

Research Article

Research and Application of Deformation Prediction Model for Deep Foundation Pit Based on LSTM

Hailin Li,^{1,2} Zhizhou Zhao ,³ and Xue Du⁴

¹School of Construction Engineering, Chongqing Creation Vocational College, Yongchuan, Chongqing 402160, China

²Chongqing Qianwanjia Construction Engineering Co., Ltd., Chongqing Yubei 401147, China

³Hehai College, Chongqing Jiaotong University, Nan'an District, Chongqing 400074, China

⁴Chongqing Fengdu County Longkong Town Primary Middle School, Chongqing Fengdu 408217, China

Correspondence should be addressed to Zhizhou Zhao; zzz@cqjtu.edu.cn

Received 12 January 2022; Revised 20 February 2022; Accepted 16 March 2022; Published 6 July 2022

Academic Editor: Xin Ning

Copyright © 2022 Hailin Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Deep foundation pit is a door with a long history, but it has new disciplines; in this paper, firstly, the modeling method and process of LSTM (long short-term memory) network are discussed in detail, then the optimization algorithm used in the model is described in detail, and the parameter selection methods such as initial learning rate, activation function, and iteration number related to LSTM network training are introduced in detail. LSTM network is used to process the deformation data of deep foundation pit, and random gradient descent, momentum, Nesterov, RMSProp, AdaGmd, and Adam algorithms are selected in the same example for modeling prediction and comparison. Two examples of horizontal displacement prediction of pile and vertical displacement prediction of column in deep foundation pit show that the LSTM network model established by different optimization algorithms has different prediction accuracy, and the LSTM network model established by Adam optimization algorithm has the highest accuracy, which proves that the selection of optimization algorithm plays an important role in LSTM and also verifies the feasibility of LSTM network in the data processing and prediction of deep foundation pit deformation.

1. Introduction

With the progress of human society and the rapid development of national economy, the process of urbanization in China is accelerating, and the land available in urban areas is sharply reduced. Therefore, urban areas pay more attention to the development and use of underground space. A large number of underground projects have appeared, such as underground foundations for high-rise buildings, subway stations, tunnels, and underground warehouses. Most of these underground projects adopt the open excavation method with simple construction and low cost, and these underground projects are generally excavated to more than 6 meters in the construction process, so more and more deep foundation pit projects emerge. Deep foundation pit is a door with a long history but it has new disciplines, it has a long history due to human exploration of deep foundation pit has had for a long time, said novel because it is itself very worthy of further research topic, and, at this stage

in the deepening of deep foundation pit excavation depth, will still be presented many new problems in the process of excavation. It is worth exploring and solving.

In recent years, the number of deep foundation pits is increasing and getting deeper. In addition, most of the deep foundation pits are located in downtown areas of the city, and the influencing factors are complicated. As a result, the difficulty of deep foundation pit construction is greatly increased, and the problems caused by this are very prominent. Deep foundation pit accidents happen from time to time, which not only bring great threat to people's property and life safety but also increase construction time and cost [1].

There is usually a slow and long deformation stage before the collapse of deep foundation pit, so corresponding monitoring points should be set according to the deformation characteristics of different structures of deep foundation pit. Adopt in deep foundation pit engineering design of a safe threshold, under the set threshold, according to the historical data of

monitoring, implementation of migration and settlement deformation monitoring area point of real-time monitoring, as well as to the monitoring data for subsequent processing, structural deformation analysis of deep foundation pit, using reasonable forecast model for deep foundation pit deformation rule. Based on these, the late deformation value and deformation trend of deep foundation pit engineering can be predicted accurately. When the shape variable of deep foundation pit exceeds the allowable range of deformation, reasonable protection schemes should be adopted, such as strengthening protection and setting drainage facilities, so as to minimize the casualties and economic losses caused by the accident of deep foundation pit.

At present, the methods commonly used for deformation prediction mainly include regression analysis, grey theory, sequence analysis, and neural network [2]. As for these single prediction models, there are certain limitations. For example, regression analysis is a static data processing method, while deep foundation pit monitoring data is dynamic data, so regression analysis has great limitations. Grey theory has a high requirement for data, which requires data to be exponentially increasing or monotonic, but most of the deformation data of deep foundation pit are complex nonlinear and do not have monotone. Among the neural networks, BP neural network is the most widely used, but BP neural network is greatly affected by the initial value and is prone to local extremum problems. Other methods also have low applicability and are difficult to be used in deep foundation pit deformation prediction.

In addition, the construction process of deep foundation pit is a soil unloading process, which is affected by a variety of external forces, which are fuzzy and complex, and the deep foundation pit engineering is also affected by a variety of other uncertain factors. Under this condition, it is very difficult to build an appropriate model for deep foundation pit deformation prediction, which has not been well solved. Therefore, it is an effective method to build a dynamic prediction model and predict the overall variation trend and deformation characteristics of shape variants by exploring the internal relationship between time series data of monitoring points.

In recent years, with the rapid development of deep learning, it has created opportunities to construct high-precision dynamic prediction model for deep foundation pit and added effective means for deformation prediction of deep foundation pit. Deep learning model [3] is a deep neural network composed of a variety of nonlinear mapping layers, which can obtain the characteristics of input data layer by layer and obtain its deep internal characteristics. Among many deep learning models, LSTM network [4] has great advantages in sequential data processing and can extract dynamic features from successively associated data. It has been applied in many fields, such as machine translation [5], information retrieval [6], image processing [7], text recognition [8], intelligent question answering [9], speech recognition [10], and data prediction [11–13], and a large number of research achievements have been obtained. However, through extensive literature review, it is found that LSTM network is rarely studied in the field of deformation

prediction and has not been applied in the prediction of deep foundation pit deformation. In this paper, through study on characteristics of LSTM theory, it is found that the LSTM network has the ability of self-learning, being adaptive, and efficient memory, to the highly nonlinear time series data have better fitting ability, and can make short- and long-term prediction, can make a high precision prediction of deep foundation pit deformation, so as to make scientific and safe protection for deep foundation pit engineering, which has theoretical and practical significance. This paper mainly studied the deformation monitoring and prediction in the construction process of deep foundation pit. In this paper, by analyzing the deformation data obtained from the construction site monitoring and referring to the final prediction results, we can effectively give early warning of relevant risks and timely take the corresponding measures to avoid risks, so as to ensure the stability and order of the whole construction process.

1.1. Related Work of Deep Foundation Pit Deformation Prediction. Foundation pit engineering started early in foreign countries. As early as the mid-18th century, large-scale construction of railways, ports, and ports began abroad, and the research of foundation pit engineering began. In the 1940s, Terzaghi [14] proposed theoretical soil mechanics and wrote a book to lay a foundation for the theory of deep foundation pit engineering; Clough et al. [15] applied the measured data of foundation pit to statistically analyze its internal support wall; Ou [16] and Finno [17] analyzed the impact of foundation pit construction on surrounding buildings (structures) in different soils. In the 1960s, Peck [18] put forward the estimation rule of foundation pit surface settlement through example verification, which has been used so far; Bjerrum et al. [19] applied relevant instruments to the monitoring of a soft soil foundation pit in Mexico. Until the 1980s, foreign research on foundation pit began to become more precise and detailed. Clough and O'Rourke [20, 21] emphatically analyzed the relationship between the wall offset of the foundation pit and the depth of the foundation pit, ground settlement, and the stiffness of the supporting structure; Mayoraz [22] and Goh [23] used neural networks to predict the horizontal displacement and deformation of foundation pit slope and retaining wall, respectively; Mann [24] gives the relationship between the maximum displacement of the inner wall of the foundation pit and the antiuplift stability coefficient according to the results of 11 foundation pit examples; Liu et al. [25] added improved Mohr-Coulomb and Cam-Clay models into FLAC-3D software to simulate the surface deformation caused by foundation pit excavation and predicted the influence range of foundation pit; Sepehri et al. [26] established a three-dimensional finite model to predict the surface deformation data of foundation pit; Anthony [27] proposed a simplified method to evaluate the basic uplift coefficient for the safety of axially symmetrically supported foundation pit in clay. The research on deep foundation pit abroad has gone through more than 200 years, and the research results are continuous.

1.2. Research Status of LSTM. The standard feedforward neural network does not have the ability to analyze and predict time series data. It is only used to process the current time series data, and its historical information cannot be used for subsequent processing. The emergence of recurrent neural network (RNN) [28, 29] just makes up for this defect. The framework of RNN was proposed in the 1980s, and the biggest difference between RNN and feedforward neural network is that the hidden layer of RNN is connected in chronological order; that is, sequence features can be stored through the hidden layer and applied to later sequence output operations. Although RNN can store and use historical data features, the information in any time length cannot be used, because with the increase of time series scale, RNN will have problems such as gradient disappearance in the training process [30], which slows down the training speed, makes it difficult for the network to converge, and cannot obtain the optimal solution, which limits the popularization and application of RNN. To solve these problems, scholars have also put forward some solutions, such as adding simulated annealing method in RNN training [31] or using hierarchical compression method for time series data [32]. These methods did not effectively deal with the disappearance of RNN gradient until LSTM network appeared. LSTM network was proposed by Hochreiter and Schmidhuber in 1997. Experiments show that LSTM network can effectively avoid the disappearance of gradient, which really solves the problems of RNN.

To sum up, the prediction accuracy of models established based on LSTM network is better than that of traditional prediction methods. After continuous application and network structure adjustment, the models are widely used in many fields and many achievements have been obtained. However, LSTM network is seldom used in surveying and mapping related fields, and there is almost no research on data processing and prediction of deep foundation pit deformation. Therefore, LSTM network is introduced into the field of deep foundation pit deformation prediction in view of its efficient processing ability and good generalization ability in various nonlinear time series data, so as to process and analyze highly nonlinear time series data of deep foundation pit deformation. However, previous experience can prove that the modeling process of LSTM network involves the selection of some parameters, which will affect the training effect of the network, and LSTM network has certain requirements on timing data. Therefore, the parameter selection and model structure of LSTM network in deep foundation pit deformation prediction are emphatically analyzed in this paper, which adds a new method for deep foundation pit deformation prediction and helps to expand the application of deep learning in the field of deformation prediction.

2. LSTM Network Equation

The training process of LSTM network is the same as that of RNN, including forward transmission and error reverse transmission. The LSTM network error correction process applies the back propagation through time (BPTT) [33] algorithm, which is based on results of calculating the loss function of output layer and the error over time and in accordance with

the gradient descent method reverse calibration weights of each neuron, and the execution process is more complex than RNN.

The operation of each LSTM memory module is the same. This time, only one memory module is used as an example. Before deriving the LSTM network equation, each variable of the network is defined first. W_{ij} is the weight of the connection between neuron i and neuron j , the input result of a certain neuron j at time t is denoted as a_j^t , and the activation function of the neuron at this time is denoted as b_j^t . Input gate, forget gate, and output gate are represented with subscript l , ϕ , and ψ , and memory unit is equipped with subscript c . s_c^t represents the state value of memory unit C at time t . The activation function of gate structure is denoted as f , and g and h are successively represented as input and output activation functions of neurons.

I represents the number of input nodes, K represents the number of output nodes, and H represents the number of nodes in the hidden layer. It should be noted that only the hidden layer unit output b_c^t can be connected with other memory modules, while the remaining activation functions in the LSTM network, such as cell state, network node input, and multiplication gate activation function, can only be effective in the LSTM memory module. For the LSTM network standard hiding unit, the index h is used to refer to the unit output of other memory structures in the hiding layer (ordinary neurons can be mixed with the LSTM storage structure in the same hiding layer if necessary). Similar to standard RNN, the forward transfer process of LSTM network is to calculate the input sequence with time step T . From the initial moment $t = 1$ of this time period, the time gradually increases, and the operation equation is constantly updated until $t = T$, and the corresponding output results are obtained. The back propagation of LSTM network starts from $t = T$, and BPTT algorithm is used to calculate the neuron derivative recursively. Meanwhile, the time decreases to t . The final weight derivative can be obtained by summing the derivatives on each time step, as shown in

$$\delta_j^t \triangleq \frac{\partial O}{\partial a_j^t}. \quad (1)$$

In the formula, O is the loss function in the LSTM network training process. The loss function is used to measure the loss of the entire network during training, and the network training is judged by the result of the loss function.

The order of the forward and backward calculation formulas for the LSTM network is important and must be followed as follows. As with standard RNN, the state and activation function equivalents of all neurons at time $t = 0$ are set to 0, and the derivative δ of some weights at time $t = T + 1$ is also 0.

2.1. Projecting Forward. Input gate:

$$a_i^t = \sum_{i=1}^I w_{ii}x_i^t + \sum_{h=1}^H w_{hi}b_h^{t-1} + \sum_{c=1}^C w_{ci}s_c^{t-1}, \quad (2)$$

$$b_i^t = f(a_i^t).$$

Forget gate:

$$a_\phi^t = \sum_{i=1}^I w_{i\phi} x_i^t + \sum_{h=1}^H w_{h\phi} b_h^{t-1} + \sum_{c=1}^C w_{c\phi} s_c^{t-1}, \quad (3)$$

$$b_\phi^t = f(a_\phi^t).$$

Memory cell:

$$a_c^t = \sum_{i=1}^I w_{ic} x_i^t + \sum_{h=1}^H w_{hc} b_h^{t-1}, \quad (4)$$

$$s_c^t = b_\phi^t s_c^{t-1} + b_i^t g(a_c^t).$$

Output gate:

$$a_\psi^t = \sum_{i=1}^I w_{i\psi} x_i^t + \sum_{h=1}^H w_{h\psi} b_h^{t-1} + \sum_{c=1}^C w_{c\psi} s_c^{t-1}, \quad (5)$$

$$b_\psi^t = f(a_\psi^t).$$

Unit output:

$$b_c^t = b_\psi^t h(s_c^t). \quad (6)$$

2.2. Reverse Calculation. Suppose

$$\varepsilon_c^t \triangleq \frac{\partial O}{\partial b_c^t} \triangleq \frac{\partial O}{\partial s_c^t}. \quad (7)$$

Unit output:

$$\varepsilon_c^t = \sum_{k=1}^K w_{ck} \delta_k^t + \sum_{h=1}^H w_{ch} \delta_h^{t+1}. \quad (8)$$

Output gate:

$$\varepsilon_\psi^t = f'(a_\psi^t) \sum_{c=1}^C h(s_c^t) \varepsilon_c^t. \quad (9)$$

Current status:

$$\varepsilon_s^t = b_\psi^t h'(s_c^t) \varepsilon_c^t + b_\phi^{t+1} \varepsilon_s^{t+1} + w_{cl} \delta_l^{t+1} + w_{c\phi} \delta_\phi^{t+1} + w_{c\psi} \delta_w^t. \quad (10)$$

Memory cell:

$$\delta_c^t = b_i^t g'(a_c^t) \varepsilon_c^t. \quad (11)$$

Forget gate:

$$\delta_\phi^t = f'(a_\phi^t) \sum_{c=1}^C s_c^{t-1} \varepsilon_s^t. \quad (12)$$

Input gate:

$$\varepsilon_i^t = f'(a_i^t) \sum_{c=1}^C g(a_c^t) \varepsilon_s^t. \quad (13)$$

This chapter first introduces the related theories of RNN, the advantages and disadvantages of RNN, and how LSTM network is improved from RNN, then introduces the structure of LSTM network in detail, and finally deduces the training equation of LSTM network in detail.

3. The Construction of Deep Foundation Pit Deformation Prediction Model Based on LSTM

3.1. LSTM Network Modeling Principle and Process. The prediction process of deep foundation pit deformation is a dynamic process, and it is necessary to update the measured data constantly to forecast the deformation of deep foundation pit in real time. The advantage of LSTM network training is that the model can be saved after the LSTM network prediction model training. When the model is used again, the newly added data will not affect the overall training of the model, which can effectively reduce the training time of the model and then dynamically predict the deformation of deep foundation pit. Deep foundation pit excavation is a huge systematic engineering; its deformation changes with time, external conditions, and other factors. The monitoring value of deep foundation pit is a series of time-dependent data and is affected by external environment and other factors. The single monitoring value is fuzzy and random, but the whole time series data has certain regularity. Through the analysis of time series data, the deformation trend and certain regularity of deep foundation pit can be found. LSTM network is mainly used for modeling analysis and prediction of time series data, and LSTM network can build multi-to-multiple model structure with time difference according to actual requirements, which meets the basic requirements of deep foundation pit deformation prediction. Usually, time series data obtained from early monitoring are used to construct network training samples to predict the deformation in the middle and late period.

The key of LSTM network lies in the choice of memory module. Generally speaking, a memory unit corresponds to a hidden layer. The input sequence data needs to be segmented, and the data in each sequence needs to correspond to the memory unit.

Firstly, the preprocessed data is transferred from the network input layer, and the segmented timing data $[X_1, X_2, \dots, X_n]$ (n is the sequence length, which is determined by the segmentation scale) are input to the LSTM memory units at different times in sequence. After the LSTM memory unit

(hidden layer) transfers and extracts information layer by layer, it is transmitted to the network output layer. The output layer is added with full connected layers (FC) and dropout [34] (introduced in Chapter 4), which are, respectively, used to model regression and prevent over fitting of training. Finally, the timing data $[Y_1, Y_2, \dots, Y_n]$ is output.

The prediction process of deep foundation pit deformation based on LSTM can be seen introduction that the LSTM network model divides the prediction process of deep foundation pit deformation into three parts: acquisition and data preprocessing of deep foundation pit monitoring data, LSTM network model training, and directly obtaining the predicted value by using the trained model parameters. The main process is as follows.

3.1.1. Data Acquisition and Preprocessing. In this paper, the deformation monitoring data of a subway deep foundation pit in Wuhan is used, which is the data observed by relevant instruments during the construction period of the deep foundation pit. The data acquisition frequency varies according to the deformation situation. In the early stage of deep foundation pit excavation, the monitoring frequency was once every 3 days. As the depth of deep foundation pit deepens, the monitoring frequency began to adjust to once every 2 days. As the structure of deep foundation pit supporting and other structures changed greatly due to soil unloading, the monitoring frequency was set to once every day. The obtained monitoring data cannot be directly used for prediction, so it is necessary to conduct outlier test and eliminate (the number of monitoring samples is set as n , and PauTa rule should be selected). Chauve net criterion should be used. Grubbs rule is suitable. LSTM network is used to predict peer-to-peer time series data, so data after outlier processing should be interpolated. Interpolation methods include linear interpolation, Newton interpolation, and cubic spline interpolation. The monitoring data of deep foundation pit columns are highly nonlinear due to the influence of many uncertain factors, so linear interpolation method is not suitable for them. Cubic spline interpolation is a relatively simple interpolation method that can well reflect the nonlinear characteristics of deep foundation pit deformation monitoring data. Cubic spline interpolation has good smooth effect and fast implementation and is a common method of monitoring data interpolation at present. The interpolated data is constructed as time series data and divided into training set and test set.

3.1.2. LSTM Network Training. Like other neural networks, LSTM network also needs to normalize the training set, and the normalized data is conducive to the rapid convergence of the network. LSTM network training mostly uses a large number of data for training, but the monitoring data of deep foundation pit used in this paper are about 100 periods, and the training set data is limited. Therefore, LSTM network carries out network training in the way of iterative training, so as to obtain the global optimal solution. Before LSTM network training, all parameters need to be initialized. After each iterative training, the training error can be obtained through the loss function. When the error

meets the requirements, it can be predicted. If the tolerance is not satisfied, the LSTM network will automatically adopt BPTT algorithm to back propagate the error. The optimization algorithm is adopted (see the next section) to continuously update the parameters such as the weights of each layer of the network and finally debug the super parameters such as the number of iterative steps and global learning rate according to the loss function value. The network continuously iterates and trains to obtain the most loss value and obtain the optimal network parameter solution.

3.1.3. Network Prediction. When the LSTM network training error meets the requirements and the number of iterations is completed, in which the network parameters are obtained, it can be directly used to calculate the predicted value, and the real predicted value can be obtained after the obtained prediction results are calculated by the inverse normalization formula. Then, compare with the test set to test the prediction accuracy.

3.2. Selection of Optimization Algorithm. LSTM network is a deep neural network model, which can be fitted by single or multiple complex time functions. The independent variables of composite time function in LSTM network are the weights and bias items of each layer in the network, which directly affect the accuracy of the final output results of the whole LSTM network. In order to improve the effectiveness of LSTM network training, it is necessary to constantly update and optimize weight and bias parameters. The optimization algorithm starts to play its role at this moment, which directly affects the final training effect of LSTM network. Therefore, the selection of LSTM network optimization algorithm must be careful. Different optimization algorithms have different effects in LSTM network modeling. At the present stage, deep learning model mainly uses frequently optimization algorithms: stochastic gradient descent (SGD) algorithm, momentum algorithm, Nesterov algorithm, AdaGrad algorithm, RMSProp algorithm, and Adam algorithm [35], and to choose the suitable LSTM optimization algorithm of network model needs to be determined through examples. The principles of each algorithm will be introduced in detail below.

3.2.1. SGD Algorithm. SGD algorithm is the simplest and most commonly used optimization algorithm in deep learning. It randomly selects l small batch samples from all training samples n , which are independent of each other, and then takes the average gradient value of these small batch samples as the next gradient updating direction. The specific algorithm is as Algorithm 1 as follows.

3.2.2. Momentum Algorithm. Momentum algorithm can effectively reduce network training oscillation and make network convergence easier. The specific algorithm is as Algorithm 2 as follows.

v_i in the algorithm will accumulate the previous training gradient. Compared with the learning rate η , the larger the momentum coefficient α is, it indicates that the current network training iteration direction is more influenced by the previous gradient. The previous iteration point of SGD

Require: set the initial learning rate to η
 Require: set the initial parameter to θ
 While stop condition not satisfied do
 Randomly extract l samples $x^i = \{x^1, x^2, \dots, x^l\}$ from the training set, and y^i is the true value corresponding to x^i .
 Calculate the average gradient of l samples: $g_t = (1/l)\nabla_{\theta_{t-1}}\sum_i L(f(x^i; \theta_{t-1}), y^i)$
 Parameter updating: $\theta_t \leftarrow \theta_{t-1} - \eta g_t$
 End while

ALGORITHM 1: SGD algorithm.

Require: set the initial learning rate to η , set to momentum coefficient α
 Require: set the initial parameter to θ , set initial speed v to 0
 While stop condition not satisfied do
 Randomly extract l samples $x^i = \{x^1, x^2, \dots, x^l\}$ from the training set, and y^i is the true value corresponding to x^i .
 Calculate the average gradient of l samples: $g_t \leftarrow (1/l)\nabla_{\theta_{t-1}}\sum_i L(f(x^i; \theta_{t-1}), y^i)$
 Step size update: $v_t \leftarrow \alpha v_{t-1} - \eta g_t$
 Parameter updating: $\theta_t \leftarrow \theta_{t-1} - v_t$
 End while

ALGORITHM 2: Momentum algorithm.

algorithm is used as the gradient update orientation, while the momentum algorithm takes the sum of all previous gradients and current gradients after weight calculation as the current network training iteration orientation. The value of α represents the total proportion of all historical iterations, usually 0.9 or 0.99.

3.2.3. Nesterov Acceleration Gradient Algorithm. Nesterov algorithm first predicts the orientation of the iteration point of the next parameter θ and then obtains the gradient at the predicted point. Finally, weight calculation and sum of the current gradient and all previous gradients are carried out to obtain the updated orientation of the gradient at the next iteration point, as Algorithm 3 as follows.

3.2.4. AdaGrad Algorithm. The learning rate of the three optimization algorithms introduced above is constant, while the learning rate of the following optimization algorithms changes according to certain rules in network training, that is, the algorithm will adjust the learning rate according to certain rules in the whole network training stage. AdaGrad algorithm first gives an initial learning rate, and then uses the ratio of the learning rate to the square of all previous gradients as the learning rate of network training at this moment, as Algorithm 4 as follows.

3.2.5. RMSProp Algorithm. Similarly, given the initial learning rate firstly, RMSProp algorithm considers that the farther the iteration point from the current iteration point in the historical gradient has less impact on the current, so set an exponential attenuation rate value (the default value is 0.9), calculate the attenuation average value of the historical gradient, and then divide the initial learning rate by this value to the learning rate of the current iteration point, as Algorithm 5 as follows.

3.2.6. Adam Algorithm. As Algorithm 6, Adam algorithm combines the advantages of RMSProp and norm, which has the characteristics of easy implementation, high operation efficiency, and low storage requirements. Adam algorithm only requires to obtain one step of the loss function in the training process of deep learning model. If the parameters are inconsistent, the learning rate will also change. The learning rate is usually determined by Adam algorithm according to the first-order and second-order moment estimation of gradient.

These six algorithms involve the initialization process of initial parameter θ . Generally, in order to avoid a series of numerical problems caused by 0 in mathematical operation, it is not suitable to set a to 0. At present, the commonly used method is to give a very small random number, which will not affect the whole process of algorithm optimization LSTM network training but also solve the numerical problem caused by 0.

These six optimization algorithms are widely used and skilled algorithms in deep learning model and have their advantages and disadvantages in some aspects. Therefore, the depth learning model applicable to each optimization algorithm will be different, and the applicable data types or scenarios are also different. It is necessary to select the appropriate optimization algorithm in practical application. In order to select the LSTM network prediction model suitable for deep foundation pit deformation prediction, the above six optimization algorithms are used for LSTM network modeling, respectively, and the best selection of LSTM network optimization algorithm is determined through specific engineering examples.

3.2.7. Sequence Segmentation Scale. Before timing data is input to the LSTM network input layer, data segmentation is often required to correspond to LSTM memory units. If

Require: set the initial learning rate to η , set to momentum coefficient α (0.9 or 0.99)
 Require: set the initial parameter to θ , set initial speed v to 0
 While stop condition not satisfied do
 Randomly extract l samples $x^i = \{x^1, x^2, \dots, x^l\}$ from the training set, and y^i is the true value corresponding to x^i .
 Forecast point update: $\tilde{\theta}_t \leftarrow \theta_{t-1} + \alpha v_{t-1}$
 Calculate the gradient at the predicted point: $g_t \leftarrow (1/l) \nabla_{\tilde{\theta}_t} \sum_i L(f(x^i; \tilde{\theta}_t), y^i)$
 Speed update: $v_t \leftarrow \alpha v_{t-1} - \eta g_t$
 Parameter updating: $\theta_t \leftarrow \theta_{t-1} + v_t$
 End while

ALGORITHM 3: Nesterov algorithm.

Require: set the initial learning rate to η , and the value of δ of the initial constant maturity is 10^{-7} , the initial parameter is set to θ
 Require: set the initial gradient accumulative value v to 0
 While stop condition not satisfied do
 Randomly extract l samples $x^i = \{x^1, x^2, \dots, x^l\}$ from the training set, and y^i is the true value corresponding to x^i .
 Forecast point update: $\tilde{\theta}_t \leftarrow \theta_{t-1} + \alpha v_{t-1}$
 Calculate the average gradient of l samples: $g_t \leftarrow (1/l) \nabla_{\tilde{\theta}_t} \sum_i L(f(x^i; \tilde{\theta}_t), y^i)$
 Square of accumulated value of historical gradient: $v_t \leftarrow \alpha v_{t-1} + g_t^2$
 Update gradient value: $\Delta \theta_t \leftarrow -(\eta / (\delta + \sqrt{v_t})) \odot g_t$ (element operation)
 Parameter updating: $\theta_t \leftarrow \theta_{t-1} + \Delta \theta_t$
 End while

ALGORITHM 4: AdaGrad algorithm.

Require: set the initial learning rate to η , the decay rate is set to ρ , and the initial parameter is set to θ
 Require: set the initial gradient accumulative value v to 0, and the value of the initial constant δ is 10^{-8}
 While stop condition not satisfied do
 Randomly extract l samples $x^i = \{x^1, x^2, \dots, x^l\}$ from the training set, and y^i is the true value corresponding to x^i .
 Calculate the average gradient of l samples: $g_t \leftarrow (1/l) \nabla_{\tilde{\theta}_t} \sum_i L(f(x^i; \tilde{\theta}_t), y^i)$
 Square of accumulated value of historical gradient: $v_t \leftarrow \alpha v_{t-1} + (1 - \alpha) g_t \odot g_t$
 Update gradient value: $\Delta \theta_t \leftarrow -(\eta / (\sqrt{v_t} + \delta)) \odot g_t$ (element operation)
 Parameter updating: $\theta_t \leftarrow \theta_{t-1} + \Delta \theta_t$
 End while

ALGORITHM 5: RMSProp algorithm.

the sample data changes according to a certain period, the period is used as the scale of sequence segmentation. If no change rule of time series data can be seen, a value should be set in advance, and then, the sequence segmentation scale should be manually increased or decreased as required. The advantage of LSTM network compared with RNN is that LSTM network can capture the temporal connection of input data no matter how big the time step is. However, in the prediction of deep foundation pit deformation, the samples that can be used for training are limited, and the segmentation scale should not be set too large. Therefore, the sequence segmentation scale of LSTM network in this paper is controlled within 15.

3.2.8. Loss Function. The loss function is usually used to calculate the difference between the model output $f(x)$ and the input data Y , often represented by $L(Y, f(x))$. The smaller

the loss function, the better the robustness of the model. At present, loss functions commonly used in deep learning models include mean square error loss function, cross entropy loss function, exponential loss function, and absolute value loss function. In this paper, LSTM network is applied to the prediction of deep foundation pit deformation. In order to see the training effect more intuitively, the mean square error loss function is adopted as the loss function of LSTM network training.

$$L(Y, f(x)) = \frac{1}{2} \|Y - f(x)\|_2^2. \quad (14)$$

3.2.9. Activation Function. The function of activation function is to retain the characteristic information in the data and remove the redundant information. Nonlinear function

Require: set the initial learning rate to η , and the exponential decay rate of moment estimation of the first and second orders is set to β_1 and β_2 in turn, and the value range of β_1 and β_2 is in the interval $[0, 1)$.
 Require: set the initial parameter to θ , and initialized first and second orders are denoted as $m = 0$ and $v = 0$, respectively, initialize time step $t = 0$
 While stop condition not satisfied do
 Randomly extract l samples $x^i = \{x^1, x^2, \dots, x^l\}$ from the training set, and y^i is the true value corresponding to x^i .
 Calculate the average gradient of l samples: $g_t \leftarrow (1/l) \nabla_{\theta_{t-1}} \sum_i L(f(x^i; \theta_{t-1}), y^i)$
 Biased first-order moment estimation update: $m_t \leftarrow \beta_1 \bullet m_{t-1} + (1 - \beta_1) \bullet g_t$
 Biased second-order moment estimation update: $v_t \leftarrow \beta_2 \bullet v_{t-1} + (1 - \beta_2) \bullet g_t \odot g_t$
 First-order moment error correction: $\widehat{m}_t \leftarrow -m_t / (1 - \beta_1^t)$
 Second-order moment error correction: $\widehat{v}_t \leftarrow -v_t / (1 - \beta_2^t)$
 Update gradient value: $\Delta\theta_t \leftarrow -\eta \widehat{m}_t / (\sqrt{\widehat{v}_t + \delta})$ (element operation)
 Parameter updating: $\theta_t \leftarrow \theta_{t-1} + \Delta\theta_t$
 End while

ALGORITHM 6: Adam algorithm.

is usually used, and the key role of adding nonlinear function is to increase the nonlinear factor of neural network. It is the repeated superposition of nonlinear functions of each layer of network that makes the neural network have enough ability to extract various complex nonlinear features and approximate various functional relationships. The tanh or sigmoid function is often used as the activation function for the operation between memory units in the LSTM network, while the ReLU function is selected for the activation function between each layer of the network according to the previous experience. The ReLU activation function was proposed by Hinton et al. in 2010 and was originally used in the restricted Boltzmann machine (RBM). ReLU is essentially a function taking the maximum value, and its formula is as follows:

$$f(x) = \begin{cases} x_i, & x_i \geq 0, \\ 0, & x_i < 0. \end{cases} \quad (15)$$

ReLU has the following advantages compared with other activation functions:

- (1) Only need to judge whether the output is greater than zero, so the operation speed is fast
- (2) It is an unsaturated activation function, which will not cause the disappearance of model gradient like sigmoid and tanh saturation functions. ReLU function can converge rapidly during model training

3.3. Project Example. In order to compare the influence of different optimization functions on the output accuracy of LSTM network, six commonly used optimization algorithms are used to model and analyze LSTM network. As the external parameters are not the critical factors to determine the accuracy of the final output results of the LSTM network model, the external parameters of each model are adjusted to the optimal solution state. In order to better evaluate the prediction accuracy of each model, the average prediction accuracy (Formula (16)), the average relative error (Formula

(17)), and the total time spent on model training and prediction were adopted as the accuracy evaluation indexes.

$$\sigma_p^2 = \frac{\sum_{i=1}^n (x_i - \widehat{x}_i)^2}{n}, \quad (16)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{V_i}{x_i} \right|. \quad (17)$$

In the above formula, σ_p represents the average prediction accuracy, v_i^2 represents the sum of the squares of residuals, x_i and \widehat{x}_i represent the measured value and predicted value, respectively, and n represents the total number of prediction periods.

3.3.1. Application in Horizontal Displacement of Deep Foundation Pit Pile. The data source is the monitoring value of horizontal displacement of retaining pile ZQT-14 at 21 meters in a subway deep foundation pit in Wuhan (denoted as ZQT-14-21). In deep foundation pit, the retaining pile is not only used to protect the surrounding structures and pipelines from damage but also used to maintain the stability of the adjacent soil, so as to ensure the safety of deep foundation pit engineering. In the process of deep foundation pit construction, the retaining pile is not only affected by the force of soil inside and outside the foundation pit but also affected by many factors such as the supporting structure and traffic flow around it, so the monitoring data of ZQT-14-21 change frequently and the deformation law is uncertain. The monitoring data of ZQT-14-21 were processed according to the pretreatment method, and 118 monitoring data were obtained.

The first 98 periods of monitoring data of ZQT-14-21 were used for LSTM network training, and the deformation data of the next 20 periods were predicted by rolling and then compared with the measured values. The prediction results obtained by the six optimization algorithms are shown in Table 1 (the data with odd observation periods are listed in the table), and the average prediction accuracy is shown in Table 2.

As can be seen from the calculation results in Tables 1 and 2, the prediction results and prediction precision obtained by each model are different. From the perspective of average prediction accuracy and average relative error, Adam algorithm has the best effect among the six optimization algorithms. In terms of time consumption, Adam algorithm consumes not the least time in the whole training and prediction process of LSTM network, but at a medium level. This is because Adam algorithm needs to calculate more parameters than other algorithms and theoretically consumes more time. On the whole, the LSTM network prediction model optimized by Adam algorithm has the best accuracy, which is more consistent with the deep level displacement and deformation of deep foundation pit pile. It shows that Adam algorithm has the greatest influence on the accuracy of LSTM network output and the best optimization effect and has better applicability than other optimization algorithms. In general, Adam algorithm optimized LSTM network training and prediction effect is the best among the six algorithms.

3.3.2. Application in Settlement of Column of Deep Foundation Pit. The data source is the vertical displacement variation of column settlement monitoring point LZC-05 of a subway deep foundation pit in Wuhan, which has been observed for 95 periods. Firstly, the original column settlement monitoring data were tested and eliminated by outlier test. The total monitoring data of LZC-05 was less than 100 periods, and the test method was Grabus' rule. LZC-05 does not have outliers after inspection, but the monitoring frequency of column settlement measurement points in the first 23 days is once every two days, so interpolation is needed. The cubic spline method is used to obtain 23 periods of data, and a total of 106 periods of data are obtained before and after interpolation.

Similarly, the data of phase 86 were used for the training of each model, and the data of the last 20 phases were used for comparison with the predicted values. The prediction results of each model are shown in Table 3 (odd number of phases was selected), and the calculation results of average prediction accuracy are shown in Table 4.

Combined with Tables 3 and 4, it can be seen that LSTM network prediction accuracy and time required by different optimization algorithms have a large gap. From the perspective of prediction accuracy, Adam algorithm has the best effect, and it has the highest prediction accuracy when applied to LSTM network modeling, while SGD algorithm has the lowest accuracy. In terms of consumption time, AdaGrad algorithm consumes the most time and the network convergence is the slowest, while RMSProp algorithm updates the gradient fastest, so it consumes the least time. On the whole, the LSTM network optimized by Adam algorithm consumes not the least time in training and prediction, but it is not different from other algorithms, and its accuracy is the best among all optimization algorithms. The prediction of deep foundation pit deformation requires a high-precision prediction model, so Adam algorithm is the most suitable LSTM network trainer.

3.3.3. Analysis and Summary. Because LSTM network has excellent feature extraction, long- and short-term memory ability, and strong nonlinear fitting and generalization ability for time series data, it is applied to time series data processing of deep foundation pit deformation, and good prediction effect is obtained. In the same example, RMSProp, AdaGmd, Adam, momentum, Nesterov, and SGD algorithms are selected to optimize the LSTM network prediction model, and the prediction results of each model are compared and analyzed. Through two examples of horizontal displacement of pile and vertical displacement of column in deep foundation pit, it is found that there are nonlinear and time sequence characteristics in the data of the two examples, and the LSTM network prediction model constructed is relatively accurate, which confirms the feasibility of LSTM network. The LSTM network prediction results constructed by these six optimization algorithms are quite different, among which the LSTM network model constructed by Adam optimization algorithm has the best accuracy, and the consumption time is not much different from other algorithms, indicating that the selection of optimization algorithm plays a crucial role in LSTM network prediction results. Therefore, on the whole, LSTM network prediction model constructed by Adam optimization algorithm is the most suitable for deep foundation pit deformation prediction. Firstly, the modeling method and process of LSTM network are discussed in detail, then the optimization algorithm used in the model is described in detail, and the parameter selection methods such as initial learning rate, activation function, and iteration number related to LSTM network training are introduced in detail. Then, taking the deformation data of the monitoring body of deep foundation pit as an example, the influence of different optimization algorithms on the prediction effect of LSTM network is compared, which proves that the selection of optimization algorithm plays an important role in LSTM and also verifies the feasibility of LSTM network in the data processing and prediction of deep foundation pit deformation.

4. Conclusions

LSTM network is used to process the deformation data of deep foundation pit, and random gradient descent, momentum, Nesterov, RMSProp, AdaGmd, and Adam algorithms are selected in the same example for modeling prediction and comparison. Two examples of horizontal displacement prediction of pile and vertical displacement prediction of column in deep foundation pit show that the LSTM network model established by different optimization algorithms has different prediction accuracy, and the LSTM network model established by Adam optimization algorithm has the highest accuracy, indicating that the selection of optimization algorithm is particularly important in LSTM network modeling.

Combining LSTM network and its improved model on the part of the monitoring data of deep foundation pit deformation analysis and prediction obtained some achievements, but there are some contents that need to be perfect, as follows:

TABLE 1: Prediction results of pile horizontal displacement measurement point ZQT-14-21.

Numbers of period (s)	Measured value (mm)	Optimization algorithm					
		Adam (mm)	Momentum (mm)	SGD (mm)	RMSProp (mm)	AdaGrad (mm)	Nesterov (mm)
1	-0.44	-0.42	-0.39	-0.50	-0.45	-0.40	-0.42
3	-0.68	-0.62	-0.61	-0.70	-0.58	-0.61	-0.65
5	-0.92	-0.98	-0.81	-0.78	-0.98	-0.82	-0.84
7	-0.88	-0.79	-0.78	-0.75	-0.87	-0.79	-0.99
9	-0.66	-0.61	-0.59	-0.58	-0.56	-0.59	-0.73
11	-0.95	-1.00	-0.84	-0.80	-1.04	-0.85	-0.86
13	-0.88	-0.90	-0.78	-0.75	-0.82	-0.79	-0.81
15	-0.75	-0.78	-0.67	-0.65	-0.85	-0.67	-0.70
17	-0.68	-0.61	-0.61	-0.60	-0.72	-0.61	-0.65
19	0.12	0.14	0.10	0.07	0.07	0.09	0.08

TABLE 2: Calculation results of accuracy evaluation indexes of each ZQT-14-21 model.

Indicator	Optimization algorithm					
	Adam	Momentum	SGD	RMSProp	AdaGrad	Nesterov
Average prediction accuracy (mm)	0.05	0.08	0.11	0.07	0.08	0.07
Average relative error (%)	7.31	11.69	15.45	11.39	11.32	11.49
Time consumed (seconds)	13.40	12.30	15.20	11.70	35.90	12.00

TABLE 3: Prediction results of column vertical displacement measurement point LZC-05.

Numbers of period (s)	Measured value (mm)	Optimization algorithm					
		Adam (mm)	Momentum (mm)	SGD (mm)	RMSProp (mm)	AdaGrad (mm)	Nesterov (mm)
1	18.40	17.79	17.57	17.42	17.53	17.49	17.41
3	19.40	18.72	20.51	18.25	18.41	18.45	18.26
5	19.80	19.08	18.88	20.58	20.75	20.63	20.60
7	19.90	20.37	18.98	20.66	19.03	19.23	19.38
9	18.90	18.26	19.94	18.03	17.97	17.97	17.83
11	18.10	17.52	17.28	19.16	19.27	17.20	17.15
13	19.00	18.35	19.73	19.92	18.06	18.06	18.09
15	18.90	19.56	18.24	18.33	17.89	20.10	18.23
17	22.00	23.06	20.91	20.32	20.58	20.96	20.42
19	22.10	23.14	21.00	20.40	20.66	21.05	20.50

TABLE 4: Calculation results of accuracy evaluation indexes of LZC-05 models.

Indicator	Optimization algorithm					
	Adam	Momentum	SGD	RMSProp	AdaGrad	Nesterov
Average prediction accuracy (mm)	0.75	0.95	1.16	1.08	0.94	1.10
Average relative error (%)	3.60	4.70	5.21	5.31	4.71	5.19
Time consumed (seconds)	15.60	15.40	18.50	11.90	41.00	15.00

(1) Due to the fact that technology and ability are limited, only the LSTM network optimization algorithm is presented in this paper, method of super parameter selection and made a research on the multipoint

prediction, etc. Future research focus can be considered on LSTM network training algorithm to further improve LSTM network training and prediction performance

- (2) As the multipoint prediction model will reduce the training speed with the increase of the number of monitoring points, GPU acceleration can be considered in the future to improve the running speed of the model

Data Availability

The supporting data are available from the corresponding author, and the authors declare that they have no conflicts of interest regarding this paper.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] P. Xu, Y. Han, H. Duan, and S. Fang, "Environmental effects induced by deep subway foundation pit excavation in Yellow River alluvial landforms," *Geotechnical and Geological Engineering*, vol. 33, no. 6, pp. 1587–1594, 2015.
- [2] Z. Bao and C. Wang, "A multi-agent knowledge integration process for enterprise management innovation from the perspective of neural network," *Information Processing & Management*, vol. 59, no. 2, p. 102873, 2022.
- [3] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [4] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [5] I. Sutskever, O. Vinyals, and V. L. Quoc, "Sequence to sequence learning with neural networks," *Advances in neural information processing systems*, vol. 27, 2014.
- [6] H. Palangi, L. Deng, Y. Shen et al., "Deep sentence embedding using long short-term memory networks: analysis and application to information retrieval," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 4, pp. 694–, 2016.
- [7] F. Visin, K. Kastner, K. Cho, M. Matteucci, A. Courville, and Y. Bengio, "Renet: a recurrent neural network based alternative to convolutional networks," 2015, <https://arxiv.org/abs/1505.00393>.
- [8] S. Yousfi, S. A. Berrani, and C. Garcia, "Contribution of recurrent connectionist language models in improving LSTM-based Arabic text recognition in videos," *Pattern Recognition*, vol. 64, pp. 245–254, 2017.
- [9] D. Monner and J. A. Reggia, "Neural architectures for learning to answer questionsBiologically Inspired Cognitive Architectures," vol. 2, pp. 37–53, 2012.
- [10] A. Graves, N. Jaitly, and A. R. Mohamed, "Hybrid speech recognition with deep bidirectional LSTM," in *2013 IEEE workshop on automatic speech recognition and understanding*, pp. 273–278, Olomouc, Czech Republic, December 2013.
- [11] S. Xu and R. Niu, "Displacement prediction of Baijiabao landslide based on empirical mode decomposition and long short-term memory neural network in Three Gorges area, China," *Computers & Geosciences*, vol. 111, pp. 87–96, 2018.
- [12] D. Zhang, E. S. Hølland, G. Lindholm, and H. Ratnaweera, "Hydraulic modeling and deep learning based flow forecasting for optimizing inter catchment wastewater transfer," *Journal of Hydrology*, vol. 567, pp. 792–802, 2018.
- [13] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," *Transportation Research Part C*, vol. 54, pp. 187–197, 2015.
- [14] D. R. P. Hettiaratchi, "Theoretical soil mechanics and implement design," *Soil and Tillage Research*, vol. 11, no. 3-4, pp. 325–347, 1988.
- [15] G. W. Clough and L. A. Hansen, "Clay anisotropy and braced wall behavior," *Journal of the Geotechnical Engineering Division*, vol. 107, no. 7, pp. 893–913, 1981.
- [16] C. Y. Ou, J. T. Liao, and W. L. Cheng, "Building response and ground movements induced by a deep excavation," *Geotechnique*, vol. 50, no. 3, pp. 209–220, 2000.
- [17] R. J. Finno and L. S. Bryson, "Response of building adjacent to stiff excavation support system in soft clay," *Journal of Performance of Constructed Facilities*, vol. 16, no. 1, pp. 10–20, 2002.
- [18] R. B. Peck, "Deep excavations and tunneling in soft ground," in *7th International Conference on Soil Mechanics and Foundation Engineering*, vol. 1, pp. 225–290, Mexico City, Mexico, State-of-the-Art, 1969.
- [19] L. Bjerrum and O. Eide, "Stability of strutted excavations in clay," *Geotechnique*, vol. 6, no. 1, pp. 32–47, 1956.
- [20] G. W. Clough, "Construction induced movements of in situ walls," *Design & Performance of Earth Retaining Structures*, 1990.
- [21] G. W. Clough and M. W. Reed, "Measured behavior of braced wall in very soft clay," *Journal of Geotechnical Engineering*, vol. 110, no. 1, pp. 1–19, 1984.
- [22] A. T. C. Goh, K. S. Wong, and B. B. Broms, "Estimation of lateral wall movements in braced excavations using neural networks," *Canadian Geotechnical Journal*, vol. 32, no. 6, pp. 1059–1064, 1995.
- [23] F. Mayoraz and L. Vulliet, "Neural networks for slope movement prediction," *International Journal of Geomechanics*, vol. 2, no. 2, pp. 153–173, 2002.
- [24] C. Moormann, "Analysis of wall and ground movements due to deep excavations in soft soil based on a new worldwide database," *Soils and foundations*, vol. 44, no. 1, pp. 87–98, 2004.
- [25] H. M. Liu, J. Cao, and W. F. Zhang, "Study on effects of ground deformation based on foundation pit excavation," *Advanced Materials Research*, vol. 671-674, no. 1, pp. 357–362, 2013.
- [26] M. Sepehri, D. B. Apel, and R. A. Hall, "Prediction of mining-induced surface subsidence and ground movements at a Canadian diamond mine using an elastoplastic finite element model," *International Journal of Rock Mechanics and Mining Sciences*, vol. 100, pp. 73–82, 2017.
- [27] A. T. Goh, "Basal heave stability of supported circular excavations in clay," *Tunnelling and Underground Space Technology*, vol. 61, pp. 145–149, 2017.
- [28] J. L. Elman, "Finding structure in time," *Cognitive Science*, vol. 14, no. 2, pp. 179–211, 1990.
- [29] M. I. Jordan, "Serial order: a parallel distributed processing approach," *Advances in Psychology*, vol. 121, no. 97, pp. 471–495, 1997.
- [30] J. Schmidhuber, *Gradient Flow in Recurrent Nets: The Difficulty of Learning Long-Term Dependencies*, Wiley-IEEE Press, 2001.
- [31] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE transactions on neural networks*, vol. 5, no. 2, pp. 157–166, 1994.

- [32] Schmidhuber, "Learning complex, extended sequences using the principle of history compression," *Neural Computation*, vol. 4, no. 2, pp. 234–242, 1992.
- [33] R. Williams and J. Peng, "An efficient gradient-based algorithm for on-line training of recurrent network trajectories," *Neural Computation*, vol. 2, no. 4, pp. 490–501, 1990.
- [34] B. Dalton, S. Sozubek, M. Saldana, V. Courville, and Huawei Technologies Co Ltd, "Reduction of parameters in fully connected layers of neural networks," 2019, U.S. Patent 10,509,996.
- [35] I. Goodfellow, Y. Bengio, and A. Courville, "Deep learning," *An MIT Press book*, 2016.