Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

3D Convolution Lightweight Vision Transformer to Progressive Semantic Focusing for Magnetic Property Prediction of Additively Manufactured Components

Jhih-Cheng Huang
National Cheng Kung University
Chia-Ho Chuang
National Cheng Kung University
Yin-Yang Hsiao
National Cheng Kung University
Mi-Ching Tsai
National Cheng Kung University
Ming-Huwi Horng
horng@mail.csie.ncku.edu.tw

National Cheng Kung University

Research Article

Keywords: Selective Laser Melting, Magnetic Property Prediction, 3D Vision Transformer, Progressive Sampling, Process Parameters

Posted Date: March 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8836731/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

3D Convolution Lightweight Vision Transformer to Progressive Semantic Focusing for Magnetic Property Prediction of Additively Manufactured Components

Jhih-Cheng Huang[1,a], Chia-Ho Chuang[2,b], Yin-Yang Hsiao[1,c], Mi-Ching Tsai[2,d] and Ming-Huwi Horng[1,e]*

[1] Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan 700, Taiwan

[2] Electrical Motor Technology Research Center, National Cheng Kung University, Tainan 700, Taiwan

[a]zhihcheng032544@gmail.com, [b]chiaho.chuang@gmail.com, [c]ian5051490@gmail.com, [d]mctsai@mail.ncku.edu.tw, [e]horng@mail.csie.ncku.edu.tw

Abstract

Objective/background: Predicting the magnetic performance of selective laser melting (SLM)-fabricated components is essential for quality assurance; however, the complex nonlinear dynamics of Fe-50Ni soft magnetic alloys under high-frequency excitation (400–800 Hz) remain poorly modeled by existing deep learning approaches. This study aims to overcome the significant accuracy degradation observed in conventional models when addressing these frequency-dependent hysteresis behaviors.

Method: We propose a lightweight conv-enhanced progressive sampling vision transformer (CVPSViT), which can synergize 3D spatial feature extraction with physics-informed process parameters. The architecture incorporates three methodological innovations: (1) it ingests stacked layer-wise imagery as 3D volumes to capture interlayer continuity and microdepth textures; (2) it introduces a conv-enhanced progressive sampling module (CPSM), which employs a coarse-to-fine strategy to dynamically update sampling coordinates, focusing attention on semantically discriminative regions akin to the human visual system; (3) it executes a deep cross-modal fusion by embedding critical manufacturing parameters, specifically laser power and oxygen concentration, directly into the global representation prior to inference.

Results: Extensive experiments on five key magnetic targets demonstrate that CVPSViT consistently outperforms conventional machine learning baselines and the standard CvT architecture. The model exhibits exceptional robustness in high-frequency scenarios: for coercivity (H_c) at 800 Hz, it achieves an R^2 of 0.981, significantly surpassing the 0.876 of CvT. Furthermore, for iron loss (P_{cv}), the most frequency-sensitive indicator, CVPSViT maintains a high accuracy of 0.934 compared to 0.909 for CvT. Ablation studies confirm high efficiency, with the model requiring

24% fewer parameters (38.5M) and 12% fewer GFLOPs (22.64) than the PSViT baseline.
Conclusions: This work presents a robust, computationally efficient framework for the real-time quality monitoring of additively manufactured components. By effectively balancing high-frequency prediction accuracy with low model complexity, CVPSViT offers a viable solution for intelligent manufacturing systems requiring precise feedback on material properties.

Keywords: Selective Laser Melting, Magnetic Property Prediction, 3D Vision Transformer, Progressive Sampling, Process Parameters

1. INTRODUCTION

As the manufacturing industry continues to advance toward high efficiency and high customization, metal additive manufacturing (AM) technology has received considerable attention due to its design flexibility and rapid prototyping capabilities **[1]**. Compared to conventional fabrication methods (such as cutting and casting), AM allows for the direct construction of complex, functional parts from 3D digital models, eliminating the need for mold production **[2]**. This process significantly simplifies manufacturing steps and shortens the product development cycle. Metal additive manufacturing (MAM) is widely regarded as a revolutionary technology that overcomes the inherent limitations of traditional craftsmanship regarding product design and processing efficiency. Diverse MAM technologies are currently common, including wire-arc additive manufacturing (WAAM), directed energy deposition (DED), electron beam freeform fabrication (EBF3), and the most extensively applied laser powder bed fusion (LPBF) **[5]**. These methods are effective in producing functional alloy parts and offer high material utilization alongside geometric design flexibility.

Among these technologies, selective laser melting (SLM) is one of the most prevalent metal printing techniques **[6]**. SLM utilizes a high-power laser to selectively melt and stack metal powder layer by layer, ultimately fabricating metal components characterized by high precision and superior mechanical performance. This capability enables SLM to produce geometrically complex and highly integrated functional parts, thereby increasing design flexibility and freedom. These benefits translate into excellent tensile strength, ductility, and antifatigue properties. Given these advantages, SLM is widely employed in quality- and precision-critical industries, such as aerospace, medical, energy, and transportation, demonstrating its pivotal value in customized precision manufacturing. However, despite the significant benefits SLM offers in high-precision and complex part fabrication, practical application still faces numerous challenges. The core issue revolves around ensuring process stability and product

quality consistency. The rapid layer-by-layer stacking process, involving melting and solidification, creates a highly nonlinear and dynamic interaction between energy input and material behavior, making the final product performance highly susceptible to various factors **[7]**. Extensive research has investigated this critical challenge.

Crucially, traditional quality inspection methods (such as destructive testing or magnetic measurements) are often time-consuming, costly, and cannot satisfy the need for real-time process monitoring and control. Therefore, establishing a rapid and cost-effective online monitoring system is paramount to ensure the stable and reliable performance of AM products. Reviewing the literature reveals that process parameters significantly influence product characteristics, and manufacturing defects are often reflected in the part's surface texture **[11]**. Consequently, a monitoring system that combines process parameters and surface texture features will aid in more accurately predicting and controlling product quality. While existing research has successfully demonstrated good predictive performance for magnetic properties under low-frequency conditions (e.g., 50 and 200 Hz), a critical limitation exists: the model performance shows a significant decline when addressing high-frequency conditions (e.g., 400 and 800 Hz).

High-frequency magnetic characteristics are essential for many advanced applications, yet the complex, nonlinear dependencies introduced by the high-frequency response pose a major challenge that current deep learning models struggle to reliably capture. Therefore, to overcome this limitation and significantly enhance the prediction capability under challenging high-frequency conditions, this study proposes a novel lightweight deep learning architecture, the conv-enhanced progressive sampling vision transformer (CVPSViT). The ultimate goal of this research is to establish a deep learning architecture that possesses spatial perception, modal fusion capability, and practical value, thereby providing a forward-looking solution for real-time quality monitoring and magnetic material property prediction in the selective laser melting (SLM) process. To achieve this, the CVPSViT model introduces the following three main contributions upon the standard ViT framework:

1. **3D Volumetric Input Construction:** To exploit the intrinsic spatiotemporal correlations in SLM, we stack successive layer-wise images into a 3D volumetric structure. This design significantly enhances the model's sensitivity to interlayer continuity and micro-depth textural variations, surpassing traditional 2D analysis.
2. **Conv-enhanced Progressive Sampling:** This paper proposes a streamlined architecture that replaces the standard transformer encoder with multilayer 3D convolutions within a progressive sampling framework. This hybrid approach optimizes the extraction of local structural details while simultaneously reducing

computational overhead and improving inference efficiency.

3. **Multimodal Process Parameter Fusion:** Recognizing the deterministic role of fabrication conditions, we embed critical process parameters (e.g., laser power, scanning speed) into a latent feature space. These vectors are fused with image representations to facilitate joint learning, ensuring predictions are grounded in visual evidence and physical constraints.

Through these three key improvements, CVPSViT is designed to significantly enhance prediction stability and accuracy under high-frequency conditions, all while maintaining a lightweight and efficient structure.

2. RELATED WORK

A. Image Processing and Deep Learning in Additive Manufacturing Applications

In situ monitoring during SLM generates high-dimensional visual streams that encode complex process dynamics and thermal histories. Exploiting this data, early research focused on extracting explicit statistical descriptors to correlate process signatures with product quality. Chang et al.**[14]** pioneered this approach by analyzing layer-wise imagery through the gray-level co-occurrence matrix (GLCM) to isolate texture features. By fusing these visual descriptors with critical process parameters, specifically oxygen concentration, laser power, scanning speed, and scan spacing, they employed ensemble learning algorithms, including support vector regression (SVR)**[15]**, XGBoost **[16]** and LightGBM **[17]** , to predict magnetic properties such as permeability and core loss. Building upon this feature-engineering paradigm, Liu et al. **[18]** investigated the spatiotemporal correlations inherent in SLM. The authors extended the analysis to cross-sectional images, arguing that incorporating spatial continuity improves the predictive fidelity of magnetic characteristics compared to isolated layer analysis.

Despite the efficacy of these traditional machine learning pipelines, they face intrinsic scalability barriers. The reliance on hand-crafted features (e.g., GLCM) necessitates domain-specific preprocessing, effectively compressing rich raw imagery into low-dimensional statistics. This reductionist approach often discards subtle, nonlinear spatial dependencies crucial for generalization across diverse materials and varying process windows. Consequently, such models typically require extensive recalibration when applied to new fabrication contexts, limiting their deployment in flexible manufacturing systems.

To address the latency and information loss associated with manual feature extraction, the field has pivoted toward deep learning architectures capable of end-to-end representation learning**[19]**. Unlike traditional methods, deep models can

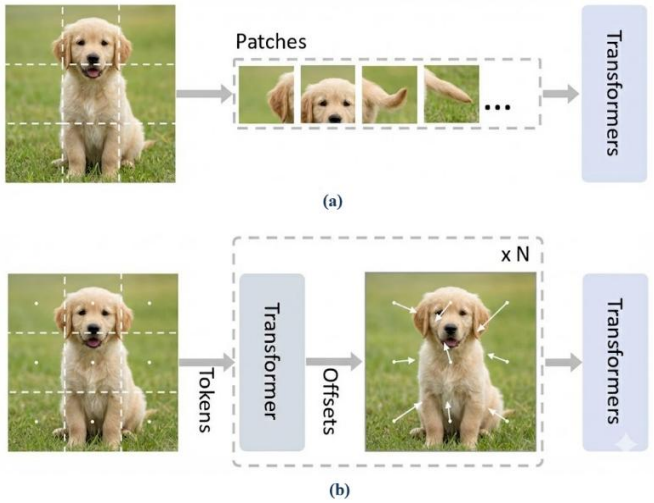


Figure 1
(a) Original ViT with Fixed, Equally Spaced Patch Partitioning; (b) Progressive Sampling That Adaptively Shifts Sampling Locations Toward Salient Regions.

autonomously capture hierarchical features directly from raw data. A notable advancement in this domain is the work of Chang et al**[20]**, who adapted the convolutional vision transformer (CvT)**[21]** for metal additive manufacturing.

Rather than analyzing single layers in isolation, the authors constructed a 3D texture stack from sequential layer images, creating a volumetric representation that implicitly encodes the part's thermal history. When integrated with process parameters, this architecture enables the model to capture complex interlayer interactions and structural heterogeneity. The results demonstrated high precision in predicting multiple magnetic performance indicators, including maximum magnetic flux density (B_m), coercivity (H_c), permeability (μ_a), remanence (b_r), and core loss (P_{cv}), underscoring the capacity of deep neural networks to decipher the latent physics within AM monitoring data without human intervention in feature engineering.

B. Vision Transformer

The ViT architecture was proposed by the Google team in 2020**[22]**. This marked the first time that the transformer model, originally designed for natural language processing (NLP)**[23]**, was directly applied to image classification tasks. Its introduction successfully challenged the long-term dominance of convolutional neural networks (CNN) in visual model design **[24]**. The success of ViT demonstrated that, even without relying on traditional convolution operations, semantic and spatial structures within images could be effectively modeled solely through the self-attention mechanism, achieving outstanding performance in various computer vision tasks. ViT's rise laid the foundation for subsequent architectures like the Swin transformer and CvT, making it one of the mainstream architectures in the field of computer vision. The

overall architecture of ViT is generally divided into three main modules, each handling different stages of feature transformation and modeling (as follows)

I. Linear Projection and Positional Embedding

The input image is first divided into several fixed-size image **patches**. Each patch is flattened into a 1D vector, which is then converted into a fixed-dimension feature vector via linear projection. To retain the spatial positional information of the image, a corresponding **position embedding** is added to each patch. A special **class token** is prepended to the overall feature sequence. This token serves as the global representation vector for the entire image.

II. Transformer Encoder Block

The core of ViT is its multilayer stack of transformer encoders, which constructs contextual relationships and global semantic representations among the patches. Each encoder block typically consists of layer normalization, the multihead self-attention (MHSA) mechanism, a feed-forward network (MLP), and dropout. The self-attention mechanism is fundamental to the entire transformer architecture. Its underlying design principle is that the representation of each patch should be determined jointly based on all other patches across the image. This is achieved by calculating the association (attention weight) between every pair of patches, which dictates the flow and integration of information. This method is effective for capturing long-range dependencies and helps the model understand more implicit or nonlocal structural features in the image. The core formula for self-attention is

$$Attention(Q, K, V) = softnax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

In Eq. (1), Q, K, and V are matrices composed of the query, key, and value vectors, respectively. The scaling factor is introduced primarily to suppress the excessively sharp softmax distribution resulting from large dot products in high-dimensional space, thereby enhancing gradient stability and training efficacy. Furthermore, the multihead attention mechanism further enhances the model's expressive capabilities. It projects the query, key, and value vectors into multiple subspaces, calculates attention separately, and then concatenates and merges all attention results to form a more comprehensive semantic understanding. This design allows the model to simultaneously focus on semantic features across different levels and regions, effectively boosting its recognition and classification abilities. The formula is given as

$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_h)W^O \quad (2)$$

where the operation for the *i*th attention head is

$$head_i = Attention\left(QW_i^Q, KW_i^K, VW_i^V\right) \quad (3)$$

In these formulas, W_i^Q, W_i^K, and W_i^V are the linear projection matrices for the query, key, and value within the *i*th head, and W^O is the final output projection matrix after concatenating the results of all attention heads. Multihead attention is a critical design component of the Transformer architecture, facilitating semantic understanding by attending to sequential information from multiple perspectives in parallel.

III. Multilayer Perceptron Head

This module is responsible for converting the output of the transformer encoder into the final classification result. Specifically, it extracts the output representation of the **class token** and passes it through one or more fully connected layers for classification prediction, yielding the corresponding probability for each category. Collectively, ViT demonstrates the powerful potential of constructing visual models solely via a pure attention mechanism, successfully moving beyond the constraints of CNNs. It has provided the foundation for subsequent architectures like the Swin transformer and CvT and has quickly established itself as a mainstream architecture in the computer vision domain.

C. Progressive Sampling Module

Traditional ViTs uniformly divide the input image into nonoverlapping patches, which are subsequently fed into the transformer encoder. However, this fixed segmentation method presents two notable limitations. First, the positions of these tokens are predefined and evenly distributed, failing to take into account the semantic structure and key regions within the image. When regions with strong semantics or object boundaries are inadvertently partitioned into different patches, the structural continuity may be compromised, leading to the fragmentation of semantic information. Second, this equidistant segmentation tends to allocate a significant number of tokens to background or non-discriminative regions, resulting in wasted attention resources. This hinders the model's ability to focus learning on critical foreground features, thereby impacting overall recognition effectiveness. In contrast, the **human visual system** possesses the ability to actively focus, progressively adjusting the areas of attention based on the importance of objects and recognition demands to construct a holistic understanding. Inspired by this phenomenon, Chen et al.**[25]** introduced the progressive sampling module (PSM). This module implements a strategy that **dynamically updates the sampling positions**, thereby enhancing the model's capability to focus on critical

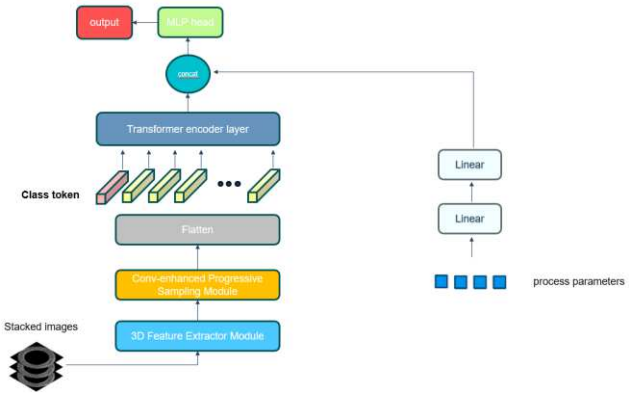


Figure 2

CVPSViT Architecture. The Overall Process Includes a 3D Feature Extraction Module, a Gradient Enhancement Module, a Transformer Encoder Layer, and a Final Prediction Module That Incorporates Manufacturing Process Parameters.

semantic regions within the image. As conceptually illustrated in Figure 1(b), after each sampling round, the module predicts the direction of the next sampling offset based on the output results of the current tokens. Through this iterative process, attention is gradually guided toward **more discriminative regions** of the image. This mechanism overcomes the limitation of fixed sampling in conventional ViT, strengthening the model's capacity to capture and model semantic targets. The design of this strategy, specifically the **CPSM** in our work, mimics the **coarse-to-fine focusing characteristic of human vision**.

3. METHODOLOGY

A. Overview of the Methodology

To enhance the prediction performance of magnetic properties for metal additive manufactured samples under high-frequency conditions, this study proposes a novel architecture named CVPSViT. This architecture integrates 3D convolutional feature extraction, spatial semantic focusing, and global contextual modeling, alongside the incorporation of process parameter information. It is dedicated to simultaneously capturing the combined effects of microstructure and processing conditions on material characteristics. The overall architecture, as illustrated in Figure 2, is divided into the following four main processing stages:

I. 3D Feature Extractor Module

The layer-by-layer images derived from the SLM process are stacked into 3D volumes. A multilayered 3D convolutional network is utilized to extract spatial texture and structural depth information, which serves as the foundational feature map for subsequent sampling.

II. Conv-enhanced Progressive Sampling Module

Mimicking the coarse-to-fine focusing characteristic observed in human vision, this

module employs multiround dynamic position updates and local convolutional operations to capture semantically salient regional features, thereby enhancing the model's attention to critical areas.

III. Transformer Encoder Layer

The sampled sequence features are input into a stacked transformer encoder to construct cross-regional global correlation and semantic integration. This layer also extracts a special class token representation that aggregates the global semantics of the entire image.

IV. Process Parameter Fusion Design

Key process parameters (such as laser power, oxygen concentration, and scanning speed) are processed through two linear layers to extract latent feature representations. These features are then concatenated with the class token vector output from the transformer encoder, forming an integrated final representation. This integrated representation is finally used by a fully connected layer to predict the magnetic properties.

This design preserves the advantage of semantic-oriented characteristic sampling provided by progressive sampling. By incorporating structural information and process parameters, it effectively enhances the prediction stability and accuracy of the model under high-frequency conditions. The subsequent subsections will provide detailed descriptions of each of these modules.

B. 3D Feature Extractor Module

This module represents the first stage within the CVPSViT architecture. Its primary purpose is to extract representative deep spatial features from the input 3D stacked image. These extracted features serve as the foundation for the subsequent modules dedicated to semantic focusing and discriminative learning. Compared to traditional 2D convolutions that process only planar images, this module utilizes 3D convolution to directly model the interlayer continuity along the depth dimension [26]. This approach offers a higher capability for identifying texture variations and microstructures between layers within the metal additive manufacturing process. The overall architecture of this module is composed of three layers of 3D convolution. A 3D MaxPooling operation is introduced between the first and second convolutional layers. Each convolutional operation is coupled with batch normalization and the ReLU activation function, which are implemented to enhance the model's stability and nonlinear representation capacity. The specific processing flow is detailed as follows:

I. Input Image

The input consists of 3D volume data with three channels and a size of 16 × 256 × 256, corresponding to image depth, height, and width, respectively.

II. First Convolutional Layer

A convolutional kernel size of 7 × 7 × 7 and a stride of 2 are employed for initial feature extraction and spatial compression. After appropriate padding, the output size is 64 × 8 × 128 × 128, meaning the depth, height, and width are halved, and the channel count is increased to 64.

III. 3D MaxPooling

A MaxPooling operation with a kernel size of 3 × 3 × 3 and a stride of 2 is applied, further compressing the spatial dimensions to 64 × 4 × 64 × 64. This step effectively reduces the computational load for subsequent modules while preserving important spatial features.

IV. Second Convolutional Layer

A 3 × 3 × 3 convolutional kernel is used while keeping the channel count unchanged, further strengthening the learning of local textures and interlayer structures. The output size remains 64 × 4 × 64 × 64, but the feature representation becomes more refined.

V. Third Convolutional Layer

A 1 × 1 × 1 convolutional kernel is utilized for channel upscaling, increasing the channel count to 1024. The output size is 1024 × 4 × 64 × 64. This step significantly enhances the semantic capacity and discriminative power of the features, providing a sufficient semantic foundation for the subsequent progressive sampling module.

In summary, this module efficiently transforms the raw 3D images into a high level, semantically rich, and spatially compressed feature representation. This process not only improves the efficiency of subsequent processing but also lays a solid foundation for the overall model's performance in detailed modeling and semantic recognition.

C. Conv-enhanced Progressive Sampling Module

The CPSM module is a core component of the CVPSViT architecture. It integrates the progressive sampling strategy with a convolutional feature enhancement mechanism. This design establishes independent sampling coordinates for each depth slice and dynamically updates the sampling positions based on semantic features in every round. This approach mimics the process of human vision, which gradually focuses and integrates information when observing objects, enabling the model to progressively concentrate on critical areas and accumulate semantic information over multiple sampling rounds to enhance its understanding of discriminative regions, as shown in

Figure 3. The overall operation is detailed as follows.

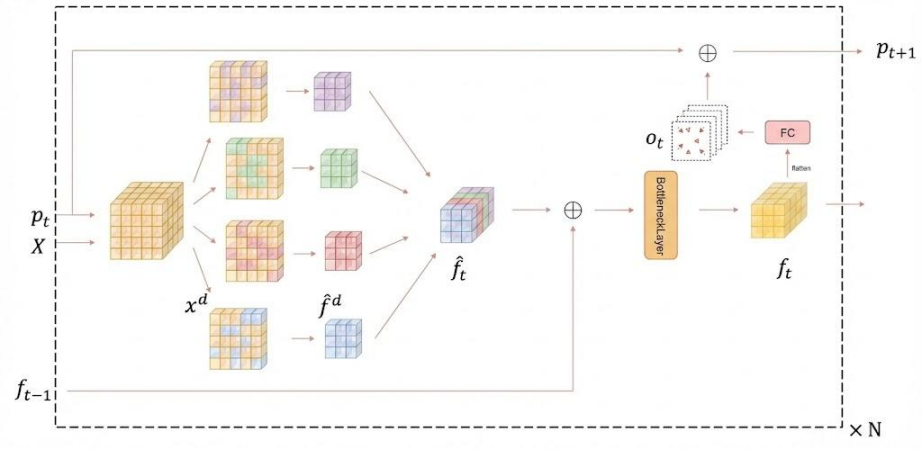


Figure 3
Conv-enhanced Progressive Sampling Module Flow Chart.

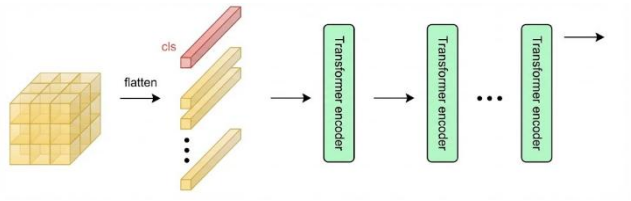


Figure 4
It Uses the ViT architecture, Omitting Traditional Patch Embedding and Position Encoding, and Directly Processing the Output Features from the CPSM module.

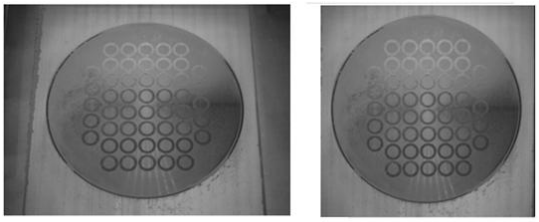


Figure 5
Image Geometric Correction: (a) Before (b) After

Slicing and Sampling

The input feature map X comes from the preceding 3D feature extractor module. It is a 4D feature map with shape $X \in R^{C\times D\times H\times W}$, where C is the number of channels, D is the depth dimension (i.e., the number of stacked slices), and H and W denote the height and width, respectively. Since X is a 4D feature map, it is sliced along the depth dimension, where each slice is treated as a 3D feature map:

$$X = \left[x^{(1)}, x^{(2)}, \ldots, x^{(D)}\right] \ , \ x^{(d)} \in \mathbb{R}^{C\times H\times W} \tag{4}$$

In each sampling round, the model receives the previous round's feature map f_{t-1}, the input feature X, and the current sampling coordinates p_t. The sampling coordinates are defined as

$$p_t = \left[{p_t}^{(1)}, x{p_t}^{(2)}, \ldots, {p_t}^{(D)}\right] \ , \ p_t \in \mathbb{R}^{D\times(H\times W)\times 2} \tag{5}$$

This indicates that $n \times n$ 2D points of interest are sampled on each slice, where $p_t^{(d)} \in R^{(n\times n)\times 2}$. In the initial stage $(t = 0)$, these coordinates are initialized with a uniform grid distribution, similar to the patch grid in the vision transformer. Features

are sampled from each slice $x^{(d)}$ according to the coordinates $p_t^{(d)}$:

$$\tilde{f}_t^{(d)} = Sample\left(x^{(d)}, p_t^{(d)}\right),\ \tilde{f}_t^{(d)} \in \mathbb{R}^{(n\times n)\times C} \quad (6)$$

These features are then reshaped into a spatial feature map:

$$\hat{f}_t^{(d)} = Reshape\left(\tilde{f}_t^{(d)}\right),\ \hat{f}_t^{(d)} \in \mathbb{R}^{n\times n\times C} \quad (7)$$

I. Feature Stacking and Dimension Adjustment

All $\hat{f}_t^{(d)}$ are stacked into a 4D feature map:

$$\hat{f}_t = Concat\left(\hat{f}_t^{(1)}, \ldots, \hat{f}_t^{(D)}\right),\ \hat{f}_t \in \mathbb{R}^{D\times n\times n\times C} \quad (8)$$

Subsequently, the dimensions are permuted to match the input format required by the convolutional layers:

$$\hat{f}_t = permute\left(\hat{f}_t\right),\ \hat{f}_t \in \mathbb{R}^{C\times D\times n\times n} \quad (9)$$

II. Residual Fusion and Bottleneck Enhancement

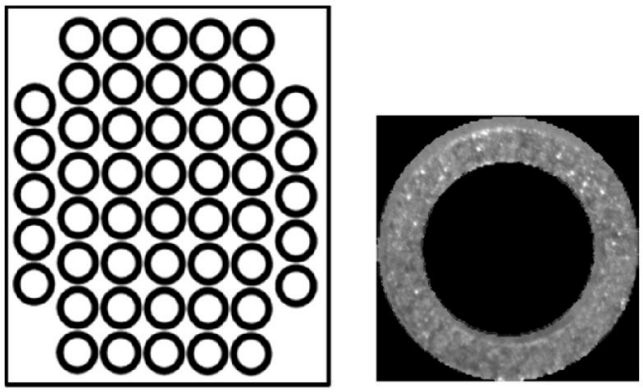
The current sampled feature $\hat{f}_t$ is residually summed with the previous round's semantic feature and subsequently fed into a bottleneck layer composed of three 3D convolutional layers that sequentially perform channel compression, spatial modeling, and channel upscaling. This residual fusion strategy is inspired by residual learning principles, which have been shown to facilitate stable optimization in deep architectures **[27]**. Moreover, the overall bottleneck design aligns with prior studies such as DenseNet **[28]**, which emphasize efficient feature reuse to reduce parameter redundancy while maintaining strong representational power:

$$f_t = BottleneckLayer(\hat{f}_t + f_{t-1}),\ f_t \in \mathbb{R}^{C\times D\times n\times n} \quad (10)$$

III. Offset Prediction and Coordinate Update

First, the feature map f_t is flattened into a 1D vector and fed into a fully connected layer (FC) to predict the 2D offset amount for each sampling point. Next, the prediction result is reshaped to have the same structure as the original coordinates p_t, forming the offset o_t:

$$o_t = Reshape(FC(Flatten(f_t))),\ o_t \in \mathbb{R}^{D\times(n\times n)\times 2} \quad (11)$$



(a) (b)

Figure 6

(a) Masks Generated by the Slicing Software Materialise Magics Define the Ideal Position and Shape of Each Workpiece in Additive Manufacturing, Serving as References for Image Alignment and Segmentation. (B) After Geometric Correction and Image Processing, the Main Circular Regions are Preserved While Background Noise is Removed.

Finally, the offset is added back to the original coordinates to update the sampling position for the next round:

$$p_{t+1} = p_t + o_t \quad (12)$$

This process achieves **round-by-round adaptive sampling** through feature feedback and offset prediction, guiding the model to focus on semantically important regions and enhancing discriminative features.

D. Transformer Encoder Layer

This module adopts the encoder architecture of the ViT and serves as the critical component for image semantic modeling in this study. As illustrated in Figure 4, we omit the traditional ViT components, specifically the patch embedding and positional encoding designs. Instead, we directly process the 3D feature map $f_t \in \mathbb{R}^{C\times D\times H\times W}$ output from the CPSM module. The processing steps are as follows:

I. Sequence Conversion and Class Token

The feature map f_t is first flattened into a sequence representation. A learnable class token is then prepended to this sequence, designed to aggregate the global information of the entire image. The resulting input sequence has a size of $\mathbb{R}^{(1+D\cdot H\cdot W)\times C}$, where the "1" denotes the class token.

II. Omitting Positional Encoding

Since the preceding module (CPSM) has already processed the features through multilayered convolutions, the receptive field coverage implicitly constructs a certain degree of spatial order. Therefore, this study omits the additional positional encoding.

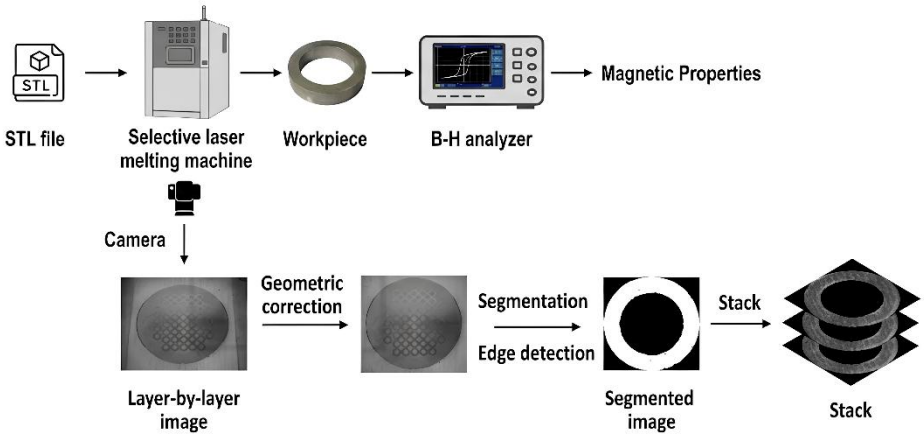


Figure 7
Data Processing Flow Chart.

III. Global Feature Modeling

The entire feature sequence is subsequently fed into multilayered transformer encoders (including multihead self-attention and the feed-forward network, FFN). These layers perform deep semantic extraction and global feature modeling.

IV. Stabilization

Layer normalization and residual connections are employed to enhance training stability and efficiency. By leveraging the transformer's global modeling capability, the model effectively integrates the spatial features focused upon by the CPSM module. This process extracts high-level information that is both semantic and discriminative, providing highly expressive features for the subsequent prediction module.

E. Process Parameter Fusion Design

Given that the mechanical properties of additively manufactured components are deterministically governed by their thermal history and processing conditions, relying solely on visual inspection may yield incomplete predictions. To bridge this gap, the proposed architecture incorporates a dedicated multimodal fusion module that synergizes scalar process parameters with high-dimensional visual features. Rather than direct injection, raw process parameters are first projected into a latent embedding space via dual linear transformations, mapping them to a feature dimension compatible with the visual representations. These encoded parameter vectors are then concatenated with the class token (CLS), a global semantic descriptor derived from the transformer encoder that aggregates holistic image information. This fused representation is subsequently processed by a fully connected layer to regress the target material properties. By integrating governing process conditions at the semantic level, this design allows the model to contextualize observed microstructural features within their fabrication settings, thereby establishing a more robust physics-informed basis for

generalization and prediction accuracy.

4. MATERIALS and EXPERIMENTAL RESULTS

A. Data Collection

The dataset utilized in this study was organized and adjusted based on existing research data. The material employed was the soft magnetic alloy Fe-50Ni, fabricated using the selective laser melting (SLM) technique. The 3D model was converted into an STL file and imported into the SLM equipment for printing. During the printing process, the equipment automatically performs powder spreading and selective laser melting, stacking the material layer by layer according to the pre-sliced model contour until the complete metal structure is constructed. Simultaneously, during the printing period, an optical camera was used to capture image data corresponding to the powder spreading and melt pool formation for every layer. To address potential limitations, which could compromise the accuracy of subsequent analysis, this study implemented geometric calibration using a calibration plate. By capturing four known spatial control points on the calibration plate and mapping them to the four vertices of an ideal rectangular region, a perspective transformation matrix was established. This matrix facilitates the derivation of the geometric mapping relationship between the world coordinate system and the camera coordinate system. Consequently, 3D points in the camera coordinates can be converted into 2D coordinates on the imaging plane and ultimately mapped to pixel coordinates in the image, ensuring precise correspondence from physical space to the digital image.

This process is crucial for subsequent image segmentation and workpiece localization. The effectiveness of the geometric calibration in eliminating perspective distortion is illustrated in Figure 5b. For the geometrically corrected images, the study combined a predesigned mask image to precisely extract the workpiece region. This mask was generated by the slicing software Materialise Magics prior to part construction (as shown in Figure 6a), providing clear and consistent information regarding the ideal position and geometry of the part. The mask was geometrically aligned with the corrected image, and the Canny edge detection algorithm was applied to extract the workpiece contour. The extracted contour was verified against the expected structure in the mask to accurately lock the location of each workpiece area. The processing workflow also included appropriate thresholding based on image noise characteristics to exclude nonstructural background information and detection errors, thereby enhancing the stability and accuracy of feature recognition. Finally, each successfully segmented workpiece area was saved as an independent image file for model training. As shown in Figure 6b, the final segmentation successfully retained the main ring structure of the workpiece while effectively removing superfluous background noise.

The layerwise images for each workpiece, constructed through the procedure noted above, were paired with the corresponding process parameters to establish a structured dataset.

The overall data encompasses 40 sets of different process parameter settings, with five ring-shaped workpieces included per set, leading to a total of 200 samples. Each workpiece corresponds to 200 layers of images, all with a size of 256 × 256 pixels. To support the requirements of 3D feature learning, the preprocessing stage constructed continuous volume images starting from the sixteenth layer, with every 16 layers grouped together. This resulted in 185 volumes generated per workpiece, culminating in a total of 37,000 3D image data entries serving as the input for the deep learning model. To acquire the magnetic material parameters necessary for model training labels, the SY-8219 B-H analyzer manufactured by IWATSU Corporation was used. By winding the tested samples, the required magnetic characteristics were measured. These parameters include maximum magnetic flux density (B_m), coercivity (H_c), permeability (μ_a), remanence (b_r), and core loss (P_{cv})**[29]**. Furthermore, tests were conducted under multiple frequency conditions, including 50, 200, 400, and 800 Hz, to observe the material's magnetic behavior. The overall data processing workflow is summarized in Figure 7.

B. Experimental Design and Training Setup

The deep learning experiments in this study were implemented using the Python programming language, and all model construction and training procedures were completed based on the PyTorch framework. All training and inference operations were executed on a single NVIDIA GeForce RTX 4090 GPU platform, ensuring high computing efficiency and stable resource allocation.

For model training, the batch size was set to 16, and the total training epoch count was 150. The initial learning rate was set to 0.0015. To ensure stable convergence and prevent oscillations, this study employed a stepwise exponential decay strategy for learning rate adjustment: the learning rate was decayed to half of the previous stage every 50 epochs. The Adam optimization algorithm was selected due to its characteristics of adaptive learning rate adjustment, which helps accelerate model convergence. The L2 regularization coefficient (weight decay) was set to 0.001 to suppress the phenomenon of overfitting, and a stability factor of $\varepsilon = 1 \times 10^{-7}$ was added to ensure numerical calculation stability. Regarding the loss function, the mean squared error (MSE) was adopted, which is defined as follows:

$$L_{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \quad (13)$$

where y_i is the actual value of the *i*th sample, $\hat{y}_i$ is the model's predicted value, and

Table 1.
Average performance metrics at 50Hz across different magnetic parameters.

Metrics	B_m	H_c	μ_a	B_r	P_{cv}
R^2	0.983 ±0.000	0.984 ±0.002	0.984 ±0.002	0.981 ±0.002	0.985 ±0.002
MSE	104.814 ±6.020	13.909 ±1.640	11.630 ±0.946	36.464 ±4.647	2.300 ±0.325
MAE	8.217 ±0.335	2.907 ±0.237	2.650 ±0.155	4.666 ±0.473	1.171 ±0.034

Table 2.
Average performance metrics at 200Hz across different magnetic parameters.

Metrics	B_m	H_c	μ_a	B_r	P_{cv}
R^2	0.984 ±0.002	0.985 ±0.001	0.985 ±0.005	0.983 ±0.001	0.985 ±0.002
MSE	94.799 ±15.001	100.504 ±7.454	8.815 ±2.302	64.528 ±4.579	179.187 ±22.071
MAE	7.817 ±0.792	7.875 ±0.408	2.378 ±0.319	6.194 ±0.312	10.490 ±0.848

Table 3.
Average performance metrics at 400Hz across different magnetic parameters.

Metrics	B_m	H_c	μ_a	B_r	P_{cv}
R^2	0.965 ±0.002	0.988 ±0.001	0.970 ±0.006	0.987 ±0.003	0.985 ±0.001
MSE	71.714 ±6.836	132.491 ±11.878	6.268 ±0.971	42.957 ±7.564	522.865 ±36.073
MAE	6.684 ±0.182	8.907 ±0.384	1.975 ±0.129	5.127 ±0.553	17.877 ±1.085

Table 4.
Average performance metrics at 800Hz across different magnetic parameters.

Metrics	B_m	H_c	μ_a	B_r	P_{cv}
R^2	0.956 ±0.003	0.981 ±0.001	0.948 ±0.005	0.941 ±0.011	0.934 ±0.007
MSE	38.994 ±1.604	74.773 ±5.270	4.410 ±0.572	32.778 ±6.231	1253.999 ±167.149
MAE	4.798 ±0.288	6.864 ±0.379	1.633 ±0.145	4.338 ±0.388	28.527 ±1.795

N is the total number of samples. To comprehensively evaluate the model's prediction accuracy and generalization performance in regression tasks, the following three performance metrics were also used for quantitative analysis:

I. Mean Absolute Error (MAE)

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \qquad (14)$$

II. Mean Squared Error (MSE) Identical to Eq. (13)

III. Coefficient of Determination (Score)

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}\ , \qquad (15)$$

where $\bar{y}$ is the average of all actual values. This metric measures the goodness of fit between the model's predictions and the true values. Furthermore, to evaluate the model's stability and generalization ability across different data partitioning scenarios, the study employed fourfold cross-validation as the experimental design strategy. Specifically, the entire dataset was divided into four equal subsets, and four rounds of training and validation were conducted. In each validation round, one subset was designated as the validation set, while the remaining three subsets served as the training set. This design ensures that every piece of data participates in the validation process in one round, thereby enhancing the representativeness and stability of the overall model performance evaluation and effectively reducing the accidental effects caused by specific data segmentation.

C. Analysis of Model Prediction Performance at Different Frequencies

Tables 1 through 4 detail the regression performance across four distinct frequency regimes (50, 200, 400, and 800 Hz), evaluating five critical magnetic parameters: maximum magnetic flux density (B_m); coercivity (H_c); permeability (u_a); remanence (B_r); and core loss (P_{cv}). In the low-frequency domain (50 Hz), the model achieved optimal fidelity; the coefficient of determination (R^2) exceeded 0.98 across all metrics, indicating a near-perfect capture of the underlying magnetic behaviors. Specifically, the prediction of core loss (P_{cv}) was exceptionally precise, yielding a minimal mean absolute error (MAE) of 1.171 ± 0.034. As the frequency increased to 200 Hz, the model maintained high robustness with R^2 values sustained between 0.983 and 0.985. However, early signs of frequency-dependent sensitivity emerged in the (P_{cv}) predictions, where the mean squared error (MSE) rose to 179.187 ± 22.071, suggesting that dynamic loss mechanisms become increasingly difficult to regress as frequency rises.

Transitioning to the medium-to-high frequency spectrum (400 and 800 Hz), the complexity of nonlinear magnetic responses posed greater challenges to the predictive architecture. At 400 Hz, while the overall R^2 remained respectable (0.965–0.988), the error metrics for P_{cv} escalated significantly (MAE reaching 17.877 ± 1.085), reflecting the intensified nonlinearity of iron loss components—hysteresis and eddy currents—at higher frequencies. This trend was exacerbated at 800 Hz, where performance degradation was most pronounced. With the exception of coercivity (H_c), R^2 values for other parameters dipped below 0.94, and the MSE for P_{cv} peaked at 1253.999 ± 167.149. This decline is attributed to the highly stochastic nature of high-frequency magnetic domains, which introduces variance that is inherently harder to model based solely on static structural features.

Despite the expected attenuation in accuracy at higher frequencies, the proposed model

exhibited commendable comprehensive robustness. The architecture sustained R^2 values consistently within the 0.94–0.97 range for the majority of cases, with no single metric falling below 0.92, even under the most demanding conditions. This consistency underscores the model's generalization capability. Further, when juxtaposed with benchmark models (detailed in Section 4.4), our approach demonstrated superior adaptability, effectively mitigating the severe performance drop-offs typically associated with high-frequency magnetic property prediction.

D. Comparison of Performance Among Different Models

To further validate the prediction performance of the proposed CVPSViT model, we conducted a comprehensive comparison against five conventional machine learning models (logistic regression, linear regression, SVR **[15]**, LightGBM **[17]**, XGBoost[16]) and the deep learning model CvT [22]. The comparison encompassed five magnetic parameters, i.e., maximum magnetic flux density (B_m), coercivity (H_c), magnetic permeability (μ_a), remnant magnetism (B_r), and iron loss (P_{cv}), utilizing the coefficient of determination (R^2) as the primary performance metric, evaluated across four frequencies: 50, 200, 400, and 800 Hz. Tables 5 to 9 illustrate that CVPSViT demonstrates a significant performance advantage, consistently surpassing conventional machine learning models across all tested frequencies. Comparison with conventional machine learning models, as follows.

- **Maximum Magnetic Flux Density (B_m):** At 50 Hz, the R^2 value for CVPSViT reached 0.986, which is notably superior to the best-performing conventional model, LightGBM (0.932). Even at the high frequency of 800 Hz, CVPSViT maintained high accuracy at 0.955, whereas the overall performance of traditional models declined, suggesting their insufficient adaptation to high-frequency signals.

- **Coercivity (H_c):** CVPSViT exhibited stable and excellent prediction capability, achieving an R^2 of 0.981 across all four frequencies. In contrast, logistic regression (0.845) and linear regression (0.635) showed significantly lower R^2 values at 800 Hz, highlighting the limitations of these models in characterizing high-frequency magnetic properties.

- **Magnetic Permeability (μ_a):** CVPSViT reached an R^2 of 0.984 at low frequencies (50 and 200 Hz) and sustained 0.948 even at 800 Hz, demonstrating

Table 5.
Comparison of R^2 implementation of each model at different B_m frequencies

B_m	CVPSViT	CvT	XGBOOST	LIGHTGBM	SVR	Logistic	Linear
50 Hz	0.986	0.987	0.924	0.932	0.912	0.896	0.586
200 Hz	0.984	0.985	0.934	0.936	0.919	0.909	0.597
400 Hz	0.965	0.969	0.924	0.936	0.900	0.894	0.543
800 Hz	0.955	0.940	0.903	0.914	0.868	0.858	0.584

Table 6.
Comparison of R^2 implementation of each model at different H_c frequencies

Hc	CVPSViT	CvT	XGBOOST	LIGHTGBM	SVR	Logistic	Linear
50 Hz	0.984	0.974	0.948	0.953	0.927	0.935	0.654
200 Hz	0.985	0.989	0.909	0.919	0.904	0.894	0.647
400 Hz	0.989	0.986	0.972	0.978	0.972	0.960	0.683
800 Hz	0.981	0.876	0.866	0.865	0.867	0.845	0.635

Table 7.
Comparison of R^2 implementation of each model at different μ_a frequencies

μ_a	CVPSViT	CvT	XGBOOST	LIGHTGBM	SVR	Logistic	Linear
50 Hz	0.984	0.984	0.929	0.942	0.910	0.902	0.605
200 Hz	0.985	0.984	0.928	0.939	0.892	0.895	0.568
400 Hz	0.970	0.961	0.927	0.934	0.890	0.895	0.525
800 Hz	0.948	0.922	0.904	0.907	0.845	0.896	0.585

Table 8.
Comparison of R^2 implementation of each model at different B_r frequencies

Br	CVPSViT	CvT	XGBOOST	LIGHTGBM	SVR	Logistic	Linear
50 Hz	0.981	0.983	0.893	0.909	0.858	0.863	0.528
200 Hz	0.983	0.989	0.918	0.933	0.892	0.897	0.584
400 Hz	0.987	0.988	0.921	0.933	0.893	0.901	0.588
800 Hz	0.940	0.928	0.882	0.894	0.804	0.841	0.517

Table 9.
Comparison of R^2 implementation of each model at different P_{cv} frequencies

Pcv	CVPSViT	CvT	XGBOOST	LIGHTGBM	SVR	Logistic	Linear
50 Hz	0.985	0.987	0.884	0.888	0.849	0.790	0.539
200 Hz	0.984	0.988	0.908	0.926	0.826	0.831	0.572
400 Hz	0.985	0.983	0.925	0.936	0.886	0.905	0.602
800 Hz	0.934	0.909	0.826	0.840	0.783	0.790	0.420

strong resistance to frequency-induced performance degradation. Conversely, linear regression achieved only 0.585, showing a remarkable performance gap.

- **Remnant Magnetism (B_r):** CVPSViT performed optimally at 400 Hz (0.987), clearly outperforming traditional models. Across all frequencies, the R^2 for linear regression remained below 0.6, indicating its inability to effectively capture the non-linear behavior associated with remnant magnetism.
- **Iron Loss (P_{cv}):** This parameter is the most nonlinear and frequency-sensitive indicator. CVPSViT achieved high R^2 values (all above 0.98) at

Table 10.
R^2 performance comparison of different models and magnetic characteristics (50Hz)

Model	B_m	H_c	μ_a	B_r	P_{cv}
ViT	0.9742 ±0.0037	0.9746 ±0.0025	0.9741 ±0.0026	0.9744 ±0.0013	0.9745 ±0.0045
PSViT	0.9854 ±0.0029	0.9833 ±0.0010	0.9835 ±0.0009	0.9808 ±0.0013	0.9848 ±0.0024
CVPSViT	0.986 ±0.001	0.984 ±0.001	0.984 ±0.002	0.981 ±0.002	0.985 ±0.002

Table 11.
R^2 performance comparison of different models and magnetic characteristics (200Hz)

Model	B_m	H_c	μ_a	B_r	P_{cv}
ViT	0.9723 ±0.0056	0.9773 ±0.0031	0.9652 ±0.0107	0.9725 ±0.0040	0.9744 ±0.0091
PSViT	0.9842 ±0.0017	0.9828 ±0.0056	0.9839 ±0.0016	0.9834 ±0.0015	0.9844 ±0.0031
CVPSViT	0.9838 ±0.0017	0.9853 ±0.0007	0.9847 ±0.0041	0.9830 ±0.0012	0.9844 ±0.0019

Table 12.
R^2 performance comparison of different models and magnetic characteristics (400Hz)

Model	B_m	H_c	μ_a	B_r	P_{cv}
ViT	0.9578 ±0.0063	0.9795 ±0.0046	0.9513 ±0.0214	0.9713 ±0.0052	0.9733 ±0.0064
PSViT	0.9602 ±0.0022	0.9840 ±0.0041	0.9659 ±0.0052	0.9816 ±0.0041	0.9821 ±0.0017
CVPSViT	0.9653 ±0.0015	0.9885 ±0.0011	0.9701 ±0.0054	0.9867 ±0.0025	0.9852 ±0.0013

Table 13.
R^2 performance comparison of different models and magnetic characteristics (800Hz)

Model	B_m	H_c	μ_a	B_r	P_{cv}
ViT	0.9432 ±0.0089	0.9731 ±0.0052	0.9349 ±0.0112	0.9314 ±0.0143	0.9044 ±0.0349
PSViT	0.9454 ±0.0081	0.9797 ±0.0044	0.9455 ±0.0049	0.9407 ±0.0031	0.9325 ±0.0162
CVPSViT	0.9553 ±0.0027	0.9811 ±0.0014	0.9480 ±0.0042	0.9403 ±0.0111	0.9340 ±0.0070

50, 200, and 400 Hz. At 800 Hz, it maintained 0.934, which significantly surpassed benchmark models like XGBoost (0.826), SVR (0.783), and logistic regression (0.790).

The comparison with the existing deep learning architecture, CvT, shows equivalent performance at low frequencies but CVPSViT's superior performance at high frequencies, indicating its distinct advantage in extracting high-frequency information.

- **Maximum Magnetic Flux Density (B_m):** At 50 Hz, CVPSViT (0.986) and CvT (0.987) showed nearly equivalent performance. However, at 800 Hz, CVPSViT improved to 0.955, surpassing CvT's 0.940, a difference of approximately 1.5%.
- **Coercivity (H_c):** CVPSViT demonstrated particularly outstanding performance at 800 Hz, achieving 0.981, significantly higher than CvT's

0.876 (a gap exceeding 10%), thus emphasizing its superiority in high-frequency modeling.

- **Magnetic Permeability (μ_a):** The difference between the two models was minimal at 50 Hz and 200 Hz, both approximating 0.984. Nevertheless, at 800 Hz, CVPSViT achieved 0.948, markedly better than CvT's 0.922.
- **Remnant Magnetism (B_r):** At 400 Hz, CvT's R^2 (0.988) was marginally higher than CVPSViT's (0.987), making it the only frequency/parameter where CvT slightly prevailed. However, under other frequency conditions, CVPSViT maintained the lead; for example, at 800 Hz, CVPSViT achieved 0.940 compared to CvT's 0.928.
- **Iron Loss (P_{cv}):** This metric is highly sensitive to high frequencies. CVPSViT achieved 0.934 at 800 Hz, surpassing CvT's 0.909, thereby demonstrating its robust stability and adaptability in challenging high-frequency scenarios.

E. Ablation Study

To rigorously dissect the contributions of individual architectural components and validate the superiority of the proposed CVPSViT, we devised a systematic ablation study. This analysis evaluates three distinct architectural paradigms: the baseline ViT, which relies solely on global self-attention; the PSViT, which integrates the progressive sampling module to enhance feature selectivity; and the refined CVPSViT, which further incorporates convolutional layers to optimize local feature extraction.

The quantitative results, summarized in Tables 10 through 13, reveal a clear performance hierarchy. The baseline ViT model exhibited the lowest coefficient of determination (R^2) across all tested frequency regimes and magnetic parameters. This underperformance suggests that pure self-attention mechanisms, while powerful in capturing long-range dependencies, struggle to isolate spatially discriminative features within the complex textures of magnetic domain imagery. In stark contrast, PSViT and CVPSViT achieved significantly higher predictive fidelity. This empirical disparity validates the hypothesis that the progressive sampling module is instrumental in guiding the network's attention toward information-rich regions, thereby filtering out background noise and enhancing the semantic density of the extracted features.

A granular comparative analysis between PSViT and CVPSViT further illuminates the specific advantages of the proposed hybrid architecture. While both models deliver comparable performance across most metrics, CVPSViT demonstrates a subtle yet critical robustness edge under challenging conditions. Specifically, in the high-frequency domain (800 Hz), where magnetic behaviors exhibit heightened nonlinearity

and stochasticity, CVPSViT outperformed PSViT in key indicators, including coercivity (H_c), remnant magnetism (B_r), and iron loss (P_{cv}). For instance, CVPSViT attained an R^2 of 0.9811 for H_c and 0.9340 for P_{cv}, surpassing PSViT's scores of 0.9797 and 0.9325, respectively. This marginal but consistent superiority implies that the introduction of convolutional biases aids in capturing local structural nuances that purely attention-based sampling might overlook, ensuring stability even when signal patterns become erratic.

Beyond predictive accuracy, the decisive advantage of CVPSViT lies in its computational efficiency. The standard PSViT architecture is computationally intensive, necessitating 50.71 million parameters and 25.77 GFLOPs. Conversely, by replacing redundant attention computations with efficient convolutional operations, the CVPSViT architecture drastically curtails resource consumption, operating with only 38.50 million parameters and 22.64 GFLOPs. This structural optimization corresponds to a roughly 24% reduction in parameter count and a 12% decrease in computational load.

Collectively, these findings confirm that the conv-enhanced progressive sampling module does not merely sustain the high-performance benchmark set by PSViT: It achieves a superior Pareto operational point. By simultaneously reducing the computational burden while enhancing robustness against high-frequency variations, the CVPSViT model presents an optimal balance between architectural lightness and regression precision. This positions the framework as a highly viable solution for deployment in resource-constrained industrial environments, where real-time monitoring demands both high throughput and minimal latency.

5. CONCLUSION

This study addressed the challenging task of predicting magnetic properties in selective laser melting (SLM) fabricated materials, particularly under high-frequency conditions. We successfully proposed the novel CVPSViT, a lightweight deep learning architecture that integrates layer-stacked image textures and corresponding manufacturing process parameters via an effective cross-modal feature fusion strategy. The CVPSViT design, incorporating an advanced progressive sampling mechanism, is guided to achieve simultaneous localized attention and global understanding, significantly enhancing its ability to perceive complex magnetic behaviors. Experimental validation across various frequencies (50 to 800 Hz) confirmed the superior performance and robustness of the CVPSViT model.

- **Performance:** CVPSViT consistently achieved R^2 scores superior to traditional machine learning models (XGBoost, LightGBM, SVR, etc.) across all five key

magnetic indicators (B_m, Hc, μa, Br, Pcv). Crucially, it maintained excellent predictive accuracy even under the complex, nonlinear high-frequency conditions (400 and 800 Hz), overcoming the performance degradation typically observed in conventional models and demonstrating a clear performance improvement over the CvT architecture in high-frequency scenarios.

- **Efficiency:** Compared to the original PSViT baseline, CVPSViT achieved comparable or better performance while utilizing significantly fewer parameters (38.50 M vs. 50.71 M) and requiring a lower computational load (22.64 GFLOPs vs. 25.77 GFLOPs).

In summary, the CVPSViT model offers a highly efficient and practical solution for real-time quality monitoring and accurate magnetic property prediction within industrial SLM smart manufacturing environments.

6. FUTURE WORK

Although the proposed CVPSViT model has demonstrated high accuracy and stability in multifrequency magnetic property prediction, suggesting potential application value in smart manufacturing environments, several directions warrant further exploration and optimization:

- **Further Enhancement of High-Frequency Prediction Performance:** Despite the CVPSViT model showing significant improvements in high-frequency prediction conditions, there remains room for optimization compared to its low-frequency performance. Future work should consider incorporating additional types of in-process monitoring data, such as thermography information, to enrich the data modalities and dimensions, thereby further enhancing the model's ability to recognize and model complex high-frequency magnetic characteristics.
- **Expansion of Sample and Material Diversity:** The current study has certain limitations regarding sample quantity and material types, which may impact the model's generalization capabilities. Future research should aim to cover more diverse magnetic alloys (e.g., silicon steel alloys) and expand the sample scale to strengthen the model's adaptability and stability under various process conditions.
- **Introduction of Cross-Frequency Learning Mechanisms:** Currently, the model requires independent training for each frequency and lacks the ability to share learning experiences across different frequencies. Subsequent research should explore mechanisms such as multitask learning, conditional modeling, or frequency embedding. These mechanisms would enable the model to build generalized capabilities across frequencies, thereby improving prediction

accuracy and adaptability under unknown frequency conditions.

In summary, this research, through innovative model design, effectively combined spatial awareness and cross-modal fusion strategies to achieve outstanding results in multi-frequency magnetic property prediction. In the future, the application can be further expanded to other fields such as the prediction of other functional material properties and the inverse estimation of process parameters, continuously extending the depth and breadth of smart manufacturing applications in materials research.

Author Contributions:
C J.-C. Huang: methodology,
C.-H. Chuang: validation, formal analysis
Y.-Y. Hsiao: writing—original draft preparation,
M.-C. Tsai: investigation andresources,
M.-H. Horng: investigation, supervision,writing and editing,
All authors have read and agreed to the published version of the manuscript.

Funding
The authors thank to the supports of National Science and Technology Concuil, ROC, NSC-111-2221-E-006-233-MY3, NSC-114-2221-E-006-079 and NSC-114-2218-E-006-020

Acknowledgments: no applicable
Data availability: Author declares that data used in this paper is available to this manuscript.
Corresponding author will share the data in a reasonable request.
Declarations Ethics Approval Not applicable.
Consent to participate Not applicable.
Consent to publish Not applicable.
Clinical Trail Not applicable.
Competing interests The authors declare no competing interests

REFERENCES

[1] Gibson D, Rosen B, Stucker M, Khorasani M (2021) Additive manufacturing technologies. 3rd edn. Springer Nature Switzerland AG. https://doi.org/10.1007/978-3-030-56127-7.
[2] Hamilton RF, Bimber BA, Palmer TA (2018) Correlating microstructure and superelasticity of directed energy deposition additive manufactured Ni-rich NiTi

alloys. J Alloys Compd 739:712–722. https://doi.org/10.1016/j.jallcom.2017.12.352.

[3] Wang C, Tan X, Du Z, Chandra S, Sun Z, Lim C, Tor S, Lim C, Wong C (2019) Additive manufacturing of NiTi shape memory alloys using pre-mixed powders. J Mater Process Technol 271:152–161. https://doi.org/10.1016/j.jmatprotec.2019.03.028.

[4] Li B, Wang L, Wang B, Li D, Oliveira J, Cui R, Yu J, Luo L, Chen R, Su Y et al (2022) Electron beam freeform fabrication of NiTi shape memory alloys: crystallography, martensitic transformation, and functional response. Mater Sci Eng A 843:143135. https://doi.org/10.1016/j.msea.2022.143135.

[5] Sing SL, Yeong WY (2020) Laser powder bed fusion for metal additive manufacturing: perspectives on recent developments. Virtual Phys Prototyp 15(3):359–370. https://doi.org/10.1080/17452759.2020.1779999.

[6] Yap CY, Chua CK, Dong ZL, Liu ZH, Zhang DQ, Loh LE, Sing SL (2015) Review of selective laser melting: materials and applications. Appl Phys Rev 2(4):041101. https://doi.org/10.1063/1.4935926.

[7] Liverani E, Toschi S, Ceschini L, Fortunato A (2017) Effect of selective laser melting (SLM) process parameters on microstructure and mechanical properties of 316L austenitic stainless steel. J Mater Sci Technol 33(4):376–385. https://doi.org/10.1016/j.jmst.2016.10.011.

[8] Liu Y, Li J, Xu K, Cheng T, Zhao D, Li W, Teng Q, Wei Q (2022) An optimized scanning strategy to mitigate excessive heat accumulation caused by short scanning lines in laser powder bed fusion process. Addit Manuf 60:103256. https://doi.org/10.1016/j.addma.2022.103256.

[9] Tucho WM, Lysne VH, Austbø H, Sjolyst-Kverneland A, Hansen V (2018) Investigation of effects of process parameters on microstructure and hardness of SLM manufactured SS316L. J Alloys Compd 740:910–925. https://doi.org/10.1016/j.jallcom.2017.12.261.

[10] Giganto S, Zapico P, Castro-Sastre M, Martínez-Pellitero S, Leo P, Perulli P (2019) Influence of the scanning strategy parameters upon the quality of the SLM parts. Procedia Manuf 41:698–705. https://doi.org/10.1016/j.promfg.2019.09.060

[11] Yang J, Zhu Q, Wang Z, Xiong F, Li Q, Yang F, Li R, Ge X, Wang M (2023) Effects of metallurgical defects on magnetic properties of SLM NiFeMo permalloy. Mater Charact 197:112672. https://doi.org/10.1016/j.matchar.2023.112672

[12] Liu S, Yang W, Shi X, Li B, Duan S, Guo H, Guo J (2019) Influence of laser process parameters on the densification, microstructure, and mechanical properties of a selective laser melted AZ61 magnesium alloy. J Alloys Compd 808:151713. https://doi.org/10.1016/j.jallcom.2019.06.261

[13] Gong H, Rafi K, Gu H, Ram GJ, Starr T, Stucker B (2015) Influence of defects on mechanical properties of Ti–6Al–4V components produced by selective laser melting and electron beam melting. Mater Des 86:545–554. https://doi.org/10.1016/j.matdes.2015.07.110

[14] Chang LK, Chen RS, Tsai MC, Lee RM, Lin CC, Huang JC, Chang TW, Horng MH (2024) Machine learning applied to property prediction of metal additive manufacturing products with textural feature extraction. Int J Adv Manuf Technol 132(1–4):1–15. https://doi.org/10.1007/s00170-023-12864-1

[15] Drucker H, Burges CJC, Kaufman L, Smola A, Vapnik V (1997) Support vector regression machines. Adv Neural Inf Process Syst 9:155–161

[16] Chen T, Guestrin C (2016) XGBoost: a scalable tree boosting system. In: Proc 22nd ACM SIGKDD Int Conf Knowl Discov Data Min. pp 785–794. https://doi.org/10.1145/2939672.2939785

[17] Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu TY (2017) LightGBM: a highly efficient gradient boosting decision tree. Adv Neural Inf Process Syst 30:3146–3154

[18] Chang LK, Chang TW, Huang JC, Liu DQ, Tsai MC, Horng MH (2025) Additive manufacturing section image features for magnetic processing characteristics prediction. Key Eng Mater 1016:99–104. https://doi.org/10.4028/p-9k8n1f

[19] Weimer D, Scholz-Reiter B, Shpitalni M (2016) Design of deep convolutional neural network architectures for automated feature extraction in industrial inspection. CIRP Ann 65(1):417–420. https://doi.org/10.1016/j.cirp.2016.04.072

[20] Chang LK, Chen PC, Tsai MC et al (2025) Enhancing prediction of magnetic properties in additive manufacturing products through a 3D convolutional vision transformer model. Int J Adv Manuf Technol 137:4503–4519. https://doi.org/10.1007/s00170-025-15381-6

[21] Wu H, Xiao B, Codella N, Liu M, Dai X, Yuan L, Zhang L (2021) CvT: introducing convolutions to vision transformers. In: Proc IEEE/CVF Int Conf Comput Vis. pp 22–31

[22] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J, Houlsby N (2021) An image is worth 16×16 words: transformers for image recognition at scale. In: Proc Int Conf Learn Represent. arXiv:2010.11929

[23] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I (2017) Attention is all you need. Adv Neural Inf Process Syst 30:5998–6008

[24] LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. Proc IEEE 86(11):2278–2324.

https://doi.org/10.1109/5.726791

[25] Chen Y, Liu Z, Zhang Z, Wang H, Wang L (2023) Progressive sampling for vision transformers. In: Proc IEEE/CVF Conf Comput Vis Pattern Recognit. pp 1–10

[26] Tran D, Bourdev L, Fergus R, Torresani L, Paluri M (2015) Learning spatiotemporal features with 3D convolutional networks. In: Proc IEEE Int Conf Comput Vis. pp 4489–4497

[27] He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proc IEEE Conf Comput Vis Pattern Recognit. pp 770–778. https://arxiv.org/abs/1512.03385

[28] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: Proc IEEE Conf Comput Vis Pattern Recognit. pp 4700–4708

[29] Fiorillo F (2004) Characterization and measurement of magnetic materials. Academic Press. https://doi.org/10.1016/B978-0-12-257251-7.X5000-X