Convolutional Transformer-Based Multiview Information Perception Framework for Lithium-Ion Battery State-of-Health Estimation

Tianyou Bai and Huan Wang, *Graduate Student Member, IEEE*

***Abstract*— As a momentous indicator for the working status of lithium-ion batteries, state of health (SOH) has been adopted by various new energy enterprises. Especially in the electric vehicle industry, SOH is an essential guarantee for the safety and robustness of operating batteries. However, working in a complex open environment, the estimation model is often confronted with dynamic changes and abrupt disturbs, which demands an outstanding global information perception ability. Therefore, this study proposes a convolution transformer-based multiview information perception framework (MVIP-Trans) for SOH estimation. MVIP-Trans is dedicated to integrating the benefits of both transformer and convolutional neural network (CNN) architectures to learn global and local features with multiview, thus achieving a comprehensive information perception. First, a local information perceptron (LIP) based on parallel multiscale attention (PMS-A) is constructed in order to extract features from local detailed information in voltage and current signals. With the multiscale attention block, the perceptron has an excellent feature screening ability, thereby enhancing valuable information and restraining useless noise. Subsequently, to get an effective global information perception, a transformer-attention architecture is introduced to encode the global dependencies between the filtered local features. Via multiview learning mechanism, MVIP-Trans adequately modeled the internal correlation between SOH and the physical signals of batteries. Experiments by working through real operation datasets such as NASA PCoE and Oxford prove that this model can perfectly accomplish the SOH estimating task. During the experiments, MVIP-Trans obtained the best performance on NASA dataset, as RMSE is 0.005 and R^2-score reaches 0.989. On oxford dataset, MVIP-Trans got an RMSE of 0.003 and an R^2-score of 0.997. These results prove that our model has a better performance than other existing methods.**

***Index Terms*— Convolutional neural network (CNN), deep learning, lithium-ion battery, state of health (SOH), transformer.**

I. Introduction

LITHIUM-ION (Li-ion) battery, as the most common choice for clean energy storage, has been broadly adopted in diversified aspects, which include electric vehicles, electronic devices, and automatic factories [1]. Since the capacity and safety of Li-ion batteries will decrease along with their aging process, it is of conspicuous importance to precisely and steadily monitor the working status and remaining lifespan of the batteries [2], [3]. Therefore, a framework, where the important battery status pointers such as state of health (SOH) can be constantly gauged, is badly needed by the new energy industry [4], [5]. Here, SOH is the ratio of current capacity and original capacity, which shows the shrinking of battery capacity. However, a multitude of difficulties have been found in the estimation of SOH. First of all, the capacity of batteries cannot be directly measured by measuring devices, which means the value of SOH can only be obtained from the prediction based on other measured signals [6]. Moreover, in a complex and changeable working environment, the estimation of SOH is often faced with dynamic changes and abrupt interference, thus limiting the accuracy of the predicted results.

To cope with these challenges, researchers have trialed a large amount of model-based and data-driven methods in the past decades. An off-line method is to measure battery parameters that directly affect the value of SOH, such as internal impedance and electrochemical impedance spectroscopy. This solution has an intuitive working principle and does not involve complex calculations [7]. However, in a practical environment, it is impossible to apply the extensive sensors and instruments required in this method [6]. The model-based methods, such as electrochemical models [8], equivalent circuit models [9], and empirical models [10], employ professional knowledge to build estimation models that describe the aging curves of Li-ion batteries. Referring to chemical professional knowledge, the electrochemical model, such as the pseudo-2-D model (P2D model), depends on information of the battery chemical and electrical activities to model the battery aging phenomenon [11]. However, the adopted equations are complex and require an accurate chemical structure of the battery [12]. The equivalent circuit model can use an active circuit model to explain the functioning procedure of Li-ion batteries [13]. With this model, the original battery circuit may be simplified and the measurement would be easier to be conducted. Nevertheless, the accuracy of the estimation is not satisfactory. The empirical model demands a large amount of experiment on the battery when the inside information of it remains unclear, which is impractical and inaccurate in estimation [14].

Manuscript received 11 March 2023; revised 26 June 2023; accepted 8 July 2023. Date of publication 1 August 2023; date of current version 21 August 2023. This work was supported by the Innovation Fund of Glasgow College, University of Electronic Science and Technology of China. The Associate Editor coordinating the review process was Dr. Zhibin Zhao. *(Corresponding author: Huan Wang.)*

Tianyou Bai is with the Glasgow College, University of Electronic Science and Technology of China, Chengdu 611731, China.

Huan Wang is with the Department of Industrial Engineering, Tsinghua University, Beijing 100084, China (e-mail: huan-wan21@mails.tsinghua.edu.cn).

Digital Object Identifier 10.1109/TIM.2023.3300451

In recent years, data-driven estimation algorithms have attracted more and more attention, for they do not involve internal mechanisms and can be adopted in expanded cases [15]. In the past decades, traditional data-driven models such as support vector machine (SVM) [16], particle filtering (PF) [17], and k-nearest neighbors (KNNs) [18], have been proposed by former research. However, as artificial intelligence develops, data-driven methods using deep learning and neural networks (NNs) have been the central topic of SOH prediction research and greatly elevated the accuracy of estimation. For example, Khumprom and Yodo [19] adopted dense neural network (DNN) to perform SOH prediction for Li-ion batteries. This introduction of deep learning has greatly promoted the efficiency of SOH predictors. Valant et al. [20] carried out a 1-D convolutional neural network (CNN) autoencoder. In [21], charging and discharging signals are fed to a 2-D wavelet CNN with multiple-error correcting block. However, these models are all limited by a low learning ability to temporal features. To improve the model performance when receiving time-series inputs, methods utilizing recurrent neural network (RNN) and its variants have been proposed by many researchers [22]. As an example, Zhao et al. [23] build a prognostic framework with a CNN-RNN architecture. In [24], battery signals including time, voltage, current and temperature information were imported to a long short-term memory (LSTM) model. Later on, a deep prognostic framework based on variant LSTM was proposed in [25]. He and Wu [26] proposed a domain adaptation method based on causal analysis, attention mechanism and Mogrifier-LSTM (CAM-LSTM) combining the attention mechanism with LSTM, and conducted a constrained adversarial domain adaptation to reduce the negative transfer created by traditional domain adaptation approaches. Meng et al. [27] battery prognostic method with LSTM processing partial incremental capacity features. The attention mechanism can improve the study of key features in battery signals. Base on that, self-attention mechanism applies attention between the elements within input sequences. This mechanism has been adopted by many researchers. For example, Wang et al. [28] introduced self-attention to study the global temporal dependencies of the input. In the work of [29], multivariate time series was fed to a dual self-attention network, which contains a local attention unit and global attention unit.

Nevertheless, there are still several limitations in the existing models. Firstly, as mentioned above, mechanical systems powered by Li-ion batteries usually operate in harsh environments. The signals gained from sensors often contain a large amount of background noise and abrupt fluctuation, which can also be found in various datasets (e.g., NASA dataset and Oxford dataset). These defective data would lead to poor performance in existing work. Moreover, since battery aging is a lengthy process, the input for the estimation model is usually a long time-series signal [30]. Though RNN and LSTM-based methods presented by Zhang et al. [31] and Wang et al. [32] are capable of learning the time dependencies mounted in features, there is still some room for improvement when processing long cycles data [33].

To cover these deficiencies, the primary work of this study is to explore a battery health monitoring framework with good noise tolerance and global information perception ability. In order to achieve global information perception, it requires a neural network with a global feature learning ability, where the total temporal correlation features can be fused. This global information perception mechanism will help the model to have a better learning performance for long-term feature correlations and alleviate the influence of local abrupt noise. Meanwhile, in order to increase the local feature learning ability, filters with a small receptive field could be helpful to filter the abrupt changes in the local region.

Based on the above motivations, this study proposes a novel convolution transformer-based multiview information perception framework (MVIP-Trans). MVIP-Trans attempts to build a multiscale information perceptive mechanism from both local and global perspectives, which can be realized by an effective integration of CNN and transformer. The local-global perception system can help us solving the mentioned issues in SOH estimation. To ignore abrupt noises, the battery data are required to be processed from a global view, which can be achieved by transformer architecture. With the multihead attention mechanism in the transformer, the model is allowed to process the overall input in one glance. This property alleviated the impact of noises. At the same time, it would be better to increase the learning ability for local features, which can be achieved from a local view perception. For the model, a local information perceptron (LIP) is constructed with CNN and parallel multiscale attention (PMS-A). With a local receptive field, the CNN part is capable of extracting the detailed information from discharging signals of Li-ion batteries; while PMS-A ameliorates the feature filtering ability of the perceptron, thus amplifying the useful information and attenuating the impact and abrupt noises. Apart from that, a transformer structure is employed by MVIP-Trans, where the global temporal dependencies from the local features are encoded. This architecture allows an effective global information perception and optimizes the learning capability of the framework when dealing with long time-series inputs.

By conducting this research, we made the following contributions:

1) A novel battery monitoring framework is constructed based on the transformer and CNN-attention mechanism. In this framework, both global and local battery information can be perceived and encoded from multiview, and the long-term dependencies among discharging cycles can be matched hierarchically.
2) This article introduces a LIP based on CNN and multiscale attention, which can significantly improve the learning of local features and optimize the traditional transformer encoder.
3) This article conducted experiments on two Li-ion battery datasets, and the results showed that the proposed algorithm has excellent battery state prediction capabilities, which can provide valuable information for battery management systems.

The article is arranged as follows. Chapter II describes the detailed architecture of this framework. In chapter III, the

effectiveness and superiority of the MVIP-Trans are verified. Chapter IV summarizes the article.

II. Methodology

A. Problem Statement

In this research, the deep correlation between discharging signals and battery state of health is modeled, so that the working status and health condition of Li-ion batteries could be monitored online. To achieve that, a prognostic model is required. Assumed to work in an open complex environment (such as electric vehicles) [34], the expected $\Im(\omega_p)$ is driven by battery signals, from which the battery capacity $\widehat{C_n^N}$ could be predicted.

From multiple internal indexes measuring platforms, we obtained the electric properties of Li-ion batteries and constructed a processed dataset $D_t = \{B\}_{n=1}^N$, where N represents the number of Li-ion batteries in the dataset, n denotes the number of discharge cycles and B_n^N can be represented as $B_n^N = \left[(V_1, I_1, T_1, t_1)^n, \ldots, (V_c, I_c, T_c, t_c)^n\right]$, $B_n^N \in \Re^{c\times 4}$, c is the sampled time, V, I, T, t are the measured voltage, current, temperature and time value, respectively. From an accurate impedance measurement system, the battery capacity in each discharging cycle was ascertained. Therefore, a training dataset $D_t = \{B_n^N, C_n^N\}$ could be prepared, where B_n^N is the input of our model, C_n^N is the actual capacity of the battery.

Analyzing this dataset, we found that the capacity data may be significantly influenced by abrupt noises, some obvious sharp interruptions can be spotted. Moreover, the input sequence of this model is a long-period temporal series, which requires a high learning ability for the long-term temporal dependencies. In order to achieve a high predicting accuracy under this circumstance, the prognostic model $\Im(\omega_p)$ is capable of filtering the noise signal from a global view and modeling the local correlations from a local view.

To generate SOH estimation, we first predict the capacity variation of Li-ion batteries. The training dataset is fed to our prognostic framework, where a deep learning model $\Im(\omega_p)$ calculates the gradients $\text{Loss}(C_n^N, \widehat{C_n^N})$, updates the learning weights ω_p and optimizes the hyper-parameters p. In this way, based on the principles of fully supervised learning, the connotative correlation between the discharging features and battery capacity is modeled by $\Im(\omega_p)$. At the terminal of this framework, the predicted capacity values of target batteries are utilized to calculate the SOH value of the battery, according to the following equation:

$$\text{SOH}^N = \frac{\widehat{C_n^N}}{C_1^N}. \tag{1}$$

Here, SOH^N denotes the Li-ion battery N, C_1^N represents the original capacity of battery N measured at the first cycle.

B. MVIP-Trans Framework

To address the mentioned issues, this research proposes a CNN-transformer-based SOH prognostic framework, MVIP-Trans, which makes use of different perception scales of CNN and transformer to synthetically encode dependencies between battery signals. As shown in Fig. 1, MVIP-Trans is composed of three sections: data processing, LIP, and global information perceptron (GIP). During the data processing, the discharging signals of multiple Li-ion batteries are collected, from which the voltage, current, temperature, and time data are selected and fused. After data preparation, MVIP-Trans utilizes a CNN-based Attention module to learn the detailed features. This module consists of a 2-D CNN, a PMS-A block, and a 1-D CNN. The PMS-A block is intended to increase the temporal dependencies modeling ability of the perceptron. Subsequently, the encoded local features are fed to an GIP constructed with a transformer encoder. This block adopts multihead self-attention mechanism to learn the long period temporal dependencies of local features. The self-attention part is capable of receiving signals from a wide time domain, thus improving the global information perceiving ability. With the local-global information perception mechanism, MVIP-Trans studies the deep relationship between SOH and discharging features of Li-ion battery from multiple views, hence the health condition of batteries can be estimated and monitored.

C. Local Information Perceptron

To increase the local information perception ability, this model utilized CNNs to extract the neighborhood dependencies and time-domain locality from upstream signals. Inspired by the CNN-LSTM design in [25] and [35], this article proposed a LIP with multiple CNNs and PMS-A block. which impels CNNs to make full contribution to local feature learning. The architecture of LIP is visualized in Fig. 2.

To begin with, a training dataset $D_t = \{B_n^N, C_n^N\}$ is derived. With a 2-D CNN, each input B_n^N is converted to a k-channel signal, where k is the output channel of the CNN layer. Each channel of the convolution result is taken as the packed slices for the positional embedding of the transformer, that is $B_n^N = [\mathcal{B}_1^n, \mathcal{B}_2^n, \ldots, \mathcal{B}_k^n]$, $\mathcal{B}_k^n \in \Re^{(c/k)\times 4}$. Each slice of the input signal contains local features of the input data, and is sequentially scanned by convolutional filter K, $K = [\mathcal{K}_1, \mathcal{K}_2, \ldots, \mathcal{K}_{n=32}]$, $\mathcal{K}_n \in \Re^{4\times 4}$. As shown in the following equation and Fig. 2, the features in local receptive regions are learned by 2-D CNN and the feature maps of input signals is outputted:

$$S_1 = \phi\left(B_n^N K\right) = \phi\left(\sum_n \sum_k \mathcal{B}_k^n \mathcal{K}_n\right) \tag{2}$$

ϕ denotes the rectified linear unit (ReLU) function, which increases the nonlinearity of the neural network output; S_1 represents the response of 2-D CNN.

While studying the local features from input signals, this model also utilizes an attention mechanism to capture the time-domain dependencies between each feature, which refines the model's studying of expected features. The attention mechanism can also improve the study of key features in battery signals and filter unrelated information. In order to sufficiently utilize the localized features in long sequences, MVIP-Trans introduces a PMS-A [36] based on the combination of CNN and attention mechanism. At first, the encoder of PMS-A takes a sequence of temporal embeddings $(s_1^1, s_1^2, \ldots, s_1^k)$ as

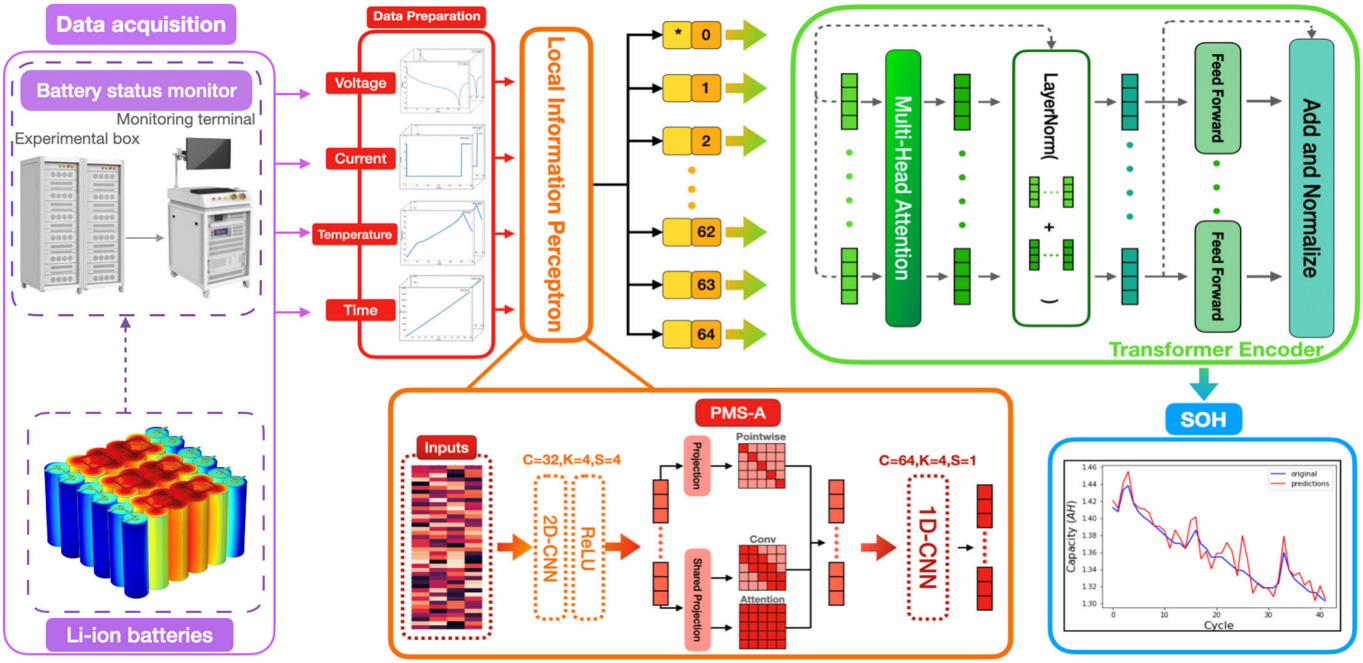

Fig. 1. Detailed architecture for MVIP-Trans.

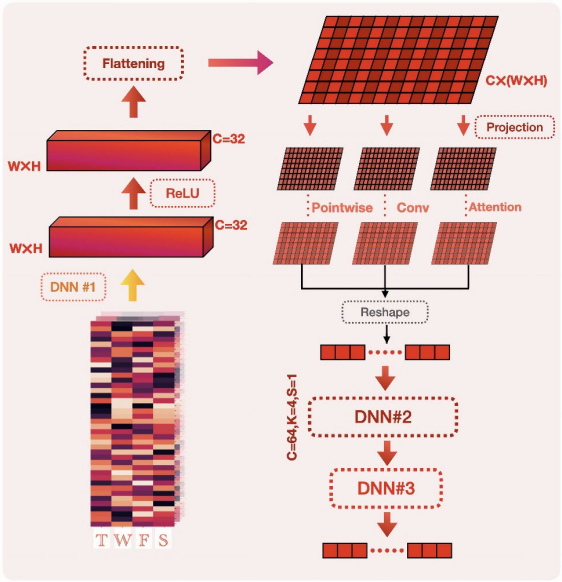

Fig. 2. Detailed workflow of PMS-A and LIP is shown in this image.

inputs, where k is the length of the input and it equals to the channel of 2-D CNN. The module then generates two copies of this input, called the "projection" and the "shared projection." The projection of input is processed by a point-wise transformation. The point-wise transformation extracts the token features of inputs with linear transformation at different points. The shared projection of the input is processed by a hybrid of CNN and self-attention layer. The output of PMS-A S_2 would be the summation of the three branches of sequence. Detailed process can be described by

$$S_2^i = S_2^{i-1} + \alpha\left(S_2^{i-1}\right) + \beta\left(S_2^{i-1}\right) + \gamma\left(S_2^{i-1}\right) \tag{3}$$

$$\alpha(X) = \max(0,\ X\ W_1 + b_1) \cdot W_2 + b_2 \tag{4}$$

$$\beta(X) = \sum_{i=1}^{n} \frac{e^{\alpha_i}}{\sum_{j=1}^{n} e^{\alpha_j}} \cdot \mathrm{Conv}_{k_i}(X) \tag{5}$$

where α means the filtering of the point-wise feed-forward network, β refers to convolution operation, γ refers to self-attention block, and X is the ideal input of these networks; the conduction of the self-attention network is demonstrated by the following equation:

$$\gamma(X) = \sigma\left(X\omega^Q, X\omega^K, X\omega^V\right)\omega^o \tag{6}$$

where ω^Q, ω^K, ω^V, ω^O denote the projection parameter of the attention layer; Q, K, V and are the linear network responses of the input signals, which can be shown as $Q, K, V = \mathrm{dense}_1(x), \mathrm{dense}_2(x), \mathrm{dense}_3(x)$; the σ operation represents the dot production of "key," "query," and "value" pairs

$$\sigma(Q, K, V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V. \tag{7}$$

After that S_2 is delivered to a 1-D CNN, which smoothly refilters the local features and fuses the temporal dependencies of self-attention responses. Here, the decoded temporal dependencies S_2 is scanned by a set of 1-D convolution kernels $\omega = [\omega_1^1, \omega_2^1, \ldots, \omega_{n=64}^1]$, $\omega_n^1 \in \Re^{4\times 1}$. As shown in the following equation, the output of LIP $S^n = [s_1^n, s_2^n, \ldots, s_k^n]$ can be derived as:

$$\begin{aligned} s_1^n(t) &= \left(S_2^1 \cdot \omega\right)(t) = \sum_{a=0}^{T} S_2^1(a)\omega(t-a) \\ &\vdots \\ s_k^n(t) &= \left(S_2^k \cdot \omega\right)(t) = \sum_{a=0}^{T} S_2^k(a)\omega(t-a). \end{aligned} \tag{8}$$

Therefore, the feature map of LIP S is obtained, which sufficiently reflects the temporal dependencies of the local features from a high-level representation of decoded battery information.

By introducing the PMS-A module, the advantage of the attention mechanism is utilized. By refining the study of local features, the important features are focused on and further studied by the network, while the impact of unrelated information such as noise is alleviated and even neglected. These properties help us to filter the noises and interrupts in the data. At the same time, unlike the traditional attention module, PMS-A hybrids both the CNN layer and attention layer, which makes the module sensitive to both long-term and short-term structures.

D. Global Information Perception Module

In LIP models, more useful local features for SOH prognostication are encoded, and better noise resistance is achieved. However, to process long-term sequential input, a neural network that receives global information and generates long-term dependencies is of necessity. Here, we propose a global information perceptron based on vision transformer. Transformer is an emerging architecture specialized at processing long-range sequential inputs. With multihead attention mechanism, it is capable of learning the temporal dependencies from arbitrary positions regardless their time-domain distance. Therefore, this article introduces transformer encoder to model the correlation between decoded local features.

1) Positional Embedding: As shown in Fig. 2, long-sequence feature maps with multiple channels are obtained from the decoder of LIP. However, this result cannot be directly adopted by GIP, for it ceases to reflect positional correlations. Therefore, the result S is first processed by a CNN layer, which divides S into multiple packs. Each pack was then attached with positional embedding sequence, $\mathrm{PE} = [p_1^n, p_2^n, \ldots, p_{64}^n]$, $p \in \Re^{e\times 1}$ which allows the transformer encoder to understand the time sequential relationship of the features

$$X^n = [S, \mathrm{PE}] = \left[x_1^n, x_2^n, \ldots, x_k^n\right], \quad \text{where}$$
$$x_k^n = \left[s_k^n, p_k^n\right] \in \Re^{(T/k+e)x_1}. \tag{9}$$

As shown in (9), the embedded inputs X^n is derived and then delivered to the self-attention network. Here, the positional embedding operation assigns a value to each input, which reflects the time-domain distribution of input signals and helps the encoder to analyze the global temporal context information.

2) Multihead Self-Attention: While receiving inputs from full view, this article also expects to maintain the multiple channel operation from CNN, which avoids excessive attention is paid to the positional correlations between features. Therefore, the transformer adopts a multihead attention block (MHA), which consists of a series of parallel scaled dot-product attention as a global scaler. In MHA, these dot-product attention receive the input projection independently, and do not share the trainable parameters. They are designed to receive complete long-scale temporal context and aggregate such information to learn the global features $Y^n = \left[y_1^n, y_2^n, \ldots, y_k^n\right]$.

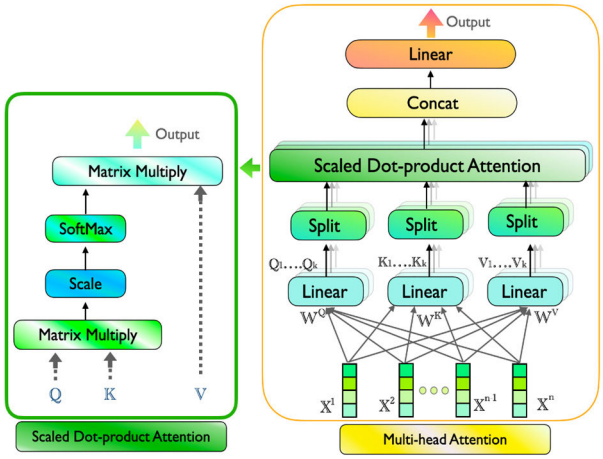

Fig. 3. Exhaustive structure of transformer encoder.

As shown in Fig. 3, the transformer encoder is fed with embedded inputs X^n, the inputs are then projected and processed by three linear neural networks, whose trainable weight matrices can be denoted as W^Q, W^K, W^V. Subsequently, the three linear transformations are employed to project X^n to Q (query vector), K (key vector), V (value vector), respectively. The operation of a single attention head can be formulated as

$$Q_k = \mathrm{dense}_q\left(X^n\right) = X^n \cdot W^Q \tag{10}$$
$$K_k = \mathrm{dense}_k\left(X^n\right) = X^n \cdot W^K \tag{11}$$
$$V_k = \mathrm{dense}_v\left(X^n\right) = X^n \cdot W^V. \tag{12}$$

In the scaled dot-product attention section, the Q_k and K_k are multiplied and scaled as

$$\mathrm{sim}_k = \frac{1}{\sqrt{d}} Q_k K_k^T \tag{13}$$
$$y_k^n = \mathrm{soft\,max}(\mathrm{sim}_k) V_k = V_k \frac{e^{\mathrm{sim}_k}}{\sum_i e^{\mathrm{sim}_k^i}}. \tag{14}$$

In (13), by conducting the dot product of Q and K, the similarity of two vectors is calculated, and then this similarity is imported to a soft max layer to attain the attention weights. Then, the learnable attention weights are coped with value vectors to obtain the product of one attention head

$$Y^n = \mathrm{dense}_l\left(\mathrm{MHA}\left(X^n\right)\right) = \left[y_1^n, y_2^n, \ldots, y_k^n\right] \cdot W^L. \tag{15}$$

As shown in Fig. 3 and (14), the encoded results of all attention heads are fused together, and delivered to a linear neural network, whose trainable parameters can be denoted as W^L. In this layer, the global features learned by all the attention heads ought to be merged, and the merged result Y^n forms the final output of MHA.

3) Position-Wise Feed-Forward Networks: After the multihead attention module, the data require successive modeling to reshape the outputs and export SOH estimation. This task will be done with a positional-wise feed-forward network (PFFN), which consists of a series of residual connections, linear transformations, and layer normalizations. To begin with, the

embedded inputs and attention outputs were matched by a residual connection. Then, a layer normalization is conducted to expedite the training of the model and decrease the internal covariant shift

$$Z_B^n = \frac{(X^n + Y^n) - \text{Ave}(X^n + Y^n)}{\left(\sigma^2 + \varepsilon\right)^{1/2}} \tag{16}$$

where Ave and σ^2, respectively calculates the mean value and variance of the encoded features.

After this, the scaled features Z_B^n are sent to a feed-forward module with two dense layers, which reshapes the output to match with training labels. In the end, the final output is obtained from another residual connection and normalization layer

$$Z_F^n = \text{dense}_p^1\left(\phi\left(\text{dense}_p^2\left(Z_B^n\right)\right)\right) \tag{17}$$

$$\text{pred} = \frac{\left(z_B^n + z_F^n\right) - \text{Ave}\left(z_B^n + z_F^n\right)}{\left(\sigma^2 + \varepsilon\right)^{1/2}}. \tag{18}$$

E. Implementation Details

As described above, a prognostic model $\Im\left(\omega_p\right)$ is completed, which is capable of generating SOH predictions by processing the external information of Li-ion batteries. The framework, MVIP-Trans, is a combination of both LIP and GIP. At the end of MVIP-Trans, a multilayer perceptron (MLP) is employed to generate the predicted values of SOH. To train this model, we first take the early 70% of the discharging cycles as training set. The training set was then shuffled. The training epoch is 157. For each epoch, an input with shape of [bs, 660, 4] was entered to the model, where bs is the batch size. The model produced a predicted capacity value based on the input information, and compared the prediction with the actual value. Here, a cost function MSE-Loss is introduced to quantify the disparity between the predicted and actual SOH values

$$\text{MSE} - \text{Loss} = \frac{1}{N}\sum_{i=1}^{N}\left(\widehat{C_n^N} - C_n^N\right)^2. \tag{19}$$

Then, based on the back propagation method, the weights in $\Im\left(\omega_p\right)$ are updated to minimize MSE-Loss. Hyperparameter setting and selection is an important issue for deep learning models. In this study, we performed a grid search on the hyperparameters of the proposed MVIP-Trans under the same dataset and experimental conditions to determine the optimal hyperparameter settings. The training of MVIP-Trans is conducted under the hyperparameter configuration $\{K_n^1 = 32, K_S^1 = 4, s_1 = 4, K_n^2 = 64, K_S^2 = 10, s_2 = 1, dk = 165, dv = 165\}$, where K_n refers to the output channel of CNNs, K_S denotes the kernel size, s means the stride of each kernel, and dk and dv denote the depth of the attention matrices; while for transformer encoder, the configuration is $\{\text{dim} = 165, \text{dim}_h = 64, \text{dim}_{\text{mlp}} = 3, h = 8\}$, where the dim, dim_h, dim_{mlp} denote the depth of the input, attention heads and MLP, and h denotes the number of attention heads. Here, Algorithm 1 describes the training process of MVIP-Trans.

Algorithm 1 Multiview Information Perceptive Convolution-Transformer Framework

Input: Li-ion battery dataset, $D_t = \left\{B_n^N, C_n^N\right\}$ trained with mini-batch

Output: SOH predicting model $\Im\left(\omega_\rho\right)$

1: Set $n = 0$ and epoch$= E$; set k, H, $\hbar = 0$

2: Initialize parameter ω_p; scale the training dataset D_t

while $n \le E$ **do**

Divide B_n^N into k segments $B_n^N = \left[\mathcal{B}_1^n, \mathcal{B}_2^n, \ldots, \mathcal{B}_k^n\right]$

Conduct
$$\begin{cases} s_1^n(t) = \left(S_2^1 \cdot \omega\right)(t) = \sum_{a=0}^{T} S_2^1(a)\omega(t-a) \\ \quad\vdots \qquad\qquad \text{to get } S^n = \left[s_1^n, s_2^n, \ldots s_k^n\right] \\ s_k^n(t) = \left(S_2^k \cdot \omega\right)(t) = \sum_{a=0}^{T} S_2^k(a)\omega(t-a) \end{cases}$$

Preform positional embedding to get $X^n = [S, PE] = \left[x_1^n, x_2^n, \ldots, x_k^n\right]$

while $\hbar \le H$ ***do***

Perform MHA to extract global dependencies $Y^n = \text{dense}_l(MHA(X^n))$

Call Feed-Forward Network to get *pred*

Set $\hbar = \hbar + 1$

end while

Calculate the predict results of $\Im\left(\omega_p\right)$: $\widehat{C_n^N} = \Im\left(B_n^N \mid \omega_p\right)$

Conduct $\partial MSE - Loss\left(C_n^N, \widehat{C_n^N}\right)/\omega_p$

Update ω_p during back propagation

Set $n = n + 1$, if an epoch is completed

end while

III. Experimental Validation

A. Experimental Setup

The MVIP-Trans is implemented with Python 3.8 and Pytorch 1.9.2. All the models mentioned in this article are trained and testified on a work station with Ubuntu 16.04 operating system, an Intel 12900K CPU and a RTX 3090Ti GPU. The models are fed with labeled data, and driven by Adam optimization algorithm and MSE critic function. To fully study the deep correlations between SOH and signals, the models are initialized as $\{\text{lr} = 0.001, \text{bs} = 64, \text{ep} = 157, \text{dr} = 0.1\}$, where lr, bs, ep, dr represent the learning rate, batch size, training epoch, and dropout rate of this model. The network weights are initialized following the Glorot normal distribution initialization method.

To discuss the effectiveness and efficiency of our model, a series of ablation experiments and comparative experiments were conducted on two different datasets (NASA PCoE dataset and Oxford dataset). In the ablation experiment, MVIP-Trans was disassembled into different networks to reveal the function of each component. We respectively disabled the transformer block attention convolutional neural network (ATT-CNN), the CNN block and attention block (transformer network), and the attention block (CNN-Trans network). All these components were trained under the same configuration and environment. In this way, the effectiveness of each part of the model can be fairly evaluated. To show the superiority of MVIP-Trans, a comparative experiment was conducted,

where MVIP-Trans and other existing methods was testified on the same dataset. Existing methods with similar ideas and different structures were carefully implemented according to the existing work, including active-state-tracking long short-term memory (AST-LSTM), auto-encoder based long short-term memory (AE-LSTM) and ViT. AST-LSTM model optimizes the original LSTM model by introducing structural variants (such as pee hole and 1-couple), which simplifies the LSTM structure and accelerates the training of this model. AE-LSTM is a combination of Auto-encoder and LSTM network, thus improving the long-term feature learning ability of the model. ViT adopts the vision transformer architecture to globally study the long-term temporal dependencies of the inputs. To ensure the fairness of our comparison, all the comparative models were operated under the best hyper-parameter configuration according to the existing works.

In order to objectively reveal the performance of MVIP-Trans and comparative methods, this research adopts four error matrices to evaluate SOH prediction, which is the mean squared error (MSE), the mean absolute error (MAE), root MSE (RMSE) and confidence of determination (R^2-score). These indexes will reflect accuracy and efficiency of these models from different respects

$$\mathrm{MAE} = \frac{1}{N}\sum_{n=1}^{N}\left|C_n^N - \widehat{C_n^N}\right| \tag{20}$$

$$\mathrm{MSE} = \frac{1}{N}\sum_{n=1}^{N}\left(C_n^N - \widehat{C_n^N}\right)^2 \tag{21}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(C_n^N - \widehat{C_n^N}\right)^2} \tag{22}$$

$$R^2\text{-score} = 1 - \frac{\sum_{n=1}^{N}\left(\widehat{C_n^N} - \overline{C_n^N}\right)^2}{\sum_{n=1}^{N}\left(C_n^N - \overline{C_n^N}\right)^2} \tag{23}$$

where C_n^N is the real battery capacity; $\widehat{C_n^N}$ is the predicted battery capacity, $\overline{C_n^N}$ is the mean value of real battery capacities. Smaller MAE, MSE, and RMSE will indicate a higher prediction accuracy, while on the contrary the larger, R^2-score the higher the prediction accuracy of the model.

B. Case 1: Validation on NASA Dataset

1) Data Description: In order to testify to the efficiency and performance of MVIP-Trans, this study chooses a series of battery-discharging signals from the NASA PCoE dataset [37] as experimental data to conduct multiple experiments. In the NASA dataset, Li-ion batteries were exposed to two different working procedures (charge and discharge spectroscopy) at varying temperatures. From the NASA dataset, we select 34 sets of Li-ion batteries (B005-007, 18, 25-32, 33-34, 36, 38-56). For each battery signal, there are multiple charging and discharging cycles. In each cycle, the voltage and current values are measured at sampled times. In this research, the voltage (V), current (I), temperature (T), capacity (C), and step-time (t) information during discharging was extracted, and for each battery, the data were reshaped into a 3-D array with a shape of [cn, 660, 4], where cn is the total cycle number of the battery dataset. Based on this aggregation, we allocated the former 70% of discharging cycles to training dataset and later 30% of cycles for testing dataset. Therefore, the shape of training data would be [0.7cn, 660, 4], the shape of testing data would be [0.3cn, 660, 4]. The training dataset was then randomly shuffled.

2) Effectiveness Verification of MVIP-Trans: In this part, the effectiveness of each mechanism in MVIP-Trans is discussed. In order to observe the actual contribution of each network, four independent network models are constructed, which are transformer, ATT-CNN, CNN-Trans, and MVIP-Trans. Transformer is constructed based on a transformer encoder and three MLP blocks, each block is a single-layer perceptron coped with a batch normalization layer. ATT-CNN adopts the same architecture with LIP block in VIP-Trans, which is a combination of 2-D-CNN, 1-D-CNN and PMS-A. CNN-Trans is composed of a CNN block and a transformer encoder, and the CNN block is made up by both 2-D-CNN and 1-D-CNN. MVIP-Trans is the proposed method combined with both the CNN-Attention block and transformer encoder. The prediction results are obtained by processing NASA B005, B006, B007 datasets, which are illustrated in Table I.

By comparing the prediction results, the proposed MVIP-Trans can be confirmed to have better accuracy than the comparative models. For example, for B005, the RMSE for ATT-CNN, transformer, CNN-Trans, MVIP-Trans are 0.027, 0.012, 0.010, 0.006, respectively. Particularly, the RMSE of MVIP-Trans is only 50.03% and 22.18% of the transformer and ATT-CNN, which proves that by introducing multiview information perceptron, the performance of MVIP-Trans is significantly optimized. Apart from that, the result has shown the superiority of transformer architecture compared with ATT-CNN. According to Table I, the accuracy of ATT-CNN is the lowest among the four methods, which is reasonable considering the local receptive field of CNN is short at learning long-term features. According to the experiment result, it can be confirmed that the LIP based on convolutional attention mechanism can be tacitly combined with the transformer structure.

To further observe the conduction of prognostic models, the SOH prediction curves (B005, B006, and B007) are displayed in Fig. 4. The blue line is the real age variation of each battery, while the red line shows the average prediction output of each model. The yellow dots in the figure represent the average predicted capacity at each cycle. The cyan region marked in the figure indicates the confidence integral of our prediction results. From the set of figures, it can be seen that while there are certain divergences between the prediction values and true capacity, MVIP-Trans is still capable of forecasting the decay trend of the SOH value. Compared with ATT-CNN, the MVIP-Trans prediction curves are closer to the true values, and the deviations are slighter. Compared with the transformer, the prediction results of MVIP-Trans are much more stable and robust. Therefore, it is intuitively proved that the adopted LIP improves the robustness and noise tolerance of MVIP-Tran, while the GIP promotes the fitting accuracy of the proposed model.

TABLE I
AVERAGE ESTIMATION ACCURACY OF EACH EXPERIMENTAL MODELS

Method	MAE	RMSE	MSE	R^2-score
NASA-B005				
ATT-CNN	4.794×10^{-2}	0.019	3.614×10^{-4}	0.886
Transformer	0.906×10^{-2}	0.011	1.25×10^{-4}	0.939
CNN-Trans	0.876×10^{-2}	0.009	0.824×10^{-4}	0.975
MVIP-Trans	0.527×10^{-2}	0.006	0.358×10^{-4}	0.987
NASA-B006				
ATT-CNN	3.924×10^{-2}	0.017	2.892×10^{-4}	0.897
Transformer	1.283×10^{-2}	0.012	1.447×10^{-4}	0.951
CNN-Trans	0.808×10^{-2}	0.009	0.824×10^{-4}	0.972
MVIP-Trans	0.489×10^{-2}	0.006	0.371×10^{-4}	0.986
NASA-B007				
ATT-CNN	4.212×10^{-2}	0.017	2.886×10^{-4}	0.893
Transformer	0.894×10^{-2}	0.010	1.104×10^{-4}	0.958
CNN-Trans	0.764×10^{-2}	0.007	0.479×10^{-4}	0.981
MVIP-Trans	0.408×10^{-2}	0.005	0.274×10^{-4}	0.989

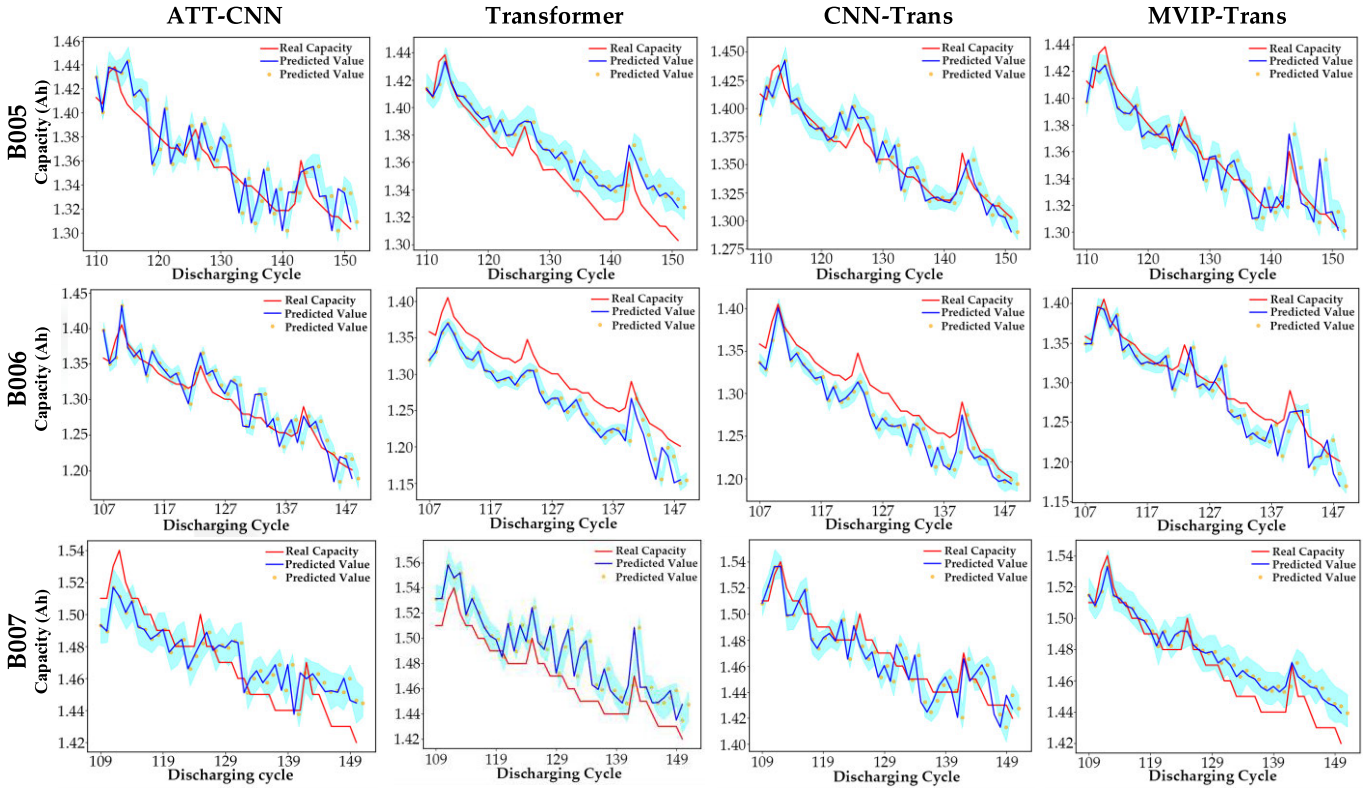

Fig. 4. Predicted capacity curves of ATT-CNN, transformer, CNN-Trans, and MVIP-Trans on B005, B006, and B007.

To summarize, by conducting this experiment, it is proved that the performance of MVIP-Trans is indeed improved by introducing a multiview information perception mechanism based on a combination of LIP and GIP. The adoption of LIP successfully improved the noise tolerance and local feature learning of MVIP-Trans, and by introducing GIP, the long-term feature learning ability is upgraded.

3) Comparison Between Existed Methods: To demonstrate the superior performance of MVIP-Trans on SOH estimating, in this section, we conduct a horizontal comparison experiment between the proposed method and other existing models. Here, the MVIP-Trans is compared with the other three deep-learning models, including AST-LSTM [38], AE-LSTM [39], and ViT [40].

All the comparative models are carefully implemented under their optimized environment and configuration and are trained with the same hyperparameters as the existing works. The prediction results are aggregated in Table II, and the prediction curves are collected in Fig. 5.

TABLE II
AVERAGE PREDICTION RESULTS OF EACH COMPARATIVE MODELS

Method	MAE	RMSE	MSE	R^2-score
NASA-B005				
AST-LSTM	1.996×10^{-2}	0.017	2.893×10^{-4}	0.898
AE-LSTM	2.803×10^{-2}	0.014	1.961×10^{-4}	0.916
ViT	1.454×10^{-2}	0.010	1.005×10^{-4}	0.970
MVIP-Trans	0.527×10^{-2}	0.006	0.361×10^{-4}	0.986
NASA-B006				
AST-LSTM	2.524×10^{-2}	0.014	1.963×10^{-4}	0.902
AE-LSTM	1.691×10^{-2}	0.012	1.444×10^{-4}	0.959
ViT	0.921×10^{-2}	0.008	0.642×10^{-4}	0.980
MVIP-Trans	0.478×10^{-2}	0.005	0.256×10^{-4}	0.989
NASA-B007				
AST-LSTM	2.307×10^{-2}	0.013	1.689×10^{-4}	0.909
AE-LSTM	1.856×10^{-2}	0.012	1.440×10^{-4}	0.958
ViT	0.963×10^{-2}	0.007	0.494×10^{-4}	0.983
MVIP-Trans	0.497×10^{-2}	0.006	0.357×10^{-4}	0.987

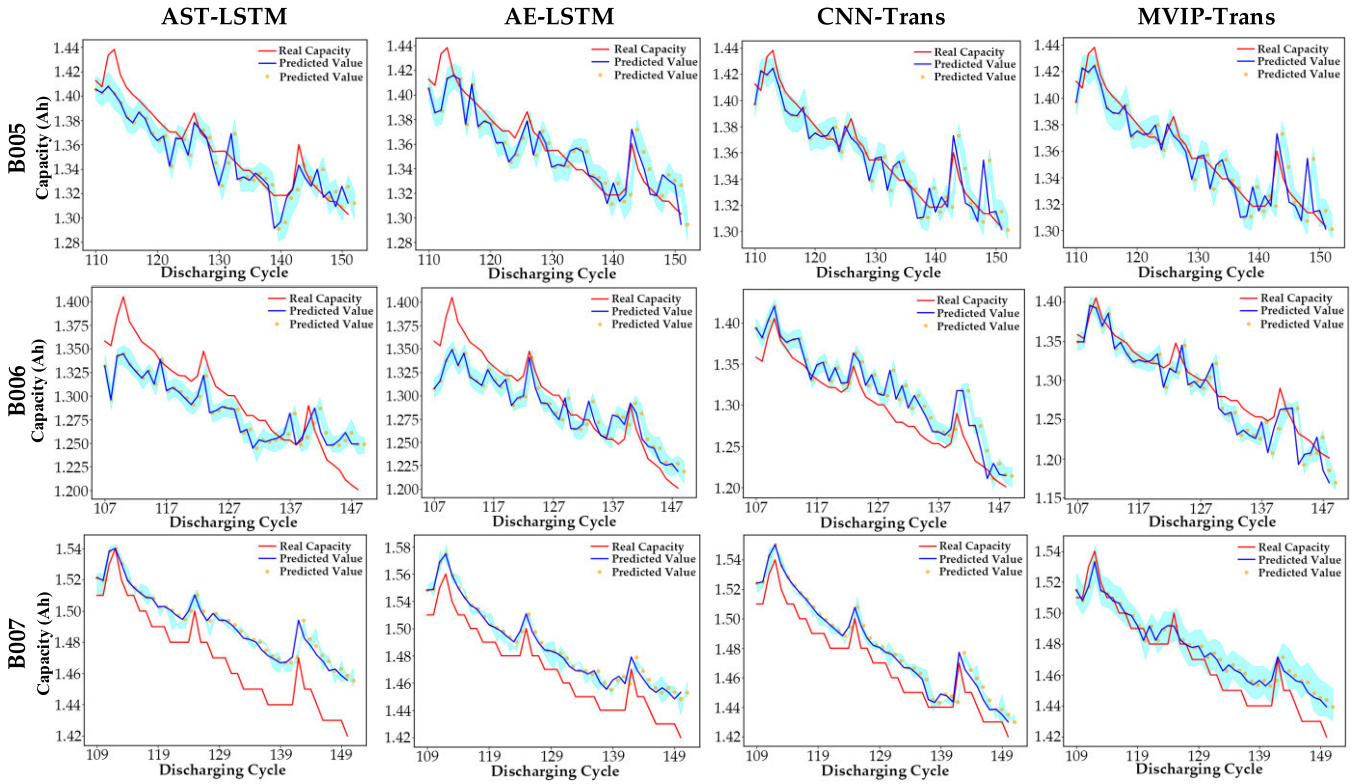

Fig. 5. Predicted capacity curves of AST-LSTM, AE-LSTM, ViT, and MVIP-Trans on B005, B006, and B007.

Obviously, the MVIP-Trans obtains the best predicting accuracy, especially in B005. For example, in B005, the RMSE of MVIP-Trans is 0.006, which is 59.97% and 35.29% of ViT and AST-LSTM models. The superiority is even more obvious in absolute value comparison, where the MAE of MVIP-Trans is only 36.24% of ViT and 26.40% of AST-LSTM. This proves that by perceiving signals from a local view, the effectiveness of transformer architecture is improved. Meanwhile, the experimental results tell that the accuracy of ViT is higher than LSTM-based model, which indirectly proves that the global information perceptron we proposed has a better long-term sequential feature learning ability.

C. Case 2: Validation on Oxford Dataset

1) Data Description: To reveal the performance of MVIP-Trans on diverse Li-ion battery types and prove the universality of our model, the same experiment was conducted using the Oxford dataset. The Oxford dataset is a composition of batteries adding measurements of 8 small Li-ion pouch cells. Each measurement was conducted in a thermal chamber

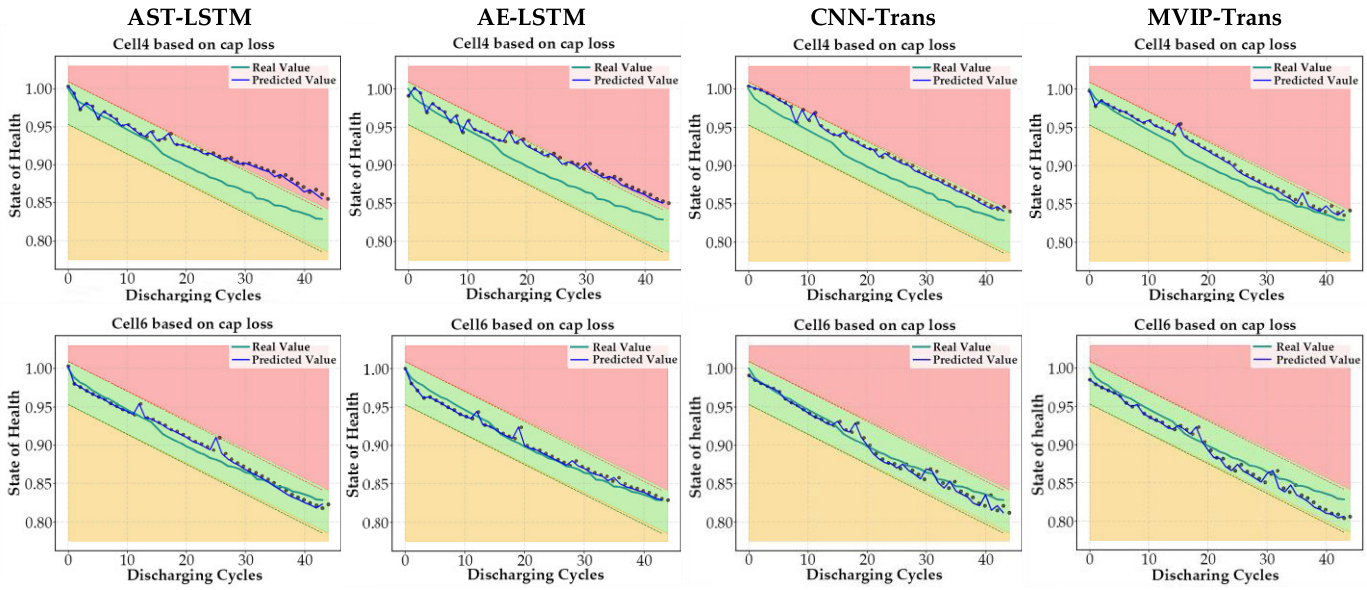

Fig. 6. Predicted aging curves of AST-LSTM, AE-LSTM, ViT, and MVIP-Trans on Oxford datasets.

TABLE III
AVERAGE PREDICTION RESULTS OF EACH MODEL

Method	MAE	RMSE	MSE	R^2-score
Cell 04				
AST-LSTM	1.427×10^{-2}	0.013	1.692×10^{-4}	0.932
AE-LSTM	0.937×10^{-2}	0.007	0.495×10^{-4}	0.969
ViT	0.512×10^{-2}	0.005	0.261×10^{-4}	0.982
MVIP-Trans	0.326×10^{-2}	0.003	0.089×10^{-4}	0.992
Cell 06				
AST-LSTM	2.524×10^{-2}	0.009	0.809×10^{-4}	0.946
AE-LSTM	0.793×10^{-2}	0.006	0.371×10^{-4}	0.972
ViT	0.534×10^{-2}	0.005	0.256×10^{-4}	0.984
MVIP-Trans	0.301×10^{-2}	0.003	0.087×10^{-4}	0.994

at 40 °C. To motivate the aging process of batteries, each cell run through a constant current charging procedure, together with a constant discharging process with multiple cycles [41]. The battery status was measured every 100 cycles. For deep learning, we select six cells for training set (Cell 1, 2, 3, 5, 7, 8) and two cells for testing (Cell 4 and 6). The maximum charges (C), voltage (V), current (I), temperature (T), and step-time (t) information during discharging were extracted, and for each battery, the data are reshaped into a 3-D array. The constructed array has similar structure with NASA dataset input, which benefits the work of MVIP-Trans. For data preparation, we assigned and shuffled the former 70% of discharging cycles to the training set and saved the later 30% of cycles for testing dataset.

2) Comparison Between Existed Methods: In order to objectively reflects the predicting accuracy of MVIP-Trans on the Oxford dataset, a comparative experiment between the above methods (AST-LSTM [38], AE-LSTM [39], ViT [40], and MVIP-Trans) was conducted. These models are implemented under the same environment and configuration, then are trained with the same hyperparameters and testified on the same dataset. Table III aggregates all the prediction results, and the prediction curves are collected in Fig. 6.

In the case of the Oxford dataset, MVIP-Trans still had the best accuracy for SOH prediction among the testified models. For example, for Cell 4, the RMSE of MVIP-Trans is 0.003, which is 60.04% and 23.07% of ViT and AST-LSTM models. The absolute value discussion shows more superiority, where the MAE of MVIP-Trans is only 33.34% of ViT and 5.89% of AST-LSTM. These results have formed solid evidence that the battery monitoring system based on MVIP-Trans can be commonly used for different kinds of Li-ion batteries and different forms of datasets. This again proves that the global-local information perceiving mechanism can significantly promote

the effectiveness of traditional transformer structures. Based on the experiment of two different datasets, it is verified that the MVIP-Trans have reached a high prognosis accuracy in all test datasets, which can sufficiently indicate the adaptability and robustness of our model.

IV. Conclusion

In this research, an end-to-end prognostic framework with a multiview information perception mechanism, MVIP-Trans, is proposed to conduct the battery health monitoring task. To attenuate the interference of background noise and abrupt signals to the feature matching process, a LIP using convolutional attention operation is constructed, which receives the local information with CNN receptive field and studies the temporal dependency between the feature mappings. Meanwhile, in order to improve the modeling ability for long-term sequential features, MVIP-Trans adopts an GIP, where the local features are integrated and encoded by a multiscale transformer encoder. This structure breaks the limitation of sequential step-length and allows the model to relate divided features regardless of temporal distance. To further study the impact of multiview information perception mechanism, a series of experiments are conducted, which includes the independent trails with separate MVIP-Trans components and parallel comparison between existing predicting models. With four evaluative indexes to reflect the performance of each predicting methods, the experiments revealed the feasibility and effectiveness of such a system in executing prognostic task with long-sequence input.

In the future, a further modification of the transformer structure will be done, which may lighten the model and solve the overconcentration problem of the attention block. More than that, a visualization block will be added, which can help us analyze the training procedure of our transformer module. Other improvement such as domain adaptation and cycle-consistency of degradation trend may be considered to increase the predicting accuracy of our model.

We believe that by introducing the MVIP-Trans framework, a novel analytical view of the Li-ion battery protection would be provided, and it would be helpful for the study of deep learning models with long-term sequential inputs.

References

[1] X. Hu, C. Zou, C. Zhang, and Y. Li, "Technological developments in batteries: A survey of principal roles, types, and management needs," *IEEE Power Energy Mag.*, vol. 15, no. 5, pp. 20–31, Sep. 2017, doi: 10.1109/MPE.2017.2708812.

[2] X. Hu, F. Feng, K. Liu, L. Zhang, J. Xie, and B. Liu, "State estimation for advanced battery management: Key challenges and future trends," *Renew. Sustain. Energy Rev.*, vol. 114, Oct. 2019, Art. no. 109334, doi: 10.1016/j.rser.2019.109334.

[3] Y. Zhang and Y.-F. Li, "Prognostics and health management of lithium-ion battery using deep learning methods: A review," *Renew. Sustain. Energy Rev.*, vol. 161, Jun. 2022, Art. no. 112282, doi: 10.1016/j.rser.2022.112282.

[4] C. Hu, H. Ye, G. Jain, and C. Schmidt, "Remaining useful life assessment of lithium-ion batteries in implantable medical devices," *J. Power Sources*, vol. 375, pp. 118–130, Jan. 2018, doi: 10.1016/j.jpowsour.2017.11.056.

[5] C. Weng, X. Feng, J. Sun, and H. Peng, "State-of-health monitoring of lithium-ion battery modules and packs via incremental capacity peak tracking," *Appl. Energy*, vol. 180, pp. 360–368, Oct. 2016, doi: 10.1016/j.apenergy.2016.07.126.

[6] P. Ding et al., "Useful life prediction based on wavelet packet decomposition and two-dimensional convolutional neural network for lithium-ion batteries," *Renew. Sustain. Energy Rev.*, vol. 148, Sep. 2021, Art. no. 111287, doi: 10.1016/j.rser.2021.111287.

[7] M. A. Hannan, M. S. H. Lipu, A. Hussain, and A. Mohamed, "A review of lithium-ion battery state of charge estimation and management system in electric vehicle applications: Challenges and recommendations," *Renew. Sustain. Energy Rev.*, vol. 78, pp. 834–854, Oct. 2017, doi: 10.1016/j.rser.2017.05.001.

[8] X. Hu, D. Cao, and B. Egardt, "Condition monitoring in advanced battery management systems: Moving horizon estimation using a reduced electrochemical model," *IEEE/ASME Trans. Mechatronics*, vol. 23, no. 1, pp. 167–178, Feb. 2018, doi: 10.1109/TMECH.2017.2675920.

[9] M. Simic, A. K. Stavrakis, V. Jeoti, and G. M. Stojanovic, "A Randles circuit parameter estimation of Li-ion batteries with embedded hardware," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–12, 2022, doi: 10.1109/TIM.2022.3183661.

[10] A. Zilio and P. Mattavelli, "An experimental validation of the adoption of DC–DC converters for the impedance measurement in Li-ion batteries," in *Proc. IEEE Int. Conf. Environ. Electr. Eng. IEEE Ind. Commercial Power Syst. Eur. (EEEIC/I&CPS Eur.)*, Sep. 2021, pp. 1–6, doi: 10.1109/EEEIC/ICPSEurope51590.2021.9584622.

[11] X. Kong et al., "Pseudo-two-dimensional model and impedance diagnosis of micro internal short circuit in lithium-ion cells," *J. Energy Storage*, vol. 27, Feb. 2020, Art. no. 101085, doi: 10.1016/j.est.2019.101085.

[12] M. Schmid and C. Endisch, "Online diagnosis of soft internal short circuits in series-connected battery packs using modified kernel principal component analysis," *J. Energy Storage*, vol. 53, Sep. 2022, Art. no. 104815, doi: 10.1016/j.est.2022.104815.

[13] Y. Xing, W. He, M. Pecht, and K. L. Tsui, "State of charge estimation of lithium-ion batteries using the open-circuit voltage at various ambient temperatures," *Appl. Energy*, vol. 113, pp. 106–115, Jan. 2014, doi: 10.1016/j.apenergy.2013.07.008.

[14] A. Shah, K. Shah, C. Shah, and M. Shah, "State of charge, remaining useful life and knee point estimation based on artificial intelligence and machine learning in lithium-ion EV batteries: A comprehensive review," *Renew. Energy Focus*, vol. 42, pp. 146–164, Sep. 2022, doi: 10.1016/j.ref.2022.06.001.

[15] N. Watrin, B. Blunier, and A. Miraoui, "Review of adaptive systems for lithium batteries state-of-charge and state-of-health estimation," in *Proc. IEEE Transp. Electrif. Conf. Expo.*, Jun. 2012, pp. 1–6.

[16] J. C. Á. Antón, P. J. G. Nieto, F. J. de Cos Juez, F. S. Lasheras, M. G. Vega, and M. N. R. Gutiérrez, "Battery state-of-charge estimator using the SVM technique," *Appl. Math. Model.*, vol. 37, no. 9, pp. 6244–6253, May 2013, doi: 10.1016/j.apm.2013.01.024.

[17] Z. Liu, G. Sun, S. Bu, J. Han, X. Tang, and M. Pecht, "Particle learning framework for estimating the remaining useful life of lithium-ion batteries," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 2, pp. 280–293, Feb. 2017, doi: 10.1109/TIM.2016.2622838.

[18] C. Hu, G. Jain, P. Zhang, C. Schmidt, P. Gomadam, and T. Gorka, "Data-driven method based on particle swarm optimization and k-nearest neighbor regression for estimating capacity of lithium-ion battery," *Appl. Energy*, vol. 129, pp. 49–55, Sep. 2014, doi: 10.1016/j.apenergy.2014.04.077.

[19] P. Khumprom and N. Yodo, "A data-driven predictive prognostic model for lithium-ion batteries based on a deep learning algorithm," *Energies*, vol. 12, no. 4, p. 660, Feb. 2019, doi: 10.3390/en12040660.

[20] C. J. Valant, J. D. Wheaton, M. G. Thurston, S. P. McConky, and N. G. Nenadic, "Evaluation of 1D CNN autoencoders for lithium-ion battery condition assessment using synthetic data," in *Proc. Annu. Conf. PHM Soc.*, Sep. 2019, vol. 11, no. 1, pp. 1–11, doi: 10.36001/phmconf.2019.v11i1.876.

[21] Y. Li, K. Li, X. Liu, and L. Zhang, "Fast battery capacity estimation using convolutional neural networks," *Trans. Inst. Meas. Control*, Nov. 2020, doi: 10.1177/0142331220966425.

[22] K. Song, D. Hu, Y. Tong, and X. Yue, "Remaining life prediction of lithium-ion batteries based on health management: A review," *J. Energy Storage*, vol. 57, Jan. 2023, Art. no. 106193, doi: 10.1016/j.est.2022.106193.

[23] F. Zhao, Y. Li, X. Wang, L. Bai, and T. Liu, "Lithium-ion batteries state of charge prediction of electric vehicles using RNNs-CNNs neural networks," *IEEE Access*, vol. 8, pp. 98168–98180, 2020, doi: 10.1109/ACCESS.2020.2996225.

[24] A. Dubey, A. Zaidi, and A. Kulshreshtha, "State-of-charge estimation algorithm for Li-ion batteries using long short-term memory network with Bayesian optimization," in *Proc. 2nd Int. Conf. Interdiscipl. Cyber Phys. Syst. (ICPS)*, May 2022, pp. 68–73, doi: 10.1109/ICPS55917.2022.00021.

[25] P. Li et al., "An end-to-end neural network framework for state-of-health estimation and remaining useful life prediction of electric vehicle lithium batteries," *Renew. Sustain. Energy Rev.*, vol. 156, Mar. 2022, Art. no. 111843, doi: 10.1016/j.rser.2021.111843.

[26] J. He and L. Wu, "Cross-conditions capacity estimation of lithium-ion battery with constrained adversarial domain adaptation," *Energy*, vol. 277, Aug. 2023, Art. no. 127559, doi: 10.1016/j.energy.2023.127559.

[27] H. Meng, M. Geng, and T. Han, "Long short-term memory network with Bayesian optimization for health prognostics of lithium-ion batteries based on partial incremental capacity analysis," *Rel. Eng. Syst. Saf.*, vol. 236, Aug. 2023, Art. no. 109288, doi: 10.1016/j.ress.2023.109288.

[28] F. Wang, Z. Zhao, J. Ren, Z. Zhai, S. Wang, and X. Chen, "A transferable lithium-ion battery remaining useful life prediction method from cycle-consistency of degradation trend," *J. Power Sources*, vol. 521, Feb. 2022, Art. no. 230975, doi: 10.1016/j.jpowsour.2022.230975.

[29] H. Wang, J. Li, X. Liu, J. Rao, Y. Fan, and X. Tan, "Online state of health estimation for lithium-ion batteries based on a dual self-attention multivariate time series prediction network," *Energy Rep.*, vol. 8, pp. 8953–8964, Nov. 2022, doi: 10.1016/j.egyr.2022.07.017.

[30] J. Yao and T. Han, "Data-driven lithium-ion batteries capacity estimation based on deep transfer learning using partial segment of charging/discharging data," *Energy*, vol. 271, May 2023, Art. no. 127033, doi: 10.1016/j.energy.2023.127033.

[31] W. Zhang, X. Li, and X. Li, "Deep learning-based prognostic approach for lithium-ion batteries with adaptive time-series prediction and on-line validation," *Measurement*, vol. 164, Nov. 2020, Art. no. 108052, doi: 10.1016/j.measurement.2020.108052.

[32] C. Wang, N. Lu, S. Wang, Y. Cheng, and B. Jiang, "Dynamic long short-term memory neural-network-based indirect remaining-useful-life prognosis for satellite lithium-ion battery," *Appl. Sci.*, vol. 8, no. 11, p. 2078, Oct. 2018, doi: 10.3390/app8112078.

[33] Y. Shichun et al., "All-climate state-of-charge estimation and equilibrium management for lithium-ion batteries based on diffusion equivalent model," *J. Energy Storage*, vol. 52, Aug. 2022, Art. no. 104700, doi: 10.1016/j.est.2022.104700.

[34] D. Xu, L. Wang, and J. Yang, "Research on Li-ion battery management system," in *Proc. Int. Conf. Electr. Control Eng.*, Jun. 2010, pp. 4106–4109, doi: 10.1109/iCECE.2010.998.

[35] Y. Ma, C. Shan, J. Gao, and H. Chen, "A novel method for state of health estimation of lithium-ion batteries based on improved LSTM and health indicators extraction," *Energy*, vol. 251, Jul. 2022, Art. no. 123973, doi: 10.1016/j.energy.2022.123973.

[36] G. Zhao, X. Sun, J. Xu, Z. Zhang, and L. Luo, "MUSE: Parallel multi-scale attention for sequence to sequence learning," 2019, *arXiv:1911.09483*.

[37] B. Saha and K. Goebel, "Battery data set," NASA AMES Prognostics Center of Excellence Data Set Repository, Washington, DC, USA, Tech. Rep., 2007.

[38] P. Li et al., "State-of-health estimation and remaining useful life prediction for the lithium-ion battery based on a variant long short term memory neural network," *J. Power Sources*, vol. 459, May 2020, Art. no. 228069, doi: 10.1016/j.jpowsour.2020.228069.

[39] M. Fasahat and M. Manthouri, "State of charge estimation of lithium-ion batteries using hybrid autoencoder and long short term memory neural networks," *J. Power Sources*, vol. 469, Sep. 2020, Art. no. 228375, doi: 10.1016/j.jpowsour.2020.228375.

[40] P. Fu, L. Chu, Z. Hou, J. Hu, Y. Huang, and Y. Zhang, "Transfer learning and vision transformer based state-of-health prediction of lithium-ion batteries," 2020, *arXiv.2209.05253*.

[41] M. Landi and G. Gross, "Measurement techniques for online battery state of health estimation in vehicle-to-grid applications," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 5, pp. 1224–1234, May 2014, doi: 10.1109/TIM.2013.2292318.

Tianyou Bai was born in Hebei, China, in 2002. He is currently pursuing the B.S. degree in electronic and electrical engineering with the University of Electronic Science and Technology of China, Chengdu, China.

His research interests include deep learning, attention mechanism, computer vision, and defect-recognition.

Huan Wang (Graduate Student Member, IEEE) was born in Hunan, China, in 1994. He received the B.S. and M.S. degrees in mechanical engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2016 and 2021, respectively. He is currently pursuing the Ph.D. in management science and engineering with Tsinghua University, Beijing, China.

He has authored or coauthored over 40 articles, including 20 journal articles indexed by the Science Citation Index (SCI) and three Essential Science Indicators (ESIs) highly cited articles. He is also a peer-reviewed reviewer for more than ten international journals. His research interests include prognostics and health management (PHM) for electromechanical systems, remaining useful life estimation, deep learning, and machine learning.

Dr. Wang has won two best conference paper awards and five international data challenge awards.