Attention Is Magnetic Core Loss Modelling Need?

Yaohua Li, *Student Member, IEEE*, Sicheng Wang, *Student Member, IEEE*, Yue Wu, *Student Member, IEEE*, Yongbin Jiang, *Member, IEEE*, Ziheng Xiao, *Member, IEEE*, and Yi Tang, *Senior Member, IEEE*

***Abstract*—Given the complex mechanism of magnetic core behaviors, the traditional equation-based models are usually difficult to achieve high accuracy under various excitations and operating conditions. In this report, we explore the potential of state-of-the-art transformer models in magnetic core loss modelling. The result shows that the proposed CNN-transformer model can achieve a higher accuracy than the existed transformer model without sacrificing the computational efficiency and parameter size. However, the attention-based transformer model needs abundant training data to achieve a good generalization ability and predicted performance. Therefore, for a small dataset, the attention mechanism may not be a good choice for magnetic core loss modelling.**

***Index Terms*—Magnetic core loss modelling, Transformer, CNN, MagNet.**

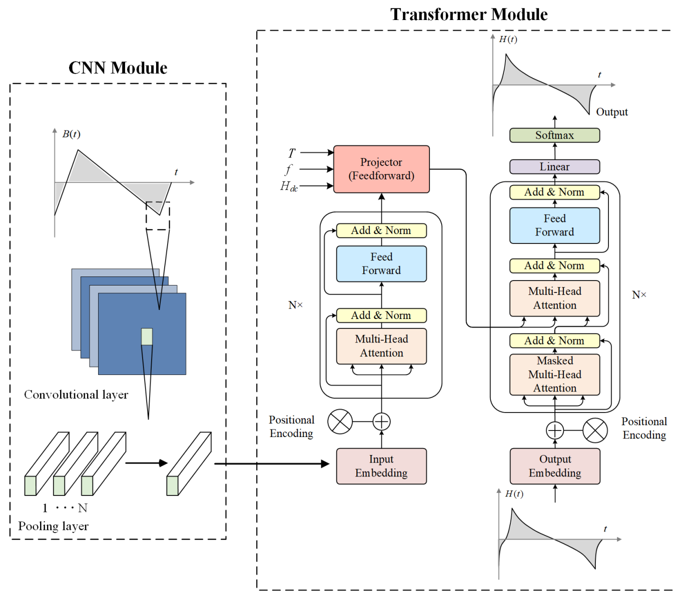

Fig. 1. Proposed CNN-Transformer structure.

I. Introduction

DEEP learning provides a powerful tool for magnetic core loss modelling. Compared with traditional equation-based models, data-driven models can achieve higher accuracy and better generalization ability in consideration of sufficient training data and model parameters.

Deep learning-based prediction and recognition technologies commonly utilize Convolutional Neural Networks (CNNs) for the purpose of feature learning and classification. The use of convolution operations in CNN models facilitates the sharing of parameters, thereby reducing the total number of parameters that require training. This effectively diminishes the complexity of the model. CNNs possess the capability to autonomously learn representations of input data features, which significantly lightens the load of manual feature extraction. Nonetheless, when dealing with prediction or recognition subjects with intricate features, it becomes essential to augment the number of convolution layers in the CNN to grasp more sophisticated and advanced features. This, however, results in increased computational expenses.

Over the past few years, Transformer models have made significant strides in both natural language processing and computer vision. These models excel at identifying long-term dependencies in input sequences, facilitating comprehensive global feature extraction. The inherent parallel processing capability of Transformers enhances their efficiency in handling input sequences, thereby accelerating both training and inference processes. However, their reliance on self-attention mechanisms results in a reduced capacity for in-depth feature exploration, which in turn elevates the model's complexity and necessitates higher quality input data.

Due to the highly nonlinear characteristics of magnetic materials, we adopt a strategy that combines Transformers and CNNs to compensate for the drawbacks of both mainstream deep learning models. Specifically, a CNN layer is used as the regression head, while the Transformer serves as the backbone of the model. This modeling approach takes into account the complementary nature of global and local information, which helps to significantly reduce training time while enhancing the model's multi-scale feature extraction capabilities.

II. Model Architectures

During the entire Magnet Challenge process, we experimented with the following types of models.

A. Vision Transformer Model

The Vision Transformer (ViT) is essentially tailored for visual tasks, utilizing a Transformer-based architecture. It operates on images by breaking them down into numerous small pieces, referred to as "patches," and then handling these patches in a sequential manner. This approach is same as how traditional Transformers process textual data, considering each word in a sequence for natural language tasks. In this case, the input magnetic flux denstiy B(t) is treated as a one-dimensional(1-D) image for ViT, where it undergoes patch embedding before being fed through a Transformer structure, enabling the prediction of the magnetic field intensity sequence H(t). However, our experiments indicate that ViT's head is relatively intricate, and it only performs comparably to a standalone Transformer when provided with sufficient input data.

TABLE I
COMPARISON OF DIFFERENT MODELS WITH BASELINE

Name	Model	Parameters	CNN In	Transformer In	Decoder Layer	Average Error	95-Prct Error	Running Time
Baseline	Transformer	28481	NA	128*1	1	3.05%	9.64%	129mins
Model 1	Transformer	28481	NA	256*1	1	2.47%	8.38%	304mins
Model 2	CNN+Transformer	28564	256	128*2	1	2.4%	7.94%	152mins
Model 3	CNN+Transformer	28564	512	128*2	1	2.07%	6.97%	152mins
Model 4	CNN+Transformer	62788	512	128*2	4	1.45%	4.72%	502mins

Furthermore, segmenting the input into multiple patches can disrupt the inherent relationships within the input sequences. Adequate parameters are crucial to ensure sufficient feature learning, which is necessary to reduce prediction errors.

B. Long Short-Term Memory Network Model

LSTM stands out as an apt solution when dealing with shorter data sequences or in environments with constrained resources. Yet, in light of the complex nonlinear attributes of magnetic materials, LSTM's capacity for parallel processing is markedly less effective compared to that of transformers. Moreover, transformers present a broader horizon for improvements. As the database expands with increasing amounts of data, the advantages of transformers will become more distinctly observable.

C. CNN-Transformer Model

Based on our experience with ViT experiments, our next consideration is how to replace the head of ViT with a more lightweight structure that can better extract features from the input sequence. Undoubtedly, CNN offers such a capability. The input sequence can first be transformed into a multi-dimensional sequence through CNN before entering the transformer. This approach not only allows for the input of more sampling points to enhance the model's prediction accuracy but also reduces the model's runtime.

III. EXPERIMENTAL RESULTS

Tables I and Fig. 2 illustrate the training performance of five different models on the magnetic material N87. All models are tested on the same GPU(Nvidia RTX 4090). The Baseline model, the Transformer provided from the MagNet Challenge tutorial, stands out for its shortest training time, showcasing the parallel processing advantage of Transformers. However, this model also exhibits the highest error, likely due to Transformers' dependency on extensive and high-quality labeled data, a limitation in certain applications and tasks. Model 1 indicates that expanding the number of sequences fed into the transformer is able to increase the accuracy of predicted core loss, though at the cost of doubling the runtime. This led us to explore balancing enhanced accuracy with reduced runtime. Model 2 reveals that incorporating an additional CNN layer not only furthers accuracy improvement but also substantially cuts down simulation time, nearing Baseline levels. Leveraging CNN's robust feature extraction, Model 3 increases input size from 256 to 512, resulting in a further drop in the 95th

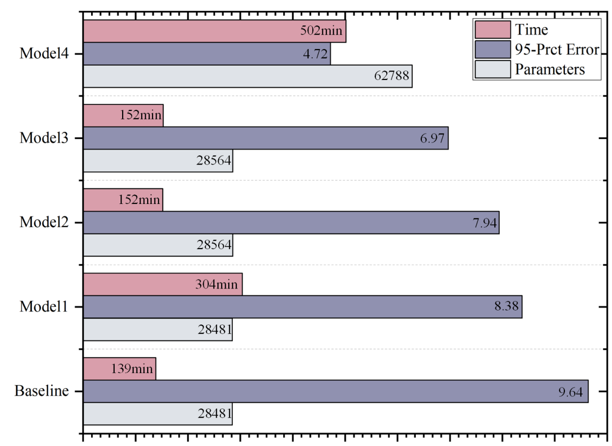

Fig. 2. Comparison of runtime, 95-Prct error and parameters for five models.

percentile error while maintaining the same runtime as Model 2. In summary, Model 3 optimally demonstrates the advantages of a combined model approach, aligning well with our goals.

To push for even lower error rates, Model 4, an extension of Model 3, upscales the parameter count from 28,564 to 62,788. Comparative data from Fig. 2 clearly shows that a higher parameter count markedly lowers the error rate, but also leads to a significant increase in computational time.

IV. DISCUSSION

A. Model Size and Model Parameters

The model size and model parameters are two important factors that affect the performance of the model. For attention-based transformer model, the model size is crucial than CNN or LSTM model. Based on our experimental results, the model size of transformer should be no less than 10000 and the parameters should be around 30000. In this case, the accuracy and runtime of the model can be guaranteed.

B. Transfer Learning Ability

Although Transformer models excel in many aspects, their success in transfer learning largely depends on factors such as the amount of data, the similarity between materials, and model adjustment strategies. The low performance of our model in transfer learning can be attributed to several reasons. Firstly, CNN-Transformer model has a strong dependency on data. In transfer learning scenarios with limited data, Transformers may struggle to effectively learn key features of the new magnetic material, leading to decreased performance. Secondly, when there is a significant difference in data distribution or features between the source and target tasks, Transformers might find it challenging to transfer the learned knowledge effectively. This issue arises because their self-attention mechanism might overly focus on specific patterns from the source task that are irrelevant or misleading for the new task.

Moreover, due to their large parameter count, Transformer models are prone to overfitting on limited datasets, making it difficult to adapt to new tasks. Furthermore, effective transfer learning involves not just the transfer of model parameters but also requires fine-tuning according to the characteristics of the target task. Without targeted adjustments, Transformers might not perform optimally on the new task.

V. Conclusion

In conclusion, this report has explored the effectiveness of advanced transformer models in the realm of magnetic core loss modeling, addressing the limitations of traditional equation-based models that struggle with accuracy under diverse excitations and operating conditions. The results clearly demonstrate that the proposed CNN-transformer model outperforms existing transformer models in terms of accuracy, while maintaining computational efficiency and a manageable parameter size. However, it is important to note that the attention-based transformer model, despite its strengths, requires a substantial amount of training data to develop robust generalization capabilities and achieve high predictive performance. Consequently, when dealing with smaller datasets, the attention mechanism might not be the most suitable approach for magnetic core loss modeling. This insight is crucial for guiding future research and application of machine learning techniques in magnetic core modeling, especially in scenarios where data availability is limited.

Acknowledgments

We would like to express our gratitude to organizers for organizing the MagNet Challenge 2023 and providing valuable equation-based insights and machine learning based solutions. We sincerely hope that through our globally collective efforts, we can make advancements in the modelling of magnetic losses. Thank you for this opportunity to collaborate and work towards furthering our understanding in this field.