CNN and transfer learning based online SOH estimation for lithium-ion battery

Yang Li[1,2], Jili Tao[1]

1. Ningbo Institute of technology, Zhejiang University, Ningbo, 315100, China
E-mail: tjl810@126.com

2. College of Control Science and Engineering, Zhejiang University, Hangzhou, 310000, China
E-mail: 1185071644@qq.com

Abstract: Accurate estimation of state of health (SOH) is extremely important for lithium-ion (Li-ion) rechargeable batteries. An improved strategy based on convolutional neural network (CNN) architecture is proposed for online SOH estimation, in which the features can be automatically extracted, instead of manual extraction. Accelerated aging data from different dynamic conditions, including overcharging and over-discharging, is utilized to pretrain a base model. And then by transfer learning method, the base model can be fine-tuned with just 15% normal-speed aging data and migrated as a new model for testing on the remaining 85% normal-speed aging data. The transfer learning method can reduce the laboratory costs for large amount of cycling data. Only the constant current charging data is selected as the input of model. And results show that the proposed deep learning method owns great generalization ability between different aging scenarios.

Key Words: State of health, Lithium-ion batteries, Convolutional neural network, Transfer learning

1 INTRODUCTION

Accurate SOH estimation of lithium-ion (Li-ion) rechargeable batteries plays an important role in battery management system (BMS), which ensures the safety and reliability of cell operating. Recent studies can be generally divided into two kinds of approaches: model-based methods [1, 2] and data-driven methods [3–12]. Model-based methods require professional knowledge of complex internal physical mechanisms of the battery, so more and more studies pay attention to data-driven methods. Incremental capacity analysis (ICA) method [6–8] has been popularly adopted in SOH estimation, using the differentiation of battery capacity over its terminal voltage, dQ/dV. However, tradition machine learning approaches manually extract characteristic features, which requires heavy human labor. And it's quite hard to determine what are the most suitable and useful features for capacity estimation. A deep learning model can automatically learn from large scale charging and discharging data, eliminating the complicated manual feature extraction. Shen et al. [9] presented a deep learning method with convolutional neural network,which was the first outstanding attempt to apply deep learning method to battery SOH estimation. The slightly regretful thing is that Shen's study was based on 10-year daily cycling data. Cycling experiments are time-consuming and the goal of our study is to improve SOH model rather than depend too much on data.

Due to those analysis, this study applies a convolutional neural network (CNN) model in SOH estimation. In this paper, transfer learning based on accelerated aging data and partial normal-speed aging data is implemented, which saves lots of experimental time. Different dynamic accelerated aging protocols, including overcharging and over-discharging, are utilized to verify the generalization performance of the proposed deep learning model. Two types of batteries, i.e., SONY US18650VTC6 and FST-2000 are performed under accelerated and normal-speed aging conditions.

The specific contributions of this paper are as follows:

(1) The proposed deep learning model eliminates the need to manually extract features from large amounts of data, which reduces human labor and prevents the loss of important information.

(2) Accelerated aging data is utilized to pretrain a source model and a group of parameters are learned. Then partial normal-speed aging cycles(about 15%) are adopted to train further and fine-tune the established base model. And finally the remaining normal-speed aging cycles(about 85%) are used to test the migrated model. The transfer learning method can reduce the laboratory costs for large amount of cycling data, because accelerated aging experiments under stress factors greatly reduce battery lifespan.

(3) Results show that the proposed deep learning method owns great generalization ability between different aging scenarios. The base model trained on accelerated scenarios can be quickly migrated to the target model for normal-speed estimation.

(4) Only partial charging data(constant current), instead of the complete charging process(containing constant current and constant voltage), is selected as the input of model. It eliminates the need for the device to measure constant

978-1-7281-5855-6/20/$31.00 ©2020 IEEE

voltage charging data, and the model can load samples as input faster.

The remainder of this paper is divided into 3 sections. Section 2 introduces the architecture of the proposed CNN model and transfer learning method. And some implementation details are presented in this section. Section 3 exhibits the experimental datasets and explains the results. Final conclusions and future prospects of this research are presented in Section 4.

2 CNN and transfer learning method

2.1 The overall architecture of CNN

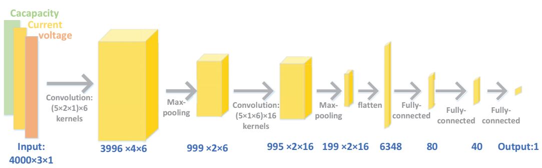


Figure 1: The overall structure of CNN

The overall structure of CNN is described in Figure 1. It mainly contains two kinds of layers with weights: convolutional layers and fully-connected layers [13, 14]. Two convolutional layers are constructed before three fully-connected layers. Rectified linear units (ReLU) follow every convolutional layer and fully-connected layer. The output of the last fully-connected layer is transmitted to a single neuron, which represents the degree of battery aging.

The input of the model is a matrix with fixed size $h \times w \times c$, where h, w, c denote the height, weight and channel numbers, respectively. Here $h = 4000$, $w = 3$ and $c = 1$.For each cycle, 4000 data points are discretely sampled at regular time intervals from the constant current charging curve. Since the input channel number is 1, the input is equivalent to a matrix, as shown in Eq.(1), each column of which represents one parameter, i.e., voltage, current and capacity, respectively.

$$\mathrm{X} = \begin{bmatrix} \mathrm{V}_1 & I_1 & C_1 \\ \mathrm{V}_2 & I_2 & C_2 \\ \cdots & \cdots & \cdots \\ \mathrm{V}_{4000} & I_{4000} & C_{4000} \end{bmatrix}_{4000\times 3} \tag{1}$$

The first convolutional layer filters the $4000 \times 3 \times 1$ input with 6 kernels of size $5 \times 2 \times 1$ and stride of size 1×1. The second convolutional layer takes as input the max-pooled output of the first convolutional layer and filters it with 16 kernels of size $5 \times 1 \times 6$. The first fully-connected layer contains 80 neurons and is fully connected to the last max-pooling layer. The second fully-connected layer has 40 neurons and the output has one single neuron.

2.2 Model building

2.2.1 Training algorithm

For fully connected layers, the forward propagation is described in Eq.(2), where f denotes the activation function (ReLU is adopted in this study, as shown in Eq.(3)).For convolutional layers, the matrix multiplication calculation in Eq.(2) is replaced by a kind of linear operation named convolution.

$$a^l = f(z^l) = f(w^l a^{l-1} + b^l) \tag{2}$$

$$\mathrm{ReLU}(x) = \begin{cases} x & x > 0 \\ 0 & x \leq 0 \end{cases} \tag{3}$$

In order to quantify the difference between the network output and real value, an optimization method, stochastic gradient descent (SGD) with momentum [15], is utilized in this study. The cost function is defined in Eq.(4). The training goal is to update the parameters of the network, making the network output $\widehat{y}_i(x)$ closer and closer to the true value $y_i(x)$, that is, to minimize the $J(w, b)$:

$$\begin{aligned} &J(w,b){=}J_0(w,b) + \lambda\Omega(w) \\ &= \tfrac{1}{2n}\sum_{i=1}^{n}\left(\widehat{y}_i(x) - y_i(x)\right)^2 + \tfrac{\lambda}{2}w^T w \end{aligned} \tag{4}$$

where $\Omega(w)$ denotes the norm penalty term, λ denotes the L2 regularization factor that weighed the relative contribution of the norm penalty term, n denotes the number of samples in each iteration, $\widehat{y}_i(x)$ denotes the network output and $y_i(x)$ denotes the true value.

The parameters θ (weight w and bias b) are iteratively updated as follows:

$$\widehat{g} = \frac{1}{2m}\sum_{i=1}^{m}\left(\widehat{y}_i(x)^j - y_i(x)^j\right)^2 \tag{5}$$

$$\theta_{j+1} = \theta_j - \alpha\widehat{g} + \gamma(\theta_j - \theta_{j-1}) - \lambda\alpha\theta_j \tag{6}$$

where m denotes the number of samples in a mini-batch, $\widehat{y}_i(x)^j$ denotes the output of the i^{th} input matrix of the mini-batch in the j^{th} iteration, $y_i(x)^j$ denotes the corresponding true value, θ_j denotes the parameters in the j^{th} iteration, α denotes the learning rate, and γ denotes the momentum that determined the contribution of gradients from the previous iteration to the current iteration.

2.2.2 Model evaluation index

In the process of model training, we mainly observe the curves of loss value and accuracy to make a judgment on the convergence effect. Generally, the loss on the test set is higher than the loss on the training set. When both no longer drop, we can consider that the model has reached the best. The loss function is defined in Eq.(4). Accuracy is defined as follows:

$$ACC = \left(1 - \frac{1}{n}\sum_{i=1}^{n}\frac{|\widehat{y}_i(x) - y_i(x)|}{y_i(x)}\right) \times 100\% \tag{7}$$

where n denotes the number of samples, $\widehat{y}_i(x)$ denotes the network output and $y_i(x)$ denotes the true value.

Root mean squared error (RMSE) is used to evaluate the trained deep learning model:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{y}_i(x) - y_i(x)\right)^2} \times 100\% \tag{8}$$

2.3 Learning rate decay strategy

Appropriate learning rate adjustment strategy is crucial for the rapid convergence of the model. Piecewise constant decay strategy is adopted here to adjust the learning rate parameter, setting different learning rate constants in the predefined training range. At the beginning, the learning rate is large, and then it became smaller and smaller. The initial value is set to 1×10^{-5}, and decreases by 0.7 times for every 5k epochs. The batch size is set to 64, which means that there are 64 samples in a mini-batch.

2.4 Transfer learning

This study uses a common method in deep learning area, called transfer learning. Transfer learning, as the name implies, is to migrate the learned model parameters to help the new model training. Transfer learning applies to two highly relevant tasks, using the model developed for one source task and then developing a new model for another target task. In transfer learning, we first train a base model with base datasets, then we fine-tune the learned parameters, and finally we migrate them to a target model with target datasets.

The overall strategy is illustrated in Figure 2. The SOH estimation task with accelerated aging data is selected as the source task, and the normal-speed is selected as the target task. Although these two scenarios are inconsistent in some parameters such as the cutoff voltage, there is some similarity included in the mapping relationship between the corresponding input and output.

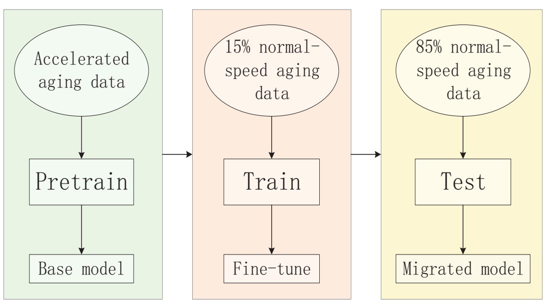


Figure 2: The overall strategy of transfer learning

Accelerated aging data is utilized to pretrain a base model. Because the capacity cannot be directly obtained so it is required to be calculated using the coulomb counting method, which integrates the charging current with respect to time during the partial charging cycle. Then the voltage, current, and charging capacity of constant current charging stage are denoted as the inputs of the proposed deep learning network. The corresponding output is the discharging capacity, which integrates the discharging current over time for the entire discharging cycle via coulomb counting method. Then, those learned model parameters are utilized to continue training with 15% normal-speed aging data. The model is fine-tuned through iterations and then a set of internal parameters are obtained. After the processes of pretraining and training, the model is used to online estimate the unknown discharging capacity of test data (the remaining 85% normal-speed aging cycles), whose voltage and current measurements during charging are available.

Since complete aging experiments of batteries require a long period of time, the experimental data is very limited, and training a model from scratch has to pay a heavy price. Accelerated aging experiments can greatly shorten the battery lifespan, and we can quickly get a batch of data, which is conducive to reducing the cost of experiments. In this way, we only need to use accelerated aging data to pretrain a base model, and then adopt partial normal-speed aging data to train further and fine-tune the established base model.

3 Results and discussions

3.1 Experimental settings

The datasets were collected from three SONY US18650VTC6 batteries (rated capacity:3 Ah) and two FST-2000 batteries (rated capacity:2 Ah). The experiments were carried out under relatively stable room temperature at about 25 $degC$. There are five different scenarios with different aging mode, i.e., accelerated overcharging aging, accelerated over-discharging aging and normal-speed aging. The corresponding battery cells and experimental details are summarized in Table 1. Constant current-constant voltage (CC-CV) charging strategy, the most widely used charging method, was selected to charge the batteries and the constant current (CC) discharging strategy was adopted to consume the capacity. As shown in Figure 3, the batteries were firstly charged with 1C current rate (i.e.,3A and 2A) to the maximum voltage $Vmax$. Then they were charged with constant $Vmax$ voltage, at which stage the current gradually dropped to the cut-off current of 0.05C (i.e.,150mA and 100mA). Rested for 10 min, the batteries were finally discharged with 1C current rate until the voltage reached the minimum value $Vmin$. As shown in Table 1, in Scenario 1, the cutoff conditions were 2.75V and 4.4V, for upper voltage limit $Vmax$ and lower voltage limit $Vmin$, respectively, which means the battery was overcharged. In Scenario 2, $Vmin$ and $Vmax$ were 1.95V and 4.2V, respectively; in other words, the battery was over-discharged. In the same way, the battery in Scenario 4 was over-discharged to lower voltage limit of 2V. For Scenario 3 and 5, the voltage boundaries of normal-speed aging condition were set to 2.75-4.2V.

The SOH value of the current state is calculated by Eq.(9), where C_0 denotes the initial capacity, and $C_{current}$ denotes the discharging capacity of the battery under the current state. The obtained SOH values are used as labels for training data and what the network needs to do is to make the outputs as close to the real values as possible. When the SOH value drops to 80% and we believe that the battery can no longer maintain normal performance. The lifespan of the five scenarios are about 260, 265, 500, 300, 500 cycles, respectively. In normal-speed aging scenarios (Scenario 3 and 5), the lifespan of batteries can reach 500 cycles, and it drops greatly in the accelerated aging scenarios (Scenario 1, 2 and 4). The aging comparison is depicted in Figure 4

Table 1: Details of the five scenarios

Scenario	Battery	Rated capacity	Mode	CCCV charging			CC discharging		lifespan
				Charging current	Vmax	cut-off current	Discharging current	Vmin	
1	SONYUS18650VTC6	3Ah	Over-charging	1C	4.4V	0.05C	1C	2.75V	260cycles
2	SONYUS18650VTC6	3Ah	Over-discharging	1C	4.2V	0.05C	1C	1.95V	265cycles
3	SONYUS18650VTC6	3Ah	Normal-speed	1C	4.2V	0.05C	1C	2.75V	500cycles
4	FST2000	2Ah	Over-discharging	1C	4.2V	0.05C	1C	2V	300cycles
5	FST2000	2Ah	Normal-speed	1C	4.2V	0.05C	1C	2.75V	500cycles

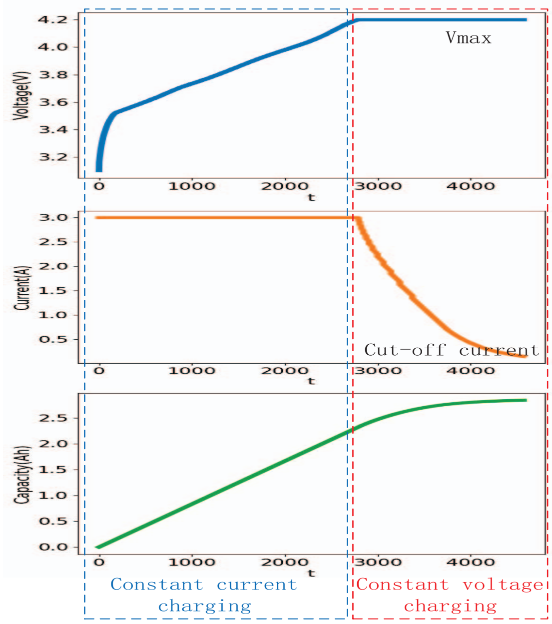


Figure 3: Constant current-constant voltage(CC-CV) charging curve of normal-speed aging

and Figure 5.

$$SOH_{current} = \frac{C_{current}}{C_0} \tag{9}$$

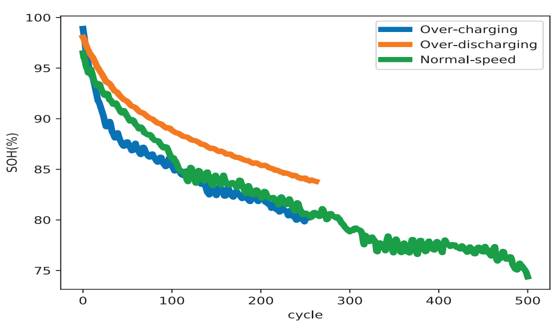


Figure 4: The aging comparison between Scenario 1-3 (SONY US18650VTC6 batteries)

In this study, three experiments based on transfer learning are carried out using the data in Table 1. The details of the three experiments are provided in Table 2. Overcharging aging data is utilized to pretrain a base model in the first and the third experiment while over-discharging aging

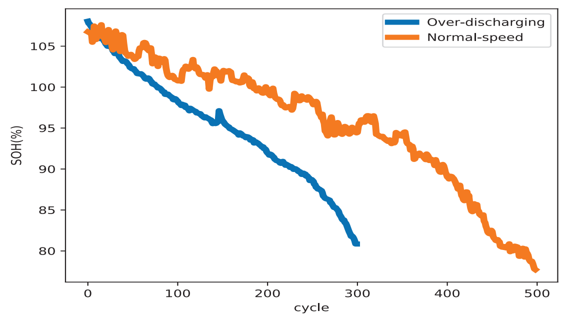


Figure 5: The aging comparison between Scenario 4-5 (FST-2000 batteries)

data is selected in the second experiment. We assume that the vehicles are regularly inspected and maintained, and the batteries are tested at regular intervals. In the maintenance test phase, the batteries are operated under complete charging and discharging tests, so we can get the discharging capacity by integrating discharging current to calculate the reference SOH values. In the non-maintenance test phase, the vehicles are in normal use on the road, so it is difficult to record the complete dynamic discharging data of voltage and current. We need to utilize the partial constant current charging data to analysis the aging trend of batteries. For better comparison, the three experiments both fine-tune their base models using 15% normal-speed aging data sampled at regular intervals in the whole lifespan. And then the remaining 85% cycles of the normal-speed aging data is used to test the three migrated model.

3.2 Result analysis

The advantages of using accelerated aging data for pretraining can be seen from the loss function. On the one hand, the model is trained directly from scratch using 15% normal-speed aging data, which means its parameters are randomly initialized. On the other hand, accelerated aging data is used to pretrain a source model, obtain a batch of parameters, and then fine-tune it based on normal-speed aging data. We can get a conclusion that the latter obviously performs better than the former, not only taking almost half the time to make the loss function value reach below 0.1 (about 60k epochs vs 120k epochs), but also obtaining a much lower final value than the former.

For Experiment 1, comparison branches are designed as follows: after pretrained on the group of SONY US18650VTC6 overcharging data, the model is

Table 2: Details of the three experiments

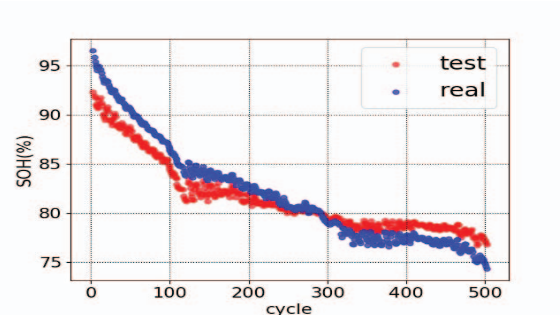
Experiment	Data			Number of epochs	
	Pre-train	Train	Test	Pretrain	Train
1	Over-charging aging data (Scenario 1)	15% normal-speed aging data (Scenario 3)	85% normal-speed aging data (Scenario 3)	2k	100k
2	Over-discharging aging data (Scenario 2)	15% normal-speed aging data (Scenario 3)	85% normal-speed aging data (Scenario 3)	5k	130k
3	Over-discharging aging data (Scenario 4)	15% normal-speed aging data (Scenario 5)	85% normal-speed aging data (Scenario 5)	10k	100k

used directly for validating on the group of SONY US18650VTC6 normal-speed aging data without any epoch of fine-tuning. As we can see in Figure 6(a), there is a certain gap between the estimated value and the real value, but it is not completely cluttered. The result after fine-tuning is shown in Figure 6(b). It can be observed from Figure 6 that the migrated model can effectively estimate SOH, even though it just uses a pretty small part of normal-speed aging cycles for training. Moreover, we do not need the complete charging data. Only the data of constant current charging is used as the input of the model, which eliminates the need for the device to measure constant voltage charging data, and the model can load samples as input faster. Figure 7 and Figure 8 represent the results of Experiment 2 and 3, respectively.

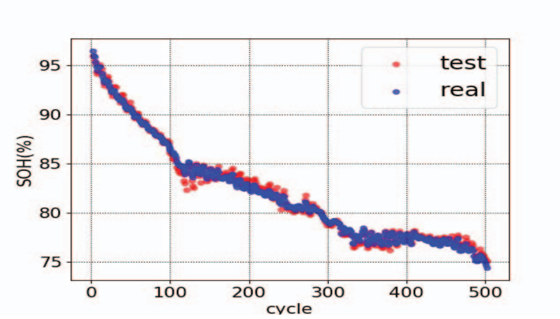
As illustrated in Table 3, the RMSE is 0.319531%, 0.293138% and 0.29796%, and the accuracy is 99.71%, 99.74% and 99.77% for the three experiment, respectively.

Table 3: RMSE of the three experiments

Experiment	RMSE(%)	ACC(%)
1	0.319531	99.71
2	0.293138	99.74
3	0.29796	99.77



(a) Validation after pretraining



(b) Validation after training

Figure 6: Result of Experiment 1:(a) Validation after pretraining (b) Validation after training

4 Conclusions

A convolutional neural network (CNN) architecture is proposed in this paper, applying deep learning method to the classical research of online SOH estimation. Different dynamic capacity consuming scenarios are utilized containing accelerated aging (i.e., overcharging and over-discharging) and normal-speed aging modes. Transfer learning method connects them together to obtain migrated models. The deep learning model automatically extracts features from the constant current charging data. Accelerated aging data and just 15% normal-speed aging data can well train a model for SOH estimation. If we utilize normal-speed aging data to train a model, it will cost much more time because accelerated aging experiments under stress factors greatly reduce battery lifespan. Only the constant current charging data, instead of the complete charging cycle, is selected as the input of model. Even though accelerated and normal-speed aging involve different degrees of internal physical aging and show obviously different SOH downtrend, the proposed migrated model owns great generalization ability, with RMSE less than 0.4%.

Our future goal is to analyze the voltage, current and capacity value collected from the actual operation of the vehicles. When hybrid electric vehicles run on the real road, the batteries are not under standard CC-CV charging or CC discharging conditions. This motivates us to further improve the proposed CNN method.

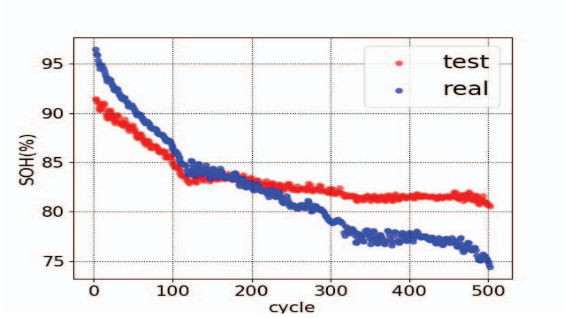
Acknowledgements

This work was partially supported by the National Natural Science Foundation of China(Grant No.61603337) and Zhejiang Provincial Natural Science Foundation (Grant No.LY19F030009).

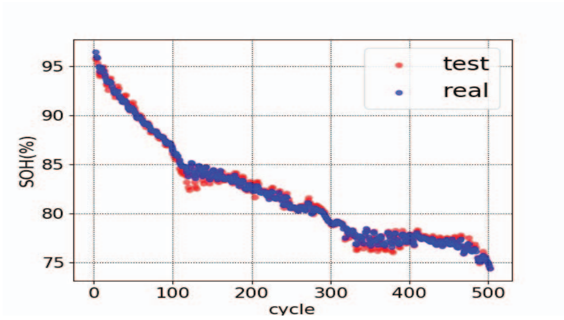
REFERENCES

[1] G. L. Plett, "Extended kalman filtering for battery management systems of lipb-based hev battery packs: Part 2. modeling and identification," *Journal of power sources*, vol. 134, no. 2, pp. 262–276, 2004.

[2] P. Ramadass, B. Haran, P. M. Gomadam, R. White, and B. N. Popov, "Development of first principles capacity fade model for li-ion cells," *Journal of the Electrochemical Society*, vol. 151, no. 2, pp. A196–A203, 2004.

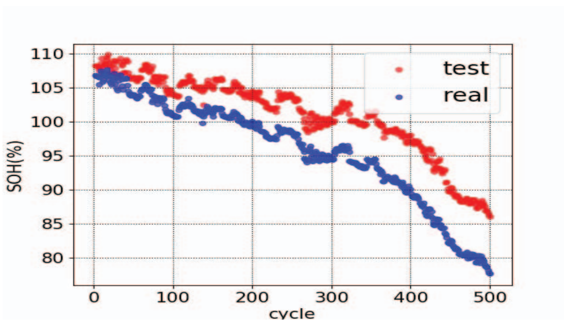


(a) Validation after pretraining

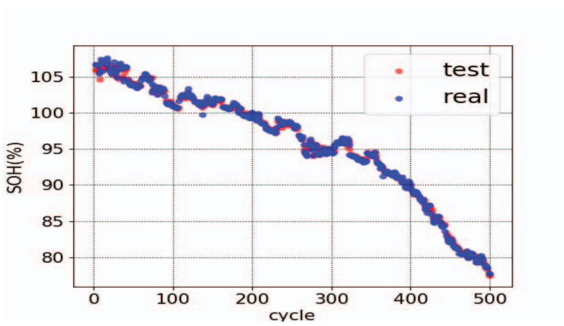


(b) Validation after training

Figure 7: Result of Experiment 2:(a) Validation after pretraining (b) Validation after training



(a) Validation after pretraining



(b) Validation after training

Figure 8: Result of Experiment 3:(a) Validation after pretraining (b) Validation after training

[3] J. Wu, C. Zhang, and Z. Chen, "An online method for lithium-ion battery remaining useful life estimation using importance sampling and neural networks," *Applied energy*, vol. 173, pp. 134–140, 2016.

[4] Z. Chen, M. Sun, X. Shu, R. Xiao, and J. Shen, "Online state of health estimation for lithium-ion batteries based on support vector machine," *Applied Sciences*, vol. 8, no. 6, p. 925, 2018.

[5] Y. Zou, X. Hu, H. Ma, and S. E. Li, "Combined state of charge and state of health estimation over lithium-ion battery cell cycle lifespan for electric vehicles," *Journal of Power Sources*, vol. 273, pp. 793–803, 2015.

[6] Y. Li, M. Abdel-Monem, R. Gopalakrishnan, M. Berecibar, E. Nanini-Maury, N. Omar, P. van den Bossche, and J. Van Mierlo, "A quick on-line state of health estimation method for li-ion battery with incremental capacity curves processed by gaussian filter," *Journal of Power Sources*, vol. 373, pp. 40–53, 2018.

[7] C. Weng, Y. Cui, J. Sun, and H. Peng, "On-board state of health monitoring of lithium-ion batteries using incremental capacity analysis with support vector regression," *Journal of Power Sources*, vol. 235, pp. 36–44, 2013.

[8] X. Tang, C. Zou, K. Yao, G. Chen, B. Liu, Z. He, and F. Gao, "A fast estimation algorithm for lithium-ion battery state of health," *Journal of Power Sources*, vol. 396, pp. 453–458, 2018.

[9] S. Shen, M. Sadoughi, X. Chen, M. Hong, and C. Hu, "A deep learning method for online capacity estimation of lithium-ion batteries," *Journal of Energy Storage*, vol. 25, p. 100817, 2019.

[10] R. Zhang and J. Tao, "A nonlinear fuzzy neural network modeling approach using an improved genetic algorithm," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 7, pp. 5882–5892, 2017.

[11] X. Tang, C. Zou, K. Yao, J. Lu, Y. Xia, and F. Gao, "Aging trajectory prediction for lithium-ion batteries via model migration and bayesian monte carlo method," *Applied Energy*, vol. 254, p. 113591, 2019.

[12] R. Zhang and J. Tao, "Data driven modeling using improved multi objective optimization based neural network for coke furnace system," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 4, pp. 3147–3155, 2016.

[13] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner *et al.*, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[15] B. T. Polyak, "Some methods of speeding up the convergence of iteration methods," *USSR Computational Mathematics and Mathematical Physics*, vol. 4, no. 5, pp. 1–17, 1964.