MAG-ViT: A ViT Backbone for Magnetic Material Modeling

Rui Zhang, Hao Song, Lie Zhang, Yibo Lu,and Lijun Hang, *Senior Member, IEEE,*

***Abstract*—Classical magnetic core loss modeling methods have limitations in terms of accuracy and comprehensiveness. Existing deep learning modeling methods also suffer from limited data size and overly simple models. Vision transformers (ViT) have shown promise in various vision tasks. In this study, we proposed MAG-ViT ,which use ViT as backbone for magnetic material modeling. MAG-ViT extends the magnetic material model to deeper dimensions in a more sensible way while using less memory through patch embedding, class token and learnable positional embeddings method. The open source MAGNET dataset was utilised for training and evaluating of MAG-ViT. The experimental results show that MAG-ViT performs well on the core loss prediction task and it ranks 8th in pretest results. Our test scripts and models are available at https://github.com/moeKedama/dg-magnet-test-script.**

***Index Terms*—deep learning, power magnetics, core loss, hysteresis loop, vision transformer (ViT).**

I. Introduction

MAGNETIC components are an important part of power electronics systems for energy buffering. The design of magnetic components relies heavily on classical loss models for magnetic materials and material datasheets tailored to specific operating conditions, such as waveforms, frequency, and temperature. Classical magnetic core loss models, including the $Steinmetz\ equation$, $iGSE$ [1], rely on empirical simplifications and physical approximations. Despite numerous upgrades since the introduction of the Steinmetz equation in 1890, these models still have limitations regarding accuracy and comprehensiveness [2]–[4]. These models may introduce significant errors by neglecting the intricate behavior of magnetic materials and their variations under diverse operating conditions. However, manufacturers' datasheets, while providing more accurate data, only cover a limited range of operating conditions. Acquiring data for all conceivable operating conditions would require significant testing efforts and resources. Obtaining an accurate and general model with a small amount of experimental data is the crux of the problem.

In recent years, there has been an increasing interest in modeling the magnetic materials with neural networks [5]. Training neural networks using both experimental and simulation data enables the construction of models for the behavior of magnetic materials. These models find applications in the design phase, aiding in the optimization and enhancement of magnetic components. Neural-network-based methods, which operate independently of physical models, demonstrate superior generalization capabilities. They excel in capturing the intricacies of complex magnetic material behavior and various operating conditions. In addition, these methods provide better coverage of the overall data distribution compared to classical datasheets. Studies have demonstrated that neural-network-based methods incur lower computational overhead compared to Preisach model-based methods [6]. Nonetheless, many existing methods are constrained by small dataset sizes, and their model structures are often too simplistic to exhibit effective generalization when meeting real-world task requirements. These challenges can be addressed by leveraging the newly developed open-source dataset, MAGNET [7]–[9]. Leveraging the MAGNET dataset facilitates greater comparability of results across related studies compared to development on disparate datasets with unknown data quality, limited size, and singular operational situations.

The Transformer has consistently stood as the state-of-the-art approach for numerous sequence-to-sequence tasks since its inception [10]. It utilizes a self-attention mechanism to capture temporal relationships between input and output sequences. Despite the limitations of the classical Transformer model in handling inputs, such as insufficient processing of local information and fixed size requirements, Vision Transformer (ViT) addresses these challenges [11]. ViT closely adheres to the original Transformer while introducing a simple yet highly extensible structure. It demonstrates out-of-the-box efficiency, providing an improved solution for handling large inputs and capturing essential local details.

We propose a ViT-based modeling method for hysteresis loop and core loss modeling, called MAG-ViT. MAG-ViT exhibits a significant improvement over Transformer-based methods, boasting reduced memory consumption, enhanced modeling accuracy, and greater scalability.

The paper is structured as follows: Section II details the specific modeling approach employed by MAG-ViT; Section III outlines the experimental results obtained with MAG-ViT on the MAGNET dataset; Section IV concludes this paper by summarizing the presented work and suggesting directions for future exploration.

II. MAG-ViT

ViT operates by segmenting the input into fixed-size patches and embedding them into vectors. It also leverages the self-attention mechanism to capture global and local relationships within the input. In order to apply ViT to magnetic material modeling, it is crucial to preprocessing the magnetic material input. The model structure is also adjusted to fit the specific modeling requirements. Here, we introduce MAG-ViT, which

This paper was produced by the team of Hangzhou Dianzi University.

Our Github repository: https://github.com/moeKedama/dg-magnet-test-script

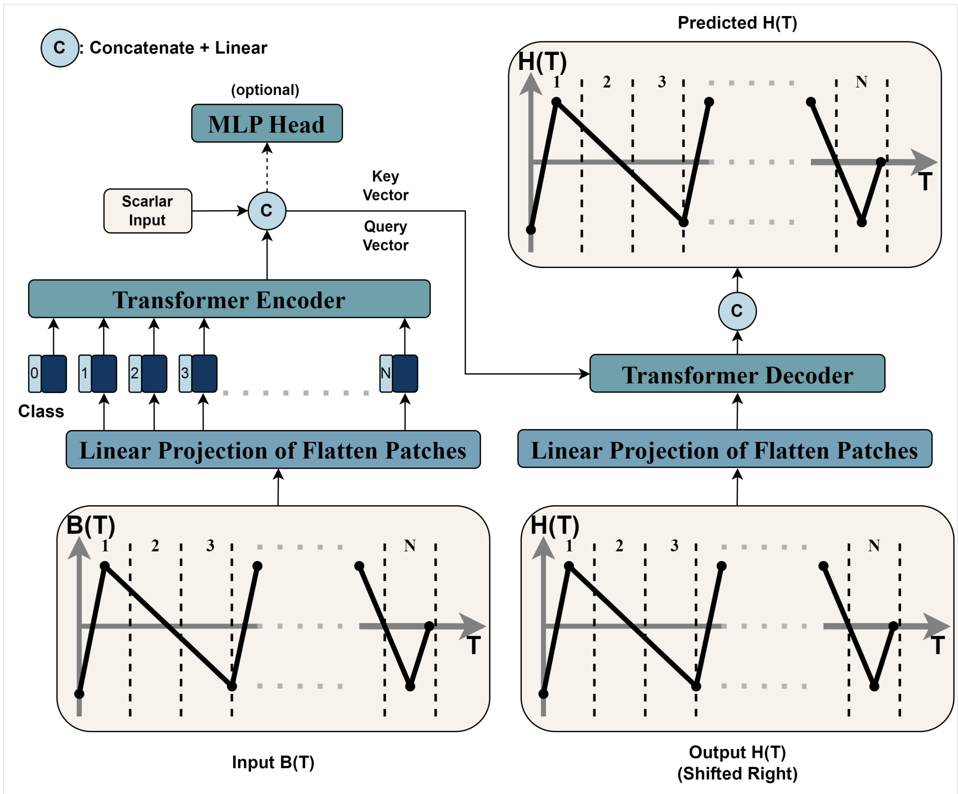

Fig. 1. Model overview and data flow of MAG-ViT.

incorporates 3 major improvements over the transformer approach.

1) **Patch Embedding:** The input to the general Transformer comprises a 1D sequence of token embeddings. To manage the flux density sequence $B(t)$ $x \in R$, we initially employ a Multi-Layer Perceptron (MLP) to linearly project $B(t)$ to C dimensions. Subsequently, the projected $B(t)$ sequence, with a C-dimensional latent vector size, undergoes transformation into a sequence of flattened 2-D patches with $x_p \in N{*}(p^2{*}C)$. In this context, the input sequence can be conceptualized as a sequence of flattened 2-D blocks. The total count of flattened 2-D blocks is $N = R/P^2$ and the dimension of each block is $(P^2 * C)$, where P denotes the size of the block.

2) **Class Token:** A BERT-like class token, represented by the learnable embedding x_{class} is introduced to the sequence of embedded patches. x_{class} shares the same dimensions as the input sequence $B(t)$, resulting in an output of $N + 1$ vectors. Functioning akin to the average pooling layer in a Convolutional Neural Network (CNN), the class token integrates information across the model input tokens. Similar to the role of the average pooling layer in CNN, which is utilized for synthesizing information from input tokens, the class token plays a crucial role. After interacting with each patch for information, it enables the model to learn specific classification information. Additionally, within the self-attention mechanism, the interaction between the class token and various patches is contingent on the degree of interaction. This approach allows the model to discern which patches exert influence on the final classification results. Moreover, it enables the model to determine the specific degree of that influence, thereby enhancing the interpretability of the model.

3) **Learnable Positional Embeddings:** Positional embeddings are imperative for Transformers, given that altering the token sequence order has no impact on the outcome. The absence of positional embeddings increases the learning cost as the model, lacking positional information, resorts to relying solely on the semantics of the patches. As an enhancement, standard learnable 1D positional embeddings, denoted as E_{pos} are incorporated as a substitute for absolute positional encodings. The resulting sequence of embedding vectors, inclusive of positional information, is then fed into the encoder. This modification contributes to the model's ability to understand and utilize the positional relationships within the sequence.

Equation (1) delineates the complete process of preprocessing $B(t)$ and $H(t)$ sequence

$$z_0 = [x_{class}; x_p^1 E; x_p^2 E; \cdots ; x_p^N E] + E_{pos}, \tag{1}$$

where $x_p \in (x_p^1, x_p^2, \ldots x_p^N)$ denotes the sequence patch,and E denotes the linearly projection operation.

The network structure of MAG-ViT is depicted in Fig. 1. Initially, the data points of each time step in the input sequence $B(t)$ undergo transformation into a representation of patch embeddings with dimension D and length $N{+}1$. This transformation is achieved through the linear projection of the flattened patches block. The resulting representation encompasses class tokens and learnable positional embeddings. Subsequently, the generated patch embeddings are inputted into the Transformer Encoder block. This block analyzes and captures the temporal

TABLE I
MAXIMUM PATH LENGTHS, PER-LAYER COMPLEXITY AND MINIMUM NUMBER OF SEQUENTIAL OPERATIONS FOR DIFFERENT LAYER TYPES

Layer Type	Complexity per Layer	Sequential Operations	Maximum Path Length
Convolution	$O(k \cdot n \cdot d^2)$	$O(1)$	$O(\log_k n)$
Recurrent	$O(n \cdot d^2)$	$O(n)$	$O(n)$
Self-Attention (Transformer)	$O(n^2 \cdot d)$	$O(1)$	$O(1)$
Patched Self-Attention (ViT)	$O(\frac{n^2}{patch\_lenth} \cdot d)$	$O(1)$	$O(\frac{n}{patch\_lenth})$

dependencies within the input sequences, yielding sequence-informative query vectors and key vectors.

The query vector and key vector are concatenated with scalar inputs (i.e., temperature, frequency, peak-to-peak value of $B(t)$). The concatenated vectors are then mapped back to their original dimensions through a linear projection layer. These vectors are further transformed to the dimensions of the original query vector and key vector, serving as new query vector and key vector inputs to the Transformer Decoder. Also, an optional MLP Head can be set up here for classification or regression tasks. The Transformer Decoder is responsible for reconstructing the output sequence.

For the implementation of output sequence reconstruction, the Transformer Decoder requires a value vector. During the network training phase, the value vector is generated using the sequence $H(t)$ as input. In the network testing phase, the value vector is cyclically generated, with the predicted output sequence initialized to 0 as input. Subsequent to further processing by the Transformer Decoder, the model produces a patched output $H(t)$ sequence with dimensions D and length $N + 1$.

The initial N patches of the sequence are extracted, and their dimensions are transformed from D to p through a linear projection. Each D-dimensional patch contains information about p original data points. The final prediction $H(t)$ is obtained by concatenating these patched sequences in order.

In contrast to the down-sampling methods of example, MAG-ViT aims to retain maximal information at high frequencies while maintaining an acceptable complexity level. It achieves this by encapsulating high-frequency features within patches and reflecting low-frequency features between patches. This model possesses a higher performance upper bound compared to down-sampling methods, albeit at the cost of increased training expenses.

It is essential to highlight that the inductive bias of ViT is notably smaller compared to Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). CNNs exhibit properties such as locality, spatial invariance, translational equivariance, whereas RNNs possess properties like sequential, time invariance. In the ViT-based approach, only the MLP layers are local and translationally equivariant, while the self-attention layers operate globally. The positional embeddings must be trained to be sufficiently expressive to handle relationships between entities effectively. This requirement also implies that the dataset used for training needs to attain a certain size.

Table I is maximum path lengths, per-layer complexity and minimum number of sequential operations for different layer types, where n is the sequence length, d is the representation dimension and k is the kernel size of convolutions. Compared to Self-Attention, Patched Self-Attention has a smaller per-layer complexity.

TABLE II
THE NUMBERS OF PARAMETERS IN THE MODEL

Model	Number of Parameters
Simple	2,396,048
Full(With MLP Head)	2,528,913

III. EXPERIMENTS

We evaluated the proposed MAG-ViT in the hysteresis loop and core loss modelling tasks. Before presenting these results, we outline the main experimental setup below. Additional details can be found in our GitHub repository.

A. Experimental Setup

Datasets. For class-conditional learning, we consider the MAGNET dataset [magnet citation], which contains 10 different ferrite materials across a wide MAGNET dataset contains 4 fields which are flux density waveform $B(t)$, field strength waveform $H(t)$, fundamental frequency f, and temperature T. In the preprocessing stage, normalization is applied to all fields within the MAGNET dataset. The normalization parameters are preserved for subsequent testing and inference.

MAG-ViT configuration. The dimension of the linear projection output, denoted as dim_val, is set to 32. For the original input $B(t)$ with a length of 1024, the shape after linear projection becomes (1024, 8). To maintain a balance between the length of the patch embedding sequence and the total number of tokens in each patch, the patch size utilized by MAG-ViT is configured as (8, 32). Consequently, for the input $B(t)$ after linear projection, the corresponding output patch embedding shape is (128, 256). In this context, every 8 original data points projected to 32 dimensions form a patch, and each patch comprises 256 tokens (represented by $embed\_dim$ in the configuration). Concerning the transformer configuration, the number of neurons in the linear layer of both the transformer encoder and decoder is set to 256. The number of attention heads is 8, and there is a single stacked encoder layer in both the encoder and decoder. Following this configuration, the total number of parameters in the models are shown in TABLE II. Both 2 models are under the 10MB parameter limitations.

Training. We employ the Adam optimizer with an initialized learning rate of 0.0001 for all materials. All models undergo

TABLE III
Core Loss Prediction Relative Errors Of The MAG-ViT

Model	AVG	95-Prct	99-Prct	Max
3C90	2.052%	**6.38%**	11%	29.66%
3C94	1.834%	**5.64%**	10.4%	44.36%
3E6	0.6745%	**1.56%**	2.24%	7.097%
3F4	3.673%	**11.4%**	21.1%	50.26%
77	1.988%	**4.77%**	8.19%	19.74%
78	1.917%	**5.65%**	9.78%	28.83%
N27	1.668%	**5.33%**	8.95%	27.24%
N30	0.6371%	**1.6%**	2.58%	19.39%
N49	3.059%	**10.4%**	18.5%	57.77%
N87	1.58%	**4.77%**	8.44%	34.98%
average	1.908%	**5.75%**	10.118%	31.9327%

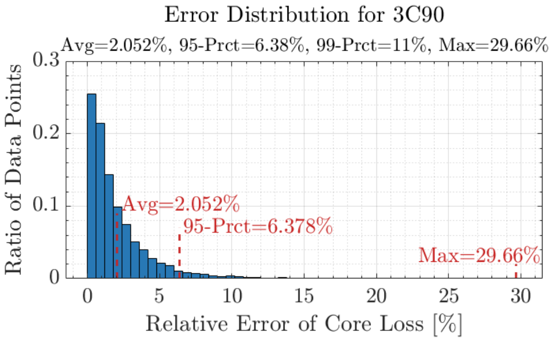

Fig. 2. Core Loss Prediction Relative Errors of Material 3C90

training for 4000 epochs, with a manual reduction of the learning rate to 0.1x at 2700 and 3600 epochs. The training process occurs at a rate of approximately 200 epochs per 5 minutes on a PC equipped with a 5.80GHz i9-13900k CPU, 128GB DDR5 4200MHz RAM, and an NVIDIA RTX A6000 GPU. The model consumes approximately 0.6GB of VRAM with a batch size of 32. MAG-ViT employs mean-square error (MSE) as the loss function for training. Each material is trained separately, resulting in 1 models for each material. To achieve the results presented in the Pretest Results, convergence should be at least on the order of 10e-5 to 10e-6 on the training set. Moreover, it should be at least on the order of 10e-4 to 10e-5 on the validation set, which does not participate in the training process.

B. Results on MAGNET pretest Dataset

The MAGNET pretest Dataset serves as a validation set accompanying the MAGNET dataset, encompassing 10 existing materials, each comprising 5,000 randomly sampled data points from the original database.

In the Core Loss Prediction task, the predicted core loss $P_{V_{pred}}$ can be directly computed through integrals with respect to the original input $B(t)$ and the predicted $H(t)$. Subsequently, the relative error between the predicted core loss $P_{V_{pred}}$ and the measured $P_{V_{mea}}$ is employed as an evaluation metric.

$$P_V = \frac{1}{T}\int_{B(0)}^{B(t)} H(t)dB(t) \tag{2}$$

TABLE IV
The Sub-domains Test Of The MAG-ViT

Model	AVG	95-Prct	99-Prct	Max
3C90	13.07%	37.2%	144%	344.7%
3C94	12.2%	34.5%	129%	383%
3E6	5.21%	19.4%	46.7%	90.4%
3F4	17.49%	62.2%	119%	407.5%
Sub1 avg	12.0175%	38.325%	109.675%	306.4%
77	2.569%	7.18%	10.1%	27.69%
78	3.591%	10.7%	15.5%	33.49%
Sub2 avg	3.08%	8.94%	12.8%	30.59%
N27	22.35%	70.9%	113%	346.2%
N30	16.8%	54.2%	75.3%	268.9%
N49	36.26%	138%	274%	599%
N87	20.3%	62%	108%	338.8%
Sub3 avg	23.9275%	81.275%	142.575%	388.225%
average	14.994%	49.628%	103.46%	283.968%

$$Relative\ Error = \frac{|P_{V_{pred}} - P_{V_{mea}}|}{P_{V_{mea}}} \tag{3}$$

Table III lists the results of our study of the 10 materials in the dataset. Here, AVG represents the average relative error, 95-Prct denotes the 95th percentile relative error, 99-Prct represents the 99th percentile relative error, and Max signifies the maximum relative error.

Taking 95-Prct-Error as an example, we achieved the best results by reaching 1.56% on material 3E6. This ranked us 3rd out of all 25 teams that submitted pre-test results. On the worst-performing material, 3F4 reached 11.39% and ranked 9th. The average 95-Prct-Error for all materials combined ranked 8th.

Fig. 2 displays the details of our study on material 3C90. The majority of the data points have an error of less than 5%. This indicates that the model is generally accurate in its predictions of the core loss. However, there is a small number of data points with a much higher error. These data points may indicate that the model is not as accurate for some types of data or they may be outliers.

After evaluating the model on 10 materials, we categorized the materials into 3 sub-domains.

- sub-domain 1 : 3C90, 3C94, 3E6, 3F4.
- sub-domain 2 : 77, 78.
- sub-domain 3 : N27, N30, N49, N87.

The division of the 3 sub-domains is based on potential similarities between the 10 materials. We posit that certain weights in the network can be shared among similar materials, particularly in the encoding layers. To test this hypothesis, we trained the sub-domain data using the MAG-ViT configuration, only altering the number of stacked encoder layers from 1 to 2. The modified model has a total of 3,586,458 parameters. Each sub-domain is trained separately.

Table IV presents the results of the MAG-ViT sub-domain tests. Materials in sub-domain 2 achieved similar results to those in Table III, whereas materials in sub-domain 1 and sub-domain 3 performed poorly in the sub-domain tests. Several possible causes are considered:

1) Parameters Limitations: In the Magnet Challenge, the size of all included models was restricted to 10 MB for one material. Despite using 14.2 MB parameters in the sub-domain test, we did not achieve satisfactory results, except for sub-domain 2. We speculate that more similar materials require fewer parameters, and sub-domains with a greater number of materials necessitate more parameters. Model quantization and model distillation methods can be employed to conserve resources in the model deployment phase. The 10MB limit may be too stringent for research-stage models, especially when the model needs to encode multiple materials.

2) Data normalization methods: The current normalization method involves computing means and standard deviations across all domains or sub-domains, relying on prior knowledge. However, when there is more than one material in the training data, this method may not normalize the data effectively. As a result, the model can face challenges in convergence, particularly in cases where prior knowledge is not fully applicable.

3) Lack of inductive bias: ViT's performance is constrained on small and medium-sized datasets due to the lack of inductive bias. The Magnet dataset, comparable in size to datasets like MNIST and CIFAR, which are considered small datasets, faces limitations in this context. Larger dataset sizes are crucial for advancing future research in this domain.

IV. Conclusion

This work introduces MAG-ViT, a straightforward and versatile ViT-based architecture for magnetic material modeling. MAG-ViT processes sequential and scalar inputs to construct embeddings and is assessed in tasks such as Hysteresis B-H Loop Prediction and Core Loss Prediction. Experimental results indicate that MAG-ViT exhibits promising performance. We believe that MAG-ViT can offer valuable insights for future research on backbones in magnetic material modeling and contribute to endeavors such as transfer learning in this field.

Acknowledgments

This paper constitutes the final submission of the Hangzhou Dianzi University team for the IEEE PELS-Google-Enphase-Princeton MagNet Challenge. The open-sourced dataset utilized in this work is sourced from the MagNet Team.

References

[1] J. Goodenough, "Summary of losses in magnetic materials," *IEEE Transactions on Magnetics*, vol. 38, no. 5, pp. 3398–3408, 2002.

[2] J. Muhlethaler, J. Biela, J. W. Kolar, and A. Ecklebe, "Improved core-loss calculation for magnetic components employed in power electronic systems," *IEEE Transactions on Power Electronics*, vol. 27, no. 2, pp. 964–973, 2012.

[3] Y. Han, G. Cheung, A. Li, C. R. Sullivan, and D. J. Perreault, "Evaluation of magnetic materials for very high frequency power applications," *IEEE Transactions on Power Electronics*, vol. 27, no. 1, pp. 425–435,2012.

[4] M. Chen, M. Araghchini, K. K. Afridi, J. H. Lang, C. R. Sullivan, and D. J. Perreault, "A systematic approach to modeling impedances and current distribution in planar magnetics," *IEEE Transactions on Power Electronics*, vol. 31, no. 1, pp. 560–580, 2016.

[5] Y. LeCun, Yann Bengio and G. Hinton, "Deep learning," *Nature*, vol. 521, p. 436–444, 2015.

[6] S. Quondam Antonio, F. Riganti Fulginei, A. Laudani, A. Faba, and E. Cardelli, "An effective neural network approach to reproduce magnetic hysteresis in electrical steel under arbitrary excitation waveforms," *Journal of Magnetism and Magnetic Materials*, vol. 528, p. 167735, 2021.

[7] H. Li, D. Serrano, S. Wang, and M. Chen, "MagNet-AI: Neural Network as Datasheet for Magnetics Modeling and Material Recommendation," *IEEE Transactions on Power Electronics*, vol. 38, no. 12, pp. 15854–15869, Dec. 2023.

[8] D. Serrano, H. Li, S. Wang, M. Luo, T. Guillod, C. R. Sullivan, and M. Chen, "Why MagNet: Quantifying the Complexity of Modeling Power Magnetic Material Characteristics," *IEEE Transactions on Power Electronics*, vol. 38, no. 11, pp. 14292–14316, Nov. 2023.

[9] H. Li, D. Serrano, T. Guillod, S. Wang, E. Dogariu, A. Nadler, M. Luo, V. Bansal, N. Jha, Y. Chen, C. R. Sullivan, and M. Chen, "How MagNet: Machine Learning Framework for Modeling Power Magnetic Material Characteristics," *IEEE Transactions on Power Electronics*, vol. 38, no. 12, pp. 15829–15853, Dec. 2023.

[10] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[11] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. H. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, ":An image is worth 16×16 words: Transformers for image recognition at scale," in *Proc. Int. Conf. Learn. Reinforcement (ICLR)*, Vienna, Austria, 2021, pp. 1–22.