. Relatively high processing accuracy is thereby maintained even under noisy conditions, addressing the challenge of noise in magnetic field images within practical industrial scenarios.

From the resolution variation robustness experimental data illustrated in Figure 7, it is observed that as image resolution increases, the peak signal-to-noise ratio, structural similarity, and mean intersection over union of the proposed method exhibit an upward trend, while the frames per second show a declining trend, consistent with fundamental principles of image processing. When the resolution is increased from 128 × 128 to 512 × 512, the peak signal-to-noise ratio improves by 3.74 dB, structural similarity improves by 0.044, and the mean intersection over union improves by 4.9 percentage points. This indicates that the proposed method adapts well to magnetic field images of different resolutions, capturing richer magnetic field details and component structural information at higher resolutions, thereby enhancing processing accuracy. At lower resolution (128 × 128), the mean intersection over union still reaches 0.876, the peak signal-to-noise ratio reaches 30.15 dB, maintaining relatively high processing performance, while the frames per second increase to 58.7, offering superior real-

time capability. In summary, the proposed method exhibits strong adaptability to variations in image resolution, accommodating industrial scenarios with different resolution requirements while balancing processing accuracy and real-time performance.

Generalization experiments were conducted to validate the general applicability of the proposed method. The method was applied to another magnetic field image dataset of different multi-phase permanent magnet motor models not involved in training. This dataset differs from the training dataset in terms of motor model and structural parameters, encompassing various load conditions, different rotational speeds, and typical fault scenarios, comprising a total of 300 samples. The image dimensions are 256 × 256, the number of channels is 3, and the semantic label categories are consistent with those of the training dataset. Core performance metrics of the proposed method on the original dataset test set and the new motor model dataset were compared to validate the generalization capability. The experimental data are presented in the following table.

From the generalization experimental data presented in Table 4, it is observed that all performance metrics of the proposed method on the new motor model dataset exhibit slight decreases compared to those on the original dataset test set. However, the magnitude of decrease is minimal, and all metrics remain at relatively high levels. Specifically, the peak

signal-to-noise ratio is 31.92 dB, a decrease of only 0.75 dB compared to the original dataset; structural similarity is 0.949, a decrease of 0.009; feature spectrum orthogonality is 0.915, a decrease of 0.008; the physical component error is 0.024, an increase of 0.003; the mean intersection over union is 0.897, a decrease of 1.5 percentage points; pixel accuracy and F1-score decrease by 1.3 and 1.5 percentage points, respectively; and the frames per second remain largely stable. These results indicate that the proposed method effectively adapts to magnetic field images of different multi-phase permanent magnet motor models. Interpretable decomposition of magnetic field physical features and accurate semantic segmentation of motor components are achieved without requiring retraining of network parameters, demonstrating strong generalization capability and general applicability. The fundamental reason lies in the incorporation of physical constraints of motor magnetic fields and structural prior information into the proposed method. The method captures general physical laws of motor magnetic fields and component structural features rather than features specific to particular motor models. Consequently, effective transfer to magnetic field image processing tasks for different motor models is enabled, meeting the application requirements of multi-model motors in industrial scenarios and further enhancing the engineering application value of the method.

Table 4. Generalization experiment performance comparison results

Dataset Type	Peak Signal-to-Noise Ratio (dB)	Structural Similarity	Feature Spectrum Orthogonality	Physical Component Error
Original dataset test set	32.67	0.958	0.923	0.021
New motor model dataset	31.92	0.949	0.915	0.024
Dataset Type	Mean Intersection over Union	Pixel Accuracy	F1-Score	Frames per Second
Original dataset test set	0.912	0.945	0.928	45.3
New motor model dataset	0.897	0.932	0.913	45.1

4. DISCUSSION

The proposed interpretable feature decomposition and semantic representation method for complex magnetic field imaging in multi-phase permanent magnet motors fundamentally innovates by overcoming the disconnection among "algorithm design, physical mechanism, and engineering application" prevalent in traditional magnetic field image processing. A technical framework integrating these three aspects is constructed. The essential distinction between this method and existing interpretable image decomposition and semantic segmentation approaches lies in its philosophy: rather than merely pursuing algorithmic performance optimization, the inherent physical laws of motor magnetic fields and topological structural priors are embedded into core image processing modules. A precise alignment between technological innovation and engineering requirements is thereby achieved. Existing methods predominantly focus on the algorithms themselves or rely on image-derived features, lacking deep integration with physical mechanisms and effective utilization of structural priors, which limits their adaptability to complex industrial scenarios.

In contrast, the proposed method incorporates harmonic distribution features of motor magnetic fields and the superposition principle of physical fields into modules such as the hybrid feature extractor and the physical-aware dynamic convolution. Structural priors are embedded into the geometry-modulated attention module. This design concurrently adheres to image processing logic and conforms to magnetic field physical principles, offering a reference paradigm for imaging processing of other physical fields in industrial applications and possessing strong potential for generalization and adoption.

Physical feature decoupling and semantic representation in magnetic field images are not independent tasks but constitute an interconnected, synergistically enhancing whole. This interrelationship is fully exploited through the proposed bidirectional feature collaboration mechanism, which constitutes a core reason for the method's superiority over existing single-task or single-branch approaches. Ablation studies validate the necessity of this mechanism. Forward collaboration utilizes physically decomposed features to provide physical support for semantic segmentation, ensuring that segmentation results align with magnetic field principles.

Reverse collaboration leverages the spatial constraints from semantic segmentation to guide and enhance the accuracy of physical decomposition, preventing component confusion. This collaborative paradigm can serve as a reference for image processing of analogous industrial physical fields, which often exhibit clear physical principles and fixed structural features. By drawing upon the dual-branch collaborative concept, performance optimization can be achieved for such tasks.

The engineering application feasibility of the proposed method was validated through experiments. Inference speed at conventional resolutions meets the real-time requirements of motor condition monitoring. Robustness and generalization experiments demonstrate its adaptability to complex industrial conditions such as noise interference, resolution variations, and different motor models. The core pathway for integrating this method into a motor condition monitoring system is as follows: interface with imaging equipment for magnetic field image preprocessing, utilization of the deep, collaborative dual-path network as the core for feature decomposition, semantic segmentation, and representation, and integration with fault diagnosis and performance evaluation models to achieve real-time motor condition monitoring. This facilitates the transition of deep learning from laboratory algorithms to practical industrial applications.

Despite the significant advantages demonstrated by the proposed method, limitations remain. Inference speed is insufficient when processing high-resolution magnetic field images, and the accuracy of physical component decomposition in scenarios involving complex multi-fault coupling requires improvement. Future research will focus on optimization and refinement aligned with research hotspots in image processing. Real-time performance will be enhanced through lightweight techniques such as network pruning and parameter quantization. Long-range dependency capture capability will be strengthened by integrating Transformer architectures to adapt to complex fault scenarios. Multi-modal fusion techniques will be explored, combining multi-source motor data to enrich feature representation. These efforts will promote the widespread application of the method in industrial scenarios.

5. CONCLUSION

Addressing the core challenge in complex magnetic field imaging of multi-phase permanent magnet motors, where traditional methods struggle to achieve interpretable decoupling of physical features and precise representation of structured semantics, an interpretable magnetic field image processing approach was investigated. An interpretable feature decomposition and semantic representation method based on a deep, collaborative dual-path architecture was proposed. Through the integration of shared encoding, dual-path decoding, physical constraints, and motor structural priors, a deep unification of magnetic field physical mechanisms, feature processing, and engineering applications was realized. The core innovations and technical contributions lie in the design of a three-dimensional convolution-axial self-attention hybrid shared encoder, a physical-aware channel modulation mechanism with spectral orthogonality regularization, a structure-guided dual-stream semantic segmentation architecture, and a bidirectional feature collaboration mechanism with multi-task joint optimization. Effective breakthroughs of traditional method limitations were

achieved, enhancing the accuracy and interpretability of magnetic field image processing. A series of experiments validated the necessity of each innovative module and confirmed that the proposed method outperformed existing approaches in magnetic field reconstruction accuracy, feature decomposition interpretability, semantic segmentation precision, and inference speed. Favorable robustness and generalization were also demonstrated. Efficient and accurate technical support for magnetic field image processing in multi-phase permanent magnet motors can be thereby provided, holding significant theoretical and engineering application value in the fields of motor condition monitoring, fault diagnosis, and performance optimization. Addressing the limitations of the method concerning real-time performance in high-resolution image processing and decomposition accuracy in complex multi-fault coupling scenarios, future research will focus on lightweight network design, integration of Transformers with existing modules, and multi-modal fusion techniques to further refine method performance and promote its industrial deployment and application.

REFERENCES

[1] Gritli, Y., Rossi, C., Pilati, A., Tani, A., Casadei, D. (2025). Detection and discrimination of mixed demagnetization and high-resistance connection faults in six-phase surface-mounted PMSM drive. IEEE Transactions on Transportation Electrification, 11(3): 8594-8603. https://doi.org/10.1109/TTE.2025.3542901

[2] Vancini, L., Mengoni, M., Rizzoli, G., Zarri, L., Tani, A. (2024). Local demagnetization detection in six-phase permanent magnet synchronous machines. IEEE Transactions on Industrial Electronics, 71(6): 5508-5518. https://doi.org/10.1109/TIE.2023.3294603

[3] Sadeghi, S., Parsa, L. (2011). Multiobjective design optimization of five-phase Halbach array permanent-magnet machine. IEEE Transactions on Magnetics, 47(6): 1658-1666. https://doi.org/10.1109/TMAG.2011.2106217

[4] Li, X.N., An, N., Li, X.J., Feng, X.Q., Bai, J.T. (2013). Two-dimensional optical imaging of artificial magnetic field in the laboratory. Acta Physica Polonica A, 123(1): 34-38. http://przyrbwn.icm.edu.pl/APP/PDF/123/a123z1p08.pdf.

[5] Zhu, F., Peck, M.A., Jones-Wilson, L.L. (2020). Reduced embedded magnetic field in type-II superconductor of finite dimension. IEEE Transactions on Applied Superconductivity, 30(6): 1-5. https://doi.org/10.1109/TASC.2020.2976592

[6] Tieng, Q.M., Vegh, V. (2011). Magnetic resonance imaging in nonlinear fields with nonlinear reconstruction. Concepts in Magnetic Resonance Part B: Magnetic Resonance Engineering, 39B(3): 128-140. https://doi.org/10.1002/cmr.b.20200

[7] Frollo, I., Andris, P., Gogola, D., Pribil, J., Valkovic, L., Szomolányi, P. (2012). Magnetic field variations near weak magnetic materials studied by magnetic resonance imaging techniques. IEEE Transactions on Magnetics, 48(8): 2334-2339. https://doi.org/10.1109/TMAG.2012.2191298

[8] Knopp, T., Them, K., Kaul, M., Gdaniec, N. (2015). Joint reconstruction of non-overlapping magnetic particle

imaging focus-field data. Physics in Medicine & Biology, 60(8): L15. https://doi.org/10.1088/0031-9155/60/8/L15

[9] Sarubbo, S., De Benedictis, A., Maldonado, I.L., Basso, G., Duffau, H. (2013). Frontal terminations for the inferior fronto-occipital fascicle: Anatomical dissection, DTI study and functional considerations on a multi-component bundle. Brain Structure & Function, 218(1): 21-37. https://doi.org/10.1007/s00429-011-0372-3

[10] Foerger, F., Boberg, M., Faltinath, J., Knopp, T., Möddel, M. (2024). Design and optimization of a magnetic field generator for magnetic particle imaging with soft magnetic materials. Advanced Intelligent Systems, 6(11): 2400017. https://doi.org/10.1002/aisy.202400017

[11] Tremsin, A.S., Kardjilov, N., Strobl, M., Manke, I., Dawson, M., McPhate, J.B., Vallerga, J.V., Siegmund, O.H.W., Feller, W.B. (2015). Imaging of dynamic magnetic fields with spin-polarized neutron beams. New Journal of Physics, 17: 043047. https://doi.org/10.1088/1367-2630/17/4/043047

[12] Yang, X.S., Du, C., Zhang, R.T., Zhang, J.Y., Chen, J., Liu, Y.T., Ding, Z.W. (2025). HEFNet: A hybrid finite element and deep learning method for magnetic field prediction in electrical equipment. Electrical Engineering, 107: 15329-15342. https://doi.org/10.1007/s00202-025-03328-9

[13] Zhu, Q.Z., Xu, L. (2024). Analysis of magnetic field regulation characteristics of novel rare earth variable flux permanent magnet synchronous motor for electric vehicles. IEICE Electronics Express, 21(15): 1-5. https://globals.ieice.org/en_publications/elex/10.1587/elex.20.20230430/_f.

[14] Lin, M.Y., Lee, O.W. (2025). Design and implementation of a nondestructive testing system for magnetic field imaging based on machine learning. Journal of Magnetism and Magnetic Materials, 614: 172638. https://doi.org/10.1016/j.jmmm.2024.172638

[15] Ammari, S., Pitre-Champagnat, S., Dercle, L., Chouzenoux, E., et al. (2021). Influence of magnetic field strength on magnetic resonance imaging radiomics features in brain imaging, an in vitro and in vivo study. Frontiers in Oncology, 10: 541663. https://doi.org/10.3389/fonc.2020.541663

[16] Ma, N.J., Gao, X.D., Wang, C.Y., Zhang, Y.X. (2021). A novel detection of weld defects by magneto-optical imaging under combined magnetic field. Insight - Non-Destructive Testing and Condition Monitoring, 63(12): 704-711. https://doi.org/10.1784/insi.2021.63.12.704

[17] Chen, Y. L., Tang, J., Guo, H., Gao, Y.J., Liu, Y.S., Shi, Z.R. (2019). Wide-field planar magnetic imaging using spins in diamond. In 2019 IEEE 19th International Conference on Nanotechnology (IEEE-NANO), Macao, China, pp. 99-102. https://doi.org/10.1109/NANO46743.2019.8993679

[18] Mihara, A. (1992). X-ray topographic observations of magnetic domain-structures and 180° walls in a (100) crystal of pure iron under an external magnetic field. Japanese Journal of Applied Physics, 31(6R): 1793. https://doi.org/10.1143/JJAP.31.1793

[19] Ma, N.J., Gu, S.C., Gao, X.D., Zhang, Y.X., Mo, L. (2025). Magneto-optical imaging analysis of welding defects under vertical combined magnetic field. IEEE Sensors Journal, 25(14): 27367-27376. https://doi.org/10.1109/JSEN.2025.3573720