

HARDCORE: H-field and power loss estimation for arbitrary waveforms with residual, dilated convolutional neural networks in ferrite cores

Nikolas Förster*, Wilhelm Kirchgässner*, Till Piepenbrock*, Oliver Schweins*, Oliver Wallscheid*

*Department of Power Electronics and Electrical Drives
Paderborn University, 33095 Paderborn, Germany

Abstract—The MagNet Challenge 2023 calls upon competitors to develop data-driven models for the material-specific, waveform-agnostic estimation of steady-state power losses in toroidal ferrite cores. The following HARDCORE (H-field and power loss estimation for Arbitrary waveforms with Residual, Dilated convolutional neural networks in ferrite COREs) approach shows that a residual convolutional neural network with physics-informed extensions can serve this task efficiently when trained on observational data beforehand. One key solution element is an intermediate model layer which first reconstructs the bh curve and then estimates the power losses based on the curve's area rendering the proposed topology physically interpretable. In addition, emphasis was placed on expert-based feature engineering and information-rich inputs in order to enable a lean model architecture. A model is trained from scratch for each material, while the topology remains the same. A Pareto-style trade-off between model size and estimation accuracy is demonstrated, which yields an optimum at as low as 1755 parameters and down to below 8 % for the 95-th percentile of the relative error for the worst-case material with sufficient samples.

Index Terms—Magnetics, machine learning, residual model

I. INTRODUCTION

The MagNet Challenge 2023 is tackled with a material-agnostic residual convolutional neural network (CNN) topology with physics-informed extensions in order to leverage domain knowledge. Topological design decisions are dictated by peculiarities found in the data sets and by the overall goal of maximum estimation accuracy at minimum model sizes.

The topology's central idea is the calculation of the area within the bh polygon based on a preceding h sequence estimate, see Fig. 3. The area within the polygon formed by the sequences $\mathbf{b}, \mathbf{h} \in \mathbb{R}^{1024}$ can be calculated using the shoelace formula or surveyor's area formula [1]. The shoelace method assigns a trapezoid to each edge of the polygon as depicted in Fig. 1. The area of these trapezoids is defined according to shoelace either with a positive or negative sign, according to the hysteresis direction. The negative areas compensate for the parts of positive trapezoids that extend beyond the boundaries of the polygon. Provided that the polygon is shifted into the first quadrant by some offsets h_{os} and b_{os} , the power loss in W m^{-3} caused by magnetic hysteresis effects can be computed with the frequency f , $M = 1024$, and circular padding by

$$\hat{p}_{\text{hyst}} = f \cdot \frac{1}{2} \sum_{i=0}^{M-1} b_i (h_{i-1} - h_{i+1}). \quad (1)$$

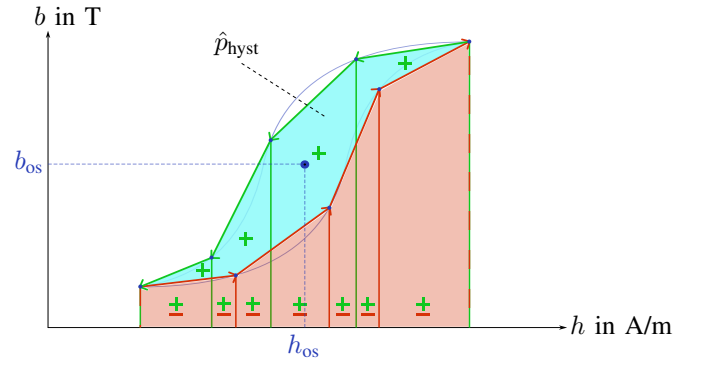


Fig. 1. Visualization of the shoelace formula applied to a bh polygon.

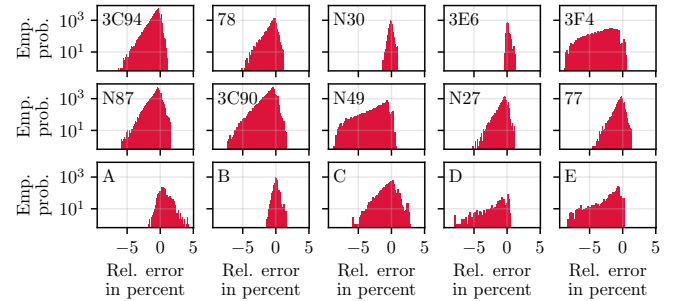


Fig. 2. Relative error $(\hat{p}_{\text{hyst}} - p)/p$ histogram between provided scalar p and \hat{p}_{hyst} calculated from the likewise provided bh polygon area.

When applying (1) on the given sequences $\mathbf{b}, \mathbf{h} \in \mathbb{R}^M$ with $M = 1024$, it becomes evident that the calculated area does not equal the provided loss measurements exactly. Fig. 2 shows the discrepancy with respect to the provided scalar loss p for all materials. The relative error ranges up to over 7 % for certain materials (e.g. 3F4, N49, D, E). Consequently, if merely an h -predicting model was to be identified, the lower bound on the rel. error would be significantly elevated by this circumstance alone.

Since the power losses calculated from neither the ground truth bh curve area (assuming ideal knowledge on the h sequence) nor the estimated area (\hat{h} reconstructed via a CNN) do perfectly match the provided loss measurement values (targets), an additional residual correction mechanism is added

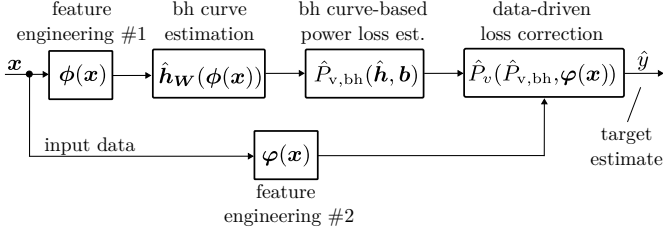


Fig. 3. Overview of the physics-inspired HARDCORE modeling toolchain.

to compensate for this. A high-level view on the proposed residual, physics-inspired modeling toolchain is depicted in Fig. 3, which is coined the HARDCORE approach (H-field and power loss estimation for Arbitrary waveforms with Residual, Dilated convolutional neural networks in ferrite COREs), and its details are discussed in the following.

II. MODEL DESCRIPTION

A residual CNN with physics-informed extensions is utilized for all materials. Such a CNN is trained for each material from scratch. Yet, the topology is unaltered across materials, signal waveforms, or other input data particularities.

A. One-dimensional CNNs for h -estimation

A 1D CNN is the fundamental building block in this contribution, which consists of multiple trainable kernels or filters per layer slid over the multi-dimensional input sequence in order to produce an activation on the following layer [2]. These activations denote the convolution (more precisely, the cross-correlation) between the learnable kernels and the previous layer's activation (or input sequence). In this stateless architecture, circular padding ensures that subsequent activation maps are of equal size. Circular padding can be utilized here instead of the common zero-padding as sequences denote complete periods of the b and h curve during steady state. Moreover, a kernel does not need to read strictly adjacent samples in a sequence at each point in time, but might use a dilated view, where samples with several samples in between are used. The dilated, temporal CNN update equation for the i -th filter's activation $a_i^{(l)}[k]$ at time k and layer l with the learnable coefficients $\mathbf{W}_i \in \mathbb{R}^{A \times \kappa}$ applied on A previous layer's filters, an uneven kernel size of $\kappa \in \{2x+1 : x \in \mathbb{N}_0\}$, and the dilation factor δ reads

$$a_i^{(l)}[k] = \sum_{p=0}^{A-1} \sum_{j=-(\kappa-1)/2}^{(\kappa-1)/2} \mathbf{W}_{i;(p,j)} \cdot a_p^{(l-1)}[k+j\delta]. \quad (2)$$

Since the task at hand does not require causality of CNN estimates along the time domain (losses are to be estimated from single b sequences), the sliding operation can be efficiently parallelized, and sequential processing happens merely along the CNN's depth. All 1D CNN layers are accompanied by weight normalization [3]. A conceptual representation of the 1D CNN for estimating \hat{h} is visualized in Fig. 6 (left part).

B. Feature engineering

The term feature engineering encompasses all preprocessing, normalization, and derivation of additional features in an observational data set. The input data contains the frequency f , the temperature T , the measured losses p as well as the 1024 sample points for the b and h waveforms. Especially the creation of new features that correlate as much as possible with the target variable (here, the h curve or the scalar power loss p) is an important part of most machine learning (ML) frameworks [4].

1) *Normalization*: As is typical in neural network training, all input and target features have to be normalized beforehand. All scalar and time series features are divided by their maximum absolute value that occurs in the material-specific data set, with the exception of the temperature and the frequency, which will be divided by 75°C and 150 kHz , respectively, regardless the material. Moreover, for an accurate h estimate, it was found to be of paramount importance to normalize each b and h curve again on a per-profile base in dependence not only on the ℓ_∞ norm of $|b|$, but also on the maximum absolute b and h appearing in the entire material-specific data set. The latter two values are denoted b_{lim} and h_{lim} , and can be understood as material-specific scaling constants. In particular, the per-profile normalized b and h curves for a certain sample read

$$b_n = \frac{b}{\max_k |b[k]|}, \quad h_n = \frac{h}{h_{\text{lim}}} \cdot \frac{b_{\text{lim}}}{\max_k |b[k]|}, \quad (3)$$

with $h_{\text{lim}} = \max_{i,k} |h_i[k]|$, $b_{\text{lim}} = \max_{i,k} |b_i[k]|$, and i being the sample index in the entire material-specific data set. Then, b_n is added to the set of input time series features, and h_n is the target variable for the h estimation task.

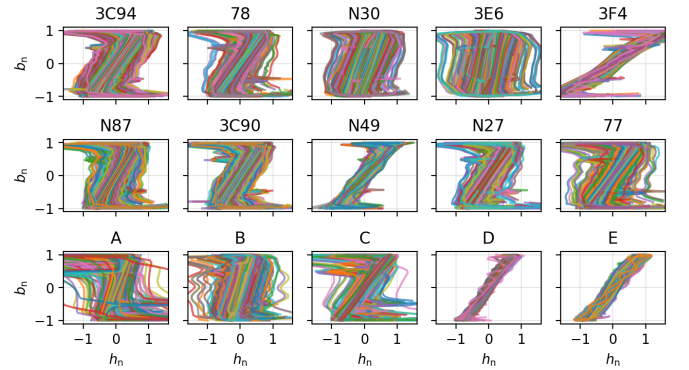


Fig. 4. Exemplary samples of the normalized b and h curves.

The b_n over h_n curves are displayed in Fig. 4, which underlines how the polygon area becomes roughly unified (no large area difference between samples). In the following, all features that get in touch with the model are normalized values without any further notational indication.

2) *Time series features (feature engineering #1)*: As discussed in Sec. II-A, 1D CNNs build the core of the implemented model. The inputs to the CNNs are the (per-profile) normalized magnetic flux density b_n and the corresponding first and second order derivatives (\dot{b}_n and \ddot{b}_n) as time series. In a macroscopic measurement circuit context, \dot{b} corresponds to

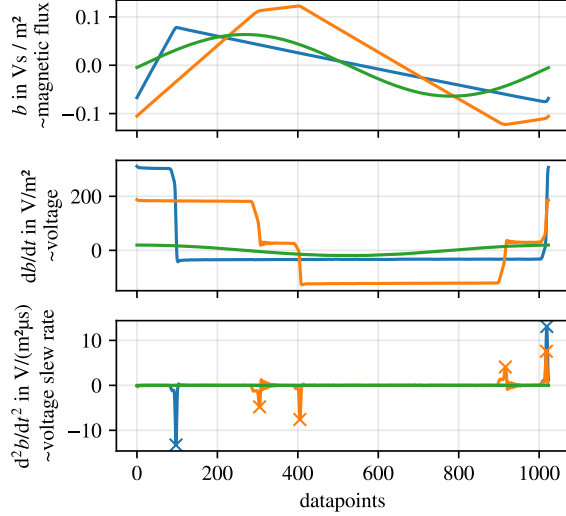


Fig. 5. Magnetic flux density examples and their first and second order derivatives for a sinusoidal, triangular and one unclassified waveform with a circuit-based interpretation in terms of their proportionality to magnetic flux, voltage and the voltage slew rate.

the applied magnetizing voltage throughout the measurement process of the data. Accordingly, \dot{b} represents the voltage slew rate during the commutation of the switches in the test setup. Consequently, the second derivative allows to detect switching events and to characterize them according to their maximum slew rate. Fig. 5 shows, that the sinusoidal waveform (green) is generated without any fast transient switching behaviour, probably with a linear signal source. The nonsinusoidal examples show typical switching behaviour with different voltage slew rates during the single transitions and voltage overshoots as well as ringing. The second derivative of b informs the ML model about switching transition events and how fast changes in time are.

3) *Scalar features (feature engineering #2)*: Although sequence-based CNNs take up the main share of the ML model size, scalar environmental variables also have a considerable impact on h and p . While the temperature T is passed to the model unaltered (but normalized), the frequency is presented by its logarithm $\ln(f)$. The sample time $1/f$ is passed directly to the model. Furthermore, some b -derived scalar features are also passed to the model to feed in a priori knowledge. For example, the peak-to-peak magnetic flux Δb as well as the mean absolute time derivative $|\dot{b}|$ are directly fed into the network. Each waveform is automatically classified into "sine", "triangular", "trapezoidal", and "other" by consulting the form and crest factors, as well as some Fourier coefficients. The waveform classification is presented to the model by one hot encoding (OHE). A summary of all expert-driven input features is presented in Tab. I.

C. Residual correction and overall topology

The model topology comprises multiple branches that end in the scalar power loss estimate \hat{p} . An overview is sketched in Fig. 6. Two main branches can be identified: an h -predictor

TABLE I
UTILIZED INPUT FEATURES.

Time series features		Scalar features	
mag. flux density	b	temperature	T
per-profile norm.	b_n	sample time	$1/f$
1st derivative	\dot{b}_n	log-frequency	$\ln(f)$
2nd derivative	\ddot{b}_n	peak2peak	Δb
tan-tan-b	$\tan(0.9 \cdot \tan(b_n))$	log peak2peak	$\ln(\Delta b)$
		mean abs dbdt	$ \dot{b} $
		log mean abs dbdt	$\ln(\dot{b})$
		waveform (OHE)	

and a wrapping p -predictor. The h -predictor utilizes both time series and scalar features with CNNs and multilayer perceptrons (MLPs), and estimates the full h sequence. The p -predictor predicts p , on the other hand, and leverages the predicted magnetic field strength \hat{h} with the shoelace formula and a scaling factor. The latter accounts for the losses inexplicable by the bh -curve (recall Fig. 2), and is predicted by a MLP that utilizes the scalar feature set only.

The h -predictor merges time series and scalar feature information by the broadcasted addition of its MLP output to a part of the first CNN layer output. This effectively considers the MLP-transformed scalar features as bias term to the time-series-based CNN structure.

On the merged feature set, two further 1D CNN layers follow that end the transformation in a 1024-element sequence. The per-profile scaled b_n sequence from the set of input time series is element-wise added to this newly obtained estimation (residual connection). This results in the CNN model to merely learn the difference between h_n and b_n [5]. Eventually, this sequence becomes the h estimation \hat{h}_n when the sequence's average along the time domain, that is, across all 1024 elements, is subtracted from each element. This is a physics-informed intervention in order to ensure a bias-free h estimate \hat{h} after denormalization. Note that all such operations are still end-to-end differentiable with an automatic differentiation framework such as PyTorch [6], [7].

Since the resulting \hat{h} can only be trained to be as close as possible to the provided h sequence, which is not leading to the correct p ground truth (cf. Fig. 2), another MLP is branched off the scalar input feature set, and denotes the start of the p -predictor. This MLP inherits two hidden layers and concludes with a single output neuron. This neuron's activation, however, is not \hat{p} but rather an area scaling factor $s \in [-1, 1]$ to be embedded in the shoelace formula (1) with

$$\hat{p} = f \cdot (0.5 + (0.1 \cdot s)) \sum_{i=0}^{M-1} b_i (\hat{h}_{i-1} - \hat{h}_{i+1}). \quad (4)$$

Consequently, the p -predictor branch can alter the shoelace formula result by up to 10 % in positive and negative direction.

The p -predictor can be justified physically, when referring to Fig. 2 again. As the comparison shows, the hysteresis loss represents the total loss within a variation of -10% to $+5\%$. The positive deviation ($\hat{p}_{\text{hyst}} > p$) indicates some measurement discrepancy between the measured loss and the given b and h curves. For parts of the negative deviations, a physical explanation can be found in eddy current losses, related to

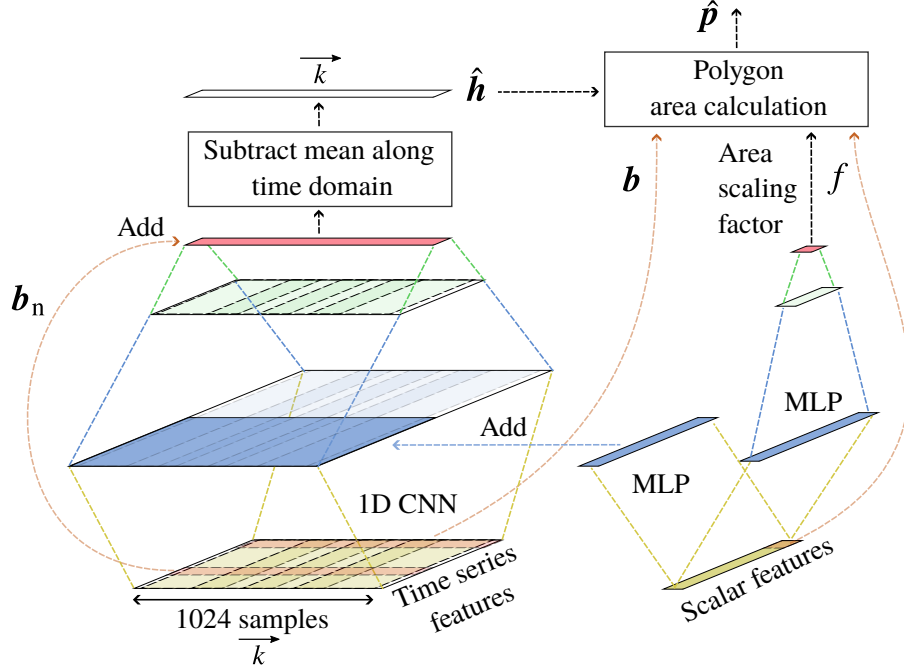


Fig. 6. The residual 1D CNN topology is shown while applied on time series and scalar features, which also contain engineered features from Tab. I.

a high dielectric constant and non-zero conductivity. Due to the small thickness of the used toroidal cores and the limited excitation frequency, eddy current losses are assumed to be of minor effect.

The physical interpretability of the intermediate estimate \hat{h} is a key advantage of the HARDCORE approach: First, it enables utilizing full h time series simulation frameworks (e.g., time domain FEM solvers). Secondly, for future designs of magnetic components with arbitrary shapes it becomes indispensable to accurately take into account also geometrical parameters of the core. This is only possible by distinguishing between the magnetic hysteresis and the (di-)electric losses.

D. Training cost functions

The training process involves two cost functions for a training data set with size N : First, the h estimation accuracy, which is assessed with the mean squared error (MSE) as

$$\mathcal{L}_{\text{MSE,H}} = \frac{1}{NM} \sum_{n=0}^{N-1} \sum_{i=0}^{M-1} (\hat{h}_{i,n} - h_{i,n})^2. \quad (5)$$

Second, the power loss estimation accuracy is to be gauged. Despite the relative error being the competition's evaluation metric, the **mean squared logarithmic error (MSLE) is selected**

$$\mathcal{L}_{\text{MSLE,P}} = \frac{1}{N} \sum_{n=0}^{N-1} (\ln \hat{p}_n - \ln p_n)^2 \quad (6)$$

in order to not overemphasize samples with a relatively low power loss [8]. As $\mathcal{L}_{\text{MSLE,P}}$ also depends on \hat{h} through (4), the question arises, how both cost functions are to be weighted. In this contribution, a scheduled weighting is applied with

$$\mathcal{L}_{\text{total}} = \alpha \mathcal{L}_{\text{MSLE,P}} + (1 - \alpha) \mathcal{L}_{\text{MSE,H}}, \quad (7)$$

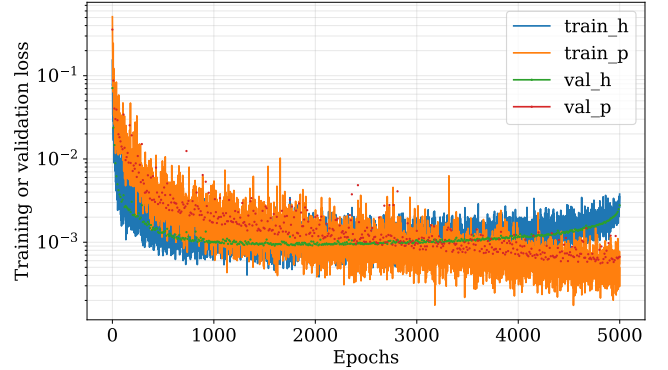


Fig. 7. Exemplary training and validation loss curve for material A, seed 0 and fold 3.

where $\alpha = (\beta \cdot i_{\text{epoch}}) / K_{\text{epoch}}$ with $\beta \in [0, 1]$ denoting a hyperparameter scaling factor, K_{epoch} being the number of training epochs, and $i_{\text{epoch}} \in \{0, 1, \dots, K_{\text{epoch}} - 1\}$ representing the current epoch index. The scheduled weighting ensures that the model focuses on \hat{h} in the beginning of the training, where more information is available. Later though, the model shall draw most of its attention to the power loss estimate, possibly at the expense of the h estimation accuracy. A training example for material A is depicted in Fig. 7.

III. HYPERPARAMETERS, PARETO FRONT AND RESULTS

The proposed topology features several degrees of freedom in form of hyperparameters. An important aspect is the model size, which is defined by the number of hidden layers and neurons in each layer. A simple trial-and-error investigation

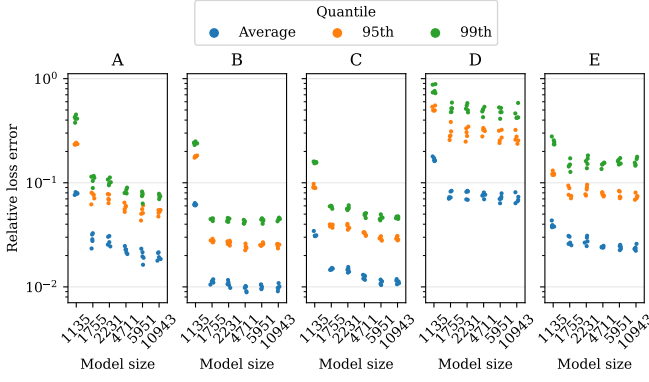


Fig. 8. Pareto front for the evaluation materials (A, B, C, D and E) showing model size (amount of parameters) vs. relative error of the power loss estimation.

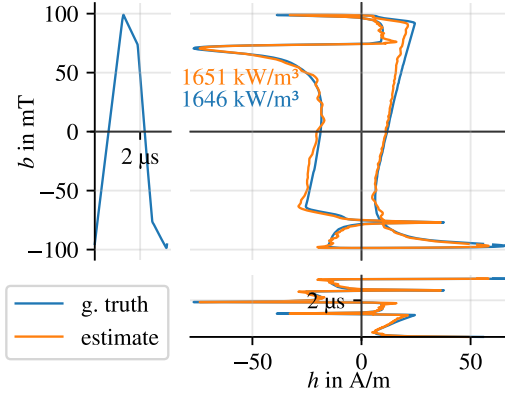


Fig. 9. Exemplary ground truth vs. estimated bh curve comparison.

can provide fast insights into the performance degradation that comes with fewer model parameters. In Fig. 8, several particularly selected model topologies are illustrated against their achieved relative error versus the inherited model size. The scatter in each quantile is due to different random number generator seeds and folds during a stratified 4-fold cross-validation. Topology variations are denoted by the amount of neurons in certain hidden layers. In addition, the largest topology has an increased kernel size with $\kappa = 17$, and the smallest topology has the second CNN hidden layer removed entirely (the green layer in Fig. 6).

A slight degradation gradient is evident as of 5 k parameters for materials A and C, whereas for the other materials the trend is visible only when removing the second hidden layer. Overall, the material performance scales strictly with the amount of training data available. Since fewer model parameters are a critical aspect, the chosen final model has 1755 parameters, which is at an optimal trade-off point on the Pareto front.

In Tab. II, the model size of a corresponding PyTorch model file dumped to disk as just-in-time (jit) compilation is reported. The exemplary bh -curve and h -curve estimation is shown in Fig. 9. Reported error rates come from the best seed out of five during a four-fold cross-validation ($\beta = 1, K_{\text{epochs}} = 5000$, Nesterov Adam optimizer). It shows effectively that any ma-

TABLE II
FINAL MODEL DELIVERY OVERVIEW

Material	Parameters	Training data	Model size	Relative error	
				Average	95-th quantile
A	1755	2432	43.13 kB	2.34 %	6.20 %
B	1755	7400	43.13 kB	1.10 %	2.68 %
C	1755	5357	43.13 kB	1.46 %	3.70 %
D	1755	580	43.13 kB	7.03 %	25.76 %
E	1755	2013	43.13 kB	2.51 %	7.10 %

terial can be modeled with the same topology at high accuracy as long as a critical training data set size is available (which is not the case for material D, see available training data in Tab. II). The final model is already a trade-off between model size and accuracy, such that in case one of the two criteria can be softened, the other can be further improved.

The final model delivery is trained on all training data samples (no repetitions with different seeds), and with $K_{\text{epoch}} = 10000, \delta = 4, \kappa = 9$. This final topology features a CNN with $12(\text{TanH}) \rightarrow 8(\text{TanH}) \rightarrow 1$ (linear) kernels, a MLP with $11(\text{TanH})$ neurons, and a p -predictor MLP with $8(\text{TanH}) \rightarrow 1(\text{TanH})$ neurons.

IV. CONCLUSION

A material-agnostic CNN topology for efficient steady-state power loss estimation in ferrite cores is presented. Since the topology remains unaltered across materials and waveforms at a steadily high accuracy, the proposed model can be considered universally applicable to plenty of materials. As long as sufficient samples of a material are available (roughly, 2000), the relative error on the 95-th quantile remains below 8%. Thus, the contributed method is proposed to become a standard way of training data-driven models for power magnetics.

REFERENCES

- [1] B. Braden, "The surveyor's area formula," *The College Mathematics Journal*, vol. 17, no. 4, pp. 326–337, 1986.
- [2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information and Processing Systems (NIPS)*, vol. 60, no. 6, 2012, pp. 84–90. [Online]. Available: <https://doi.org/10.1145/3065386>
- [3] T. Salimans and D. P. Kingma, "Weight Normalization: A Simple Reparameterization to Accelerate Training of Deep Neural Networks," ArXiv e-prints: 1602.07868 [cs.LG], 2016. [Online]. Available: <http://arxiv.org/abs/1602.07868>
- [4] P. Domingos, "A few useful things to know about machine learning," *Communications of the ACM*, vol. 55, no. 10, p. 78, 2012. [Online]. Available: <http://dl.acm.org/citation.cfm?doid=2347736.2347755>
- [5] S. Bai, J. Z. Kolter, and V. Koltun, "An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling," 2018. [Online]. Available: <http://arxiv.org/abs/1803.01271>
- [6] A. Paszke, S. Gross *et al.*, "PyTorch: An Imperative Style, High-Performance Deep Learning Library," in *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc., 2019, pp. 8024–8035.
- [7] A. G. Baydin, B. A. Pearlmutter *et al.*, "Automatic differentiation in machine learning: A survey," *Journal of Machine Learning Research*, vol. 18, pp. 1–43, 2018. [Online]. Available: <https://arxiv.org/abs/1502.05767>
- [8] C. Tofallis, "A better measure of relative prediction accuracy for model selection and model estimation," *Journal of the Operational Research Society*, vol. 66, pp. 1352–1362, 2015.