Image-based Deep Learning Models for Electric Motors

Sun, Siyuan; Wang, Ye; Koike-Akino, Toshiaki; Yamamoto, Tatsuya; Sakamoto, Yusuke; Wang, Bingnan

TR2025-088 June 24, 2025

Abstract

Recently, there are a lot efforts in developing machine learning and deep learning methods for the prediction of electric motor performances. In particular, torque profile for a given motor design, including cogging torque and torque ripple, are challenging for surrogate models to achieve high prediction accuracy. One promising approach is to represent a motor design as a 2d image, and utilize deep learning models that found success in image recognition and classification tasks for motor performance prediction. A number of deep convolutional neural networks (CNNs) have been adopted for motor applications previously, while more recently Vision Transformer (ViT) models are gaining interests. In this paper, we evaluate multiple deep CNN and ViT models on two datasets of interior permanent magnet motors, and show that ViT based models can achieve superior accuracy for cogging torque prediction compared with CNN based models, and can be jointly trained on the two different datasets and still provides prediction with low error.

International Conference on the Computation of Electromagnetic Fields (COMPUMAG) 2025

Image-based Deep Learning Models for Electric Motors

Siyuan Sun[1], Ye Wang[1], Toshiaki Koike-Akino[1], Tatsuya Yamamoto[2], Yusuke Sakamoto[2], and Bingnan Wang[1]

[1] Mitsubishi Electric Research Laboratories, Cambridge, MA 02139 USA

[3] Advanced Technology R&D Center, Mitsubishi Electric Corporation, Amagasaki, Hyogo 661-8661 Japan

Recently, there are a lot efforts in developing machine learning and deep learning methods for the prediction of electric motor performances. In particular, torque profile for a given motor design, including cogging torque and torque ripple, are challenging for surrogate models to achieve high prediction accuracy. One promising approach is to represent a motor design as a 2d image, and utilize deep learning models that found success in image recognition and classification tasks for motor performance prediction. A number of deep convolutional neural networks (CNNs) have been adopted for motor applications previously, while more recently Vision Transformer (ViT) models are gaining interests. In this paper, we evaluate multiple deep CNN and ViT models on two datasets of interior permanent magnet motors, and show that ViT based models can achieve superior accuracy for cogging torque prediction compared with CNN based models, and can be jointly trained on the two different datasets and still provides prediction with low error.

***Index Terms*—Deep Learning, Surrogate Model, Electrical Motors, Vision Transformer**

I. Introduction

The accurate analysis of a motor design typically relies on numerical simulations based on finite-element analysis (FEA), which are time-consuming, especially when various operating points are evaluated for one design. For motor design optimization tasks, parameter sweeping or iterative optimization methods are often utilized to evaluate a large number of design candidates using FEA in order to identify the optimal design. It is therefore desirable to develop surrogate models to rapidly predict motor performances. In recent years, machine learning and deep learning models have found success in many applications, and have been proposed for motor surrogate model due to their powerful capability to emulate highly nonlinear functions [1], [2]. In particular, one popular approach is to represent a motor magnetic design as a 2D image, and fed it into a convolutional neural network (CNN) based model to predict the motor performance [3]. However, it remains a challenge to accurately predict highly sensitive metrics such as cogging torque and torque ripple for permanent magnet motors. More recently, Vision Transformer (ViT) [4], [5] architectures have demonstrated very strong capabilities in image recognition. ViT adopts multi-head self-attention mechanism to grasp the correlation between different input parts of an image, and is considered to be more capable than CNN in grasping global features in an image [6].

In this work, we build deep CNN models and ViT models as surrogate for the prediction of cogging torque on two dataset of interior permanent magnet synchronous motors (IPMSMs). By comparing different variations of CNN and ViT models, we show that ViT based models generally can achieve higher performance for the task.

II. Dataset & Experiment Design

In this work, we built datasets on the evaluate the cogging torque of 4-pole 24-slot IPMSMs for two types of magnet configurations. Fig. 1 shows a representative design of one magnetic pole for each configuration. All the motor designs

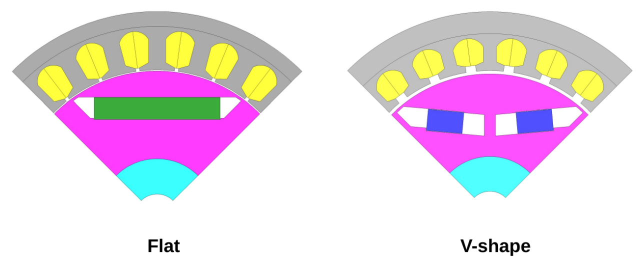

Fig. 1. Example design of IPM motors with (left) flat magnets and (right) V-shape magnets.

in the two datasets have the same inner and outer dimensions. For motors with flat magnets as shown in the left figure, a total of 19,375 designs are generated by adjusting 13 geometrical parameters, and evaluated with FEA simulations. For motors with V-shape magnets as shown in the right figure, a total of 35,001 designs are generated by adjusting 17 geometrical parameters, and evaluated with FEA simulations. The simulated torque waveform for each design is decomposed into a Fourier series including the dominating harmonic terms:

$$\mathrm{T}(\theta) = A_{12}\cos(12\theta) + B_{12}\sin(12\theta) + A_{24}\cos(24\theta) + B_{24}\sin(24\theta) \quad (1)$$

The four Fourier coefficients $A_{12}, B_{12}, A_{24}, B_{24}$ are recorded to recover the torque waveform and cogging torque. Each entry in the compiled dataset includes the value of design parameters, the corresponding 2D cross section image, and four Fourier coefficients.

Surrogate models are trained to predict the Fourier coefficients for a given motor design, which can be used to recover the torque waveform according to Eq.(1). The cogging torque is then determined by the difference between maximum and minimum value of the torque: $T_c = \max(T(\theta)) - \min(T(\theta))$. Based on previous research [7], the performance of cogging torque prediction with the Fourier series approach is generally better than directly predicting cogging torque value using the same model.

TABLE I
HYPERPARAMETERS FOR THE CNN AND VIT MODELS

Hyperparameter	CNN Models	ViT Models
Batch Size	128	128
Learning rate	0.0001	0.09 (with cosine decay)
Training epoch	300	300

Each dataset is split into 70% training set, 10% validation set, and 20% test set. During training of surrogate models, we aim to minimize the root-mean-square error (RMSE) of the cogging torque prediction. Performance of a trained model is evaluated on the test dataset.

For fair comparison, we aim to keep key hyperparameters consistent between CNN and ViT models consistent, and adjust for optimal performance for each model if needed. Table I lists some key hyperparameters of our experimental setup. Notably, for ViT models, the learning rate is initially large and gradually decreases as the training epochs progress. However, for the CNN-based models, a smaller learning rate is preferred for optimal performance.

III. RESULTS & DISCUSSIONS

We train each model on the two datasets separately, and evaluate the performance using RMSE on the test dataset. For CNN models, variations of VGG models and ResNet models are evaluated, including VGG16, VGG19, ResNet50, ResNet101, ResNet152. For ViT models, base and large models with different patch sizes are evaluated, including B/16, B/32, L/16, L/32. Due to the large number of model parameters and relatively small dataset size, all model parameters are pretrained on ImageNet before fine-tuned on our IPMSM motor datasets.

The results are shown in Table II, along with the model size in terms of trainable parameters in Millions. All RMSE values for the cogging torque are in N·m. Noticeably, all ViT based models outperform all the CNN based models on both datasets. For the dataset with flat magnets, the best performance from a CNN model is 0.905, achieved by VGG16 model, while for ViT models, the RMSE is reduced to 0.215 achieved by the L/16 model. Even the worst performing ViT model has a low RMSE of 0.272, which is still significantly more accurate than CNN models. All models perform better on the dataset with V-shape magnets, possibly due to the larger dataset size. Still, all ViT models have smaller prediction error than CNN models, with the best performing L/16 achieving a very low 0.146 N·m RMSE.

In addition to evaluating the models separately for the two datasets, we also would like to evaluate the prediction capability of a surrogate model for a given motor design without specifying the configuration of magnets of the IPM motor. To do this, we fine-tune the model with the training data from both datasets, and evaluate its performance jointly on the test data from both datasets. We expect that the new model would perform worse than the same model for a single dataset due to the more complicated task.

We perform the experiment on the smaller ViT model B/32, and the jointly trained model achieves an RMSE of 0.332 ± 0.246. While this is worse than the same model trained on the flat magnet or the V-shape magnet dataset only, the error is still very small, and it still outperforms CNN models on the flat magnet data, and comparable with CNN models on the V-shape magnet dataset. This result shows the capability of ViT models in extracting features from motor images with different magnet configuration, and the potential of generalizing to other design configurations.

TABLE II
PERFORMANCE COMPARISON FOR COGGING TORQUE PREDICTION: RMSE ON TEST DATASET

Model	Model Size	Dataset: Flat	Dataset: V-shape
ResNet152	58.8M	0.995 ± 0.897	0.299 ± 0.286
ResNet101	42.8M	1.010 ± 0.904	0.299 ± 0.283
ResNet50	23.9M	0.979 ± 0.873	0.307 ± 0.287
VGG19	144M	0.938 ± 0.829	0.280 ± 0.258
VGG16	138M	0.905 ± 0.806	0.285 ± 0.264
ViT-B/16	86M	0.258 ± 0.179	0.187 ± 0.151
ViT-B/32	88M	0.272 ± 0.201	0.176 ± 0.132
ViT-L/16	304M	0.215 ± 0.151	0.146 ± 0.115
ViT-L/32	305M	0.309 ± 0.231	0.189 ± 0.151

IV. CONCLUSIONS

In conclusion, we evaluated deep learning models for the surrogate of electric motor cogging torque prediction with motor design image as input, and compared various CNN and ViT based models on two datasets with different magnet configurations. Results show significant improvement of prediction accuracy and low RMSE with ViT models, which demonstrate their capability of extracting features from the whole image. ViT model trained jointly on the two dataset also shows low prediction error without specifying the magnet configuration, making this approach promising for electric motor surrogate modeling.

REFERENCES

[1] S. Doi, H. Sasaki, and H. Igarashi, "Multi-objective topology optimization of rotating machines using deep learning," *IEEE transactions on magnetics*, vol. 55, no. 6, pp. 1–5, 2019.

[2] A. Khan, V. Ghorbanian, and D. Lowther, "Deep learning for magnetic field estimation," *IEEE Transactions on Magnetics*, vol. 55, no. 6, pp. 1–4, 2019.

[3] T. Aoyagi, Y. Otomo, H. Igarashi, H. Sasaki, Y. Hidaka, and H. Arita, "Prediction of current-dependent motor torque characteristics using deep learning for topology optimization," in *2021 23rd International Conference on the Computation of Electromagnetic Fields (COMPUMAG)*, 2022.

[4] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, (Red Hook, NY, USA), p. 6000–6010, Curran Associates Inc., 2017.

[5] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," 2021.

[6] M. Naseer, K. Ranasinghe, S. Khan, M. Hayat, F. S. Khan, and M.-H. Yang, "Intriguing properties of vision transformers," in *Proceedings of the 35th International Conference on Neural Information Processing Systems*, NIPS '21, (Red Hook, NY, USA), Curran Associates Inc., 2024.

[7] Y. Sakamoto, B. Wang, T. Yamamoto, and Y. Nishimura, "Permanent magnet motor torque waveform prediction using learned gap flux," in *2024 IEEE 21st Biennial Conference on Electromagnetic Field Computation (CEFC)*, pp. 1–2, 2024.