*Forest Chemicals Review*
*www.forestchemicalsreview.com*
*ISSN: 1520-0191*
*Sept-Oct 2021 Page No. 1774-1785*
*Article History: Received: 10 August 2021, Revised: 25 August 2021, Accepted: 05 September 2021, Publication: 31 October 2021*

# Bus Short-Term Load Forecasting Based on Visual Source Domain and Transfer Learning

**Haihong Bian*, Guozheng Xu**
Nanjing Institute of Technology, Nanjing, China
*Corresponding Author.

***Abstract:***

Insufficient historical data seriously affects the accuracy of the Gated Recurrent Unit(GRU) in predicting bus load. Because of the single component of bus load, the accuracy of GRU load prediction can be effectively improved by transferring and utilizing the load data of similar buses. Therefore, a short-term bus load forecasting method is proposed based on visual source domain extraction and transfer learning. During the execution, the bus load is visualized and sent to the convolutional neural network in two-dimensional images for bus load classification. Colored load images help to more intuitively understand the distribution of load data and facilitate the discovery of unique attributes of different types of buses. The same type of bus load data are transferred, and the prediction model based on GRU is established. The result confirms the validity of this method to help improve the accuracy of GRU bus load prediction. At the same time, the over-fitting phenomenon is further analyzed by sharing the hidden layer.

***Keywords****: load forecasting, transfer learning, visual source domain, over-fitting.*

## I. INTRODUCTION

Bus load is the sum of the load transmitted from the main transformer in the substation to the area with a small coverage. Because the scope of power supply is small, the composition of bus load is single, usually only one or two kinds of electrical behavior, such as construction load, industrial load, financial load, accommodation, and catering industry load. Bus-level load forecasting is an essential part of load forecasting, which is widely used in power flow calculation, dynamic evaluation, reactive power optimization, and other fields of the power system[1].

At present, there are two main methods used in bus load forecasting: the method based on system load distribution and the method based on the change law of bus load itself[2-3]. The former one allocates the system load to each bus according to a specific coefficient, without considering the difference of each bus, so the prediction accuracy is not high. The latter considers each bus's characteristics, modeling respectively, and the prediction effect is good, so it becomes a standard method of bus load forecasting. With the development of artificial intelligence and big data technology, the neural network has become the mainstream load forecasting model. Because of its strong nonlinear fitting ability, it has good performance

*Forest Chemicals Review*
*www.forestchemicalsreview.com*
*ISSN: 1520-0191*
*Sept-Oct 2021 Page No. 1774-1785*
*Article History: Received: 10 August 2021, Revised: 25 August 2021, Accepted: 05 September 2021, Publication: 31 October 2021*

on many forecasting tasks[4-6]. However, excellent neural network models often need to feed large amounts of data. Although there are many buses in a region, the historical load data of the same bus or new bus are often less, which is challenging to meet the needs of neural network training. Given the problem of insufficient training sample data, some scholars used transfer learning to complete the related prediction tasks. Since the variation law of holiday load is different from that of non-holidays, [7] proposed a learning framework based on weighted knowledge transfer, which was used to predict the daily peak load during holidays to reduce the negative transfer. Reference [8] adjusted the electricity data of the source domain in real-time in the process of transfer learning, so that it maintained a high correlation with the electricity data of the building to be predicted. Reference [9] combined clustering and classification, used feature fusion strategy to determine the source domain load data, and verified the effectiveness of this method in the prediction of building electricity load in a particular region. Reference [10] proposed a comprehensive load forecasting model based on bidirectional generative adversarial network data enhancement and transfer learning technology. Because there is only one data point per month in the monthly power load forecasting and the training data are seriously insufficient, [11] selected the best source region for transfer learning from the monthly power load data of 25 districts in Seoul City based on the similarity of population, meteorology and other factors between the source region and the region to be predicted. The focus of transfer learning research is how to select training data valid for the target domain from the source domain, and these studies have proposed different methods for the selection of source domains to solve the problem of poor transfer or negative transfer[12-14]. Therefore, reasonable selection of source domain can significantly improve the effect of transfer learning.

Given these, this paper proposes a short-term bus load forecasting method that combines visual source domain selection and transfer learning. The visualization source domain adopts a load data analysis method based on pictures proposed by Hong Wang[15], which is different from traditional line graph analysis. This method displays the load data in color graphs, which enhances the interpretability of data classification. It also helps to understand the characteristics and distribution of data quickly. Because of the single component of bus load, it is naturally marked data with distinctive characteristics. This feature is conducive to visualizing bus load and constructing a convolutional neural network model(CNN) for bus load classification. The bus load in a color map is easier to understand and analyze through people's intuitive feelings. The same type of bus load often follows the same prior distribution[16-17]. These load data are transferred, and the Gated Recurrent Unit(GRU) parameters are pre-trained and retained to improve the prediction accuracy of GRU.

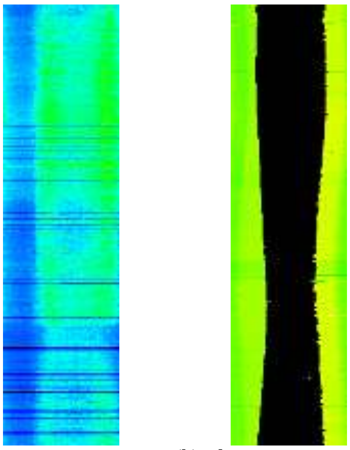
## II. ALGORITHM PRINCIPLE

### 2.1 Bus Load Visualization Method

The bus load photo is a two-dimensional pixel graphic representation of the bus load, as shown in Figure 1, where the x-axis represents the sampling interval of 15 minutes, 96 sampling time points per day, and the y-axis represents 365 days a year. Because public services and management electricity are affected

*Forest Chemicals Review*
*www.forestchemicalsreview.com*
*ISSN: 1520-0191*
*Sept-Oct 2021 Page No. 1774-1785*
*Article History: Received: 10 August 2021, Revised: 25 August 2021, Accepted: 05 September 2021, Publication: 31 October 2021*

by traffic lights that are not bright during the day, the day in the b chart is black. The load photo can be in PNG format or any other general image format, which all drawing tools or image viewers can recognize.



(a) Electricity consumption in real estate industry (b) Electricity consumption in public services and management

Fig 1: One-year load pictures of two buses in Nanjing

The specific steps of the visualization method are as follows:

(1)The original data is preprocessed to ensure that all data are sorted according to time and stored correctly. If some locations lack data or contain distorted data, the horizontal smoothing method is adopted in this paper, that is, the average value of the first 16 adjacent data points of the abnormal value is used to replace the abnormal value.

(2)The actual load is normalized by formula (1):

$$\theta_i = \frac{d_i - D_{min}}{D_{max} - D_{min}} \tag{1}$$

In formula (1), $D_{min}$ and $D_{max}$ are the minimum and maximum values of the actual load, respectively. $\theta_i$ is the normalized data of $d_i$. If $D_{min}$ equals $D_{max}$, then $\theta_i$ is set to 0.

(3)Convert the normalized load data $\theta_i$ to HSV $(h_i,s_i,v_i)$ color value. HSV is an alternative representation of RGB color pattern, which is more consistent with the way people perceive color attributes. When s=1, v=1, and $h \in [0^{\circ}, 360^{\circ}]$, the color bar is shown in Fig. 2(a). It can be seen that when $h=0^{\circ}$ and $h=360^{\circ}$, the color is too close to distinguish with the naked eye. In order to make the picture easy to distinguish, the new color bar shown in Fig. 2(b) is constructed.

(a)s=1, v=1, and $h \in [0^{\circ}, 360^{\circ}]$ (b)s=1, $v \in [0,1]$, $h \in [0^{\circ}, 250^{\circ}]$

Fig 2: Two color bars for different HSV parameters

*Forest Chemicals Review*
*www.forestchemicalsreview.com*
*ISSN: 1520-0191*
*Sept-Oct 2021 Page No. 1774-1785*
*Article History: Received: 10 August 2021, Revised: 25 August 2021, Accepted: 05 September 2021, Publication: 31 October 2021*

Use formula (2) to convert normalized load $\theta_i$ into HSV$(h_i,s_i,v_i)$ color value:

$$
\begin{aligned}
& h_i = 249*(1-\theta_i) \\
& s_i = 1 \\
& v_i = \begin{cases} 25*\theta_i & \theta_i \le 0.04 \\ 1 & \theta_i > 0.04 \end{cases}
\end{aligned}
\tag{2}
$$

(4)Convert HSV color value to RGB pixel value. HSV to RGB color conversion formula can be used in different forms and implemented through different software libraries (This article uses the built-in function of hsv2rgb in Matlab software for conversion).

(5)Arrange RGB pixel values in chronological order and draw bus load images

2.2 Transfer Learning

Transfer learning is a machine learning method that applies knowledge learned in one domain(i.e., source domain) to another domain(i.e., target domain) to improve or complete the learning effect in the target domain[18-19].

In bus load forecasting, when the historical load data of a bus(target domain) is less, it is not enough to train the network model. When there are a large number of similar types of bus (source domain) with similar electrical behavior, the accuracy of bus load forecasting can be greatly improved by using the appropriate transfer learning method. When a new bus appears, based on the classification results of the new bus, the existing network model of similar bus load forecasting can be reused to achieve rapid transfer and prediction, reduce the training model's time cost, and reflect timeliness.

2.3 Model Framework

Figure 3 shows the transfer learning prediction model. Firstly, the load data is preprocessed, drawing the load pictures. The specific steps are shown in Section 2.1. Secondly, CNN is used to determine the category of bus load pictures in the target domain and determine the source domain of transfer learning[20]. Then, according to the classification results of CNN, the bus load data in the source domain is transferred to the target domain and fused with the historical load data in the target domain. Finally, the GRU network is trained by using the fused data set.

*Forest Chemicals Review*
*www.forestchemicalsreview.com*
*ISSN: 1520-0191*
*Sept-Oct 2021 Page No. 1774-1785*
*Article History: Received: 10 August 2021, Revised: 25 August 2021, Accepted: 05 September 2021, Publication: 31 October 2021*

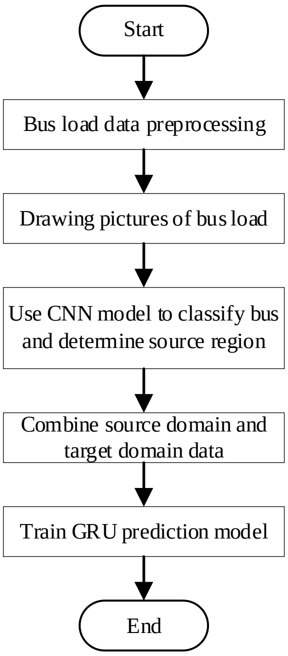



Fig 3: Bus short-term load forecasting model based on visual source domain and transfer learning

## III. EXAMPLE SIMULATIONS

### 3.1 Image Classification of Bus Load Based on CNN

(1)Load data set

The data set used in this paper is the power consumption data of some buses in May and June in Nanjing, 848 buses. The dataset contains seven types of buses, such as industry, construction, public service and management, and financial industry. Figure 4 is a randomly selected four-day industrial bus load picture, with a length of 96, a width of 61, and a cycle of 7 days. The picture clearly shows the periodicity of the load data.

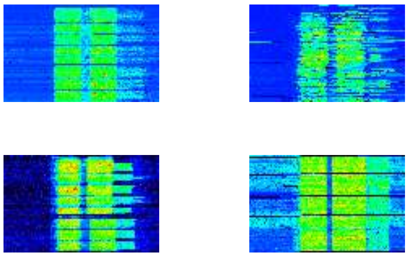


Fig 4: A four-day industrial bus load picture selected randomly

*Forest Chemicals Review*
*www.forestchemicalsreview.com*
*ISSN: 1520-0191*
*Sept-Oct 2021 Page No. 1774-1785*
*Article History: Received: 10 August 2021, Revised: 25 August 2021, Accepted: 05 September 2021, Publication: 31 October 2021*

(2)Setting parameters

The internal structure of CNN classification model is as follows:

The first layer: convolution layer, 10 convolution kernels, size 5×5, step 2, using BatchNormalization, activation function is ReLU.

The second layer: maximum pooling layer, size 2×2, step 2, dropout coefficient is 0.2.

The third layer: convolution layer, 20 convolution kernels, size 3×3, step 1, using BatchNormalization, activation function is ReLU.

The fourth layer: maximum pooling layer, the same as the second layer.

The fifth layer: flatten layer.

The sixth layer: fully connected layer, the output is 128.

The seventh layer: fully connected layer, the output is 7, corresponding to 7 types of bus.

(3)Analysis result

The load data of 600 buses are randomly selected as the training set, and 248 buses are used as the test set. The training process is shown in Figure 5.

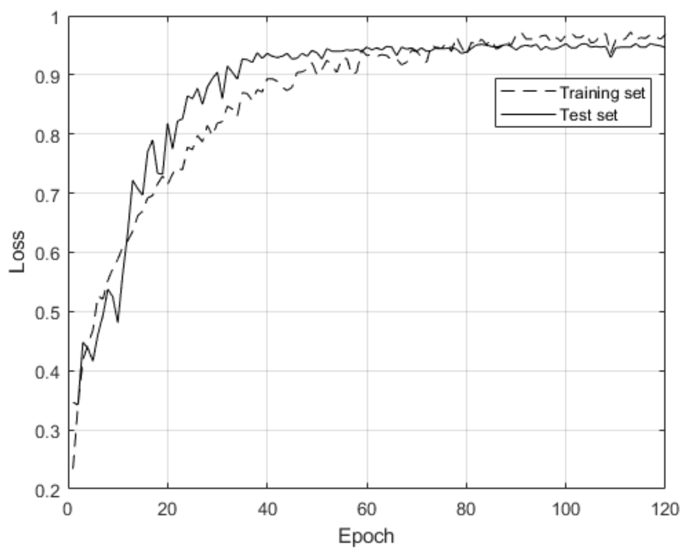



Fig 5: CNN model training process

*Forest Chemicals Review*
*www.forestchemicalsreview.com*
*ISSN: 1520-0191*
*Sept-Oct 2021 Page No. 1774-1785*
*Article History: Received: 10 August 2021, Revised: 25 August 2021, Accepted: 05 September 2021, Publication: 31 October 2021*

When the seven classification CNN model is not trained, the network parameters are randomly initialized, and the model's classification accuracy is only 1/7, which is equal to the probability of random classification. With the increase in the number of training, the classification accuracy is gradually improved. Because Dropout is added in the training stage, the accuracy of the test set is higher than that of the training set between 20 and 60 iterations. After 60 iterations, the accuracy of the training set of the model is still gradually improved, and the accuracy of the test set is unchanged. The model has a slight over-fitting phenomenon. In [15], the author compares the CNN classification model based on load images with the other five classification methods, including naive Bayes and SVM. The results show that the CNN model with load images as input has the highest recognition rate. In order to verify the effectiveness of the classification method, SVM, random forest and BP are used to classify the data set in this paper. The results are shown in Table 1.

**TABLE I. Accuracy of several different classification methods**

| Model | Accuracy |
|---|---|
| SVM | 93.21% |
| random forest | 92.77% |
| BP | 95.21% |
| CNN+load photo | 95.64% |

3.2 Bus short-term load forecasting based on visual source domain and transfer learning

This section selects the load of 112 industrial buses in the data set mentioned above as the simulation data. The sliding window was adopted in the experiment, and 96 moments on the seventh day were predicted through 576 moments on the first six days. A bus is randomly selected as the target domain and the remaining bus as the source domain. The three-week data training model from May 1 to June 18 predicts the load on June 25 in the fourth week. A three-layer neural network model is used in the experiment. The first two layers are GRU layers, and the last layer is fully connected layer.

In order to verify the prediction accuracy of the proposed method, the mean absolute percentage error(MAPE), root mean square error(RMSE) and mean square error(MSE) are selected to evaluate the prediction performance of the model. The smaller the three values, the higher the prediction accuracy.

(1)Results analysis of transfer learning

The experiment selects two methods for comparative analysis. The first method only uses the target domain load data training model, and the second method includes the target domain and source domain load data. Figure 6 shows the prediction results of the two methods, and table 2 shows the prediction error.

*Forest Chemicals Review*
*www.forestchemicalsreview.com*
*ISSN: 1520-0191*
*Sept-Oct 2021 Page No. 1774-1785*
*Article History: Received: 10 August 2021, Revised: 25 August 2021, Accepted: 05 September 2021, Publication: 31 October 2021*

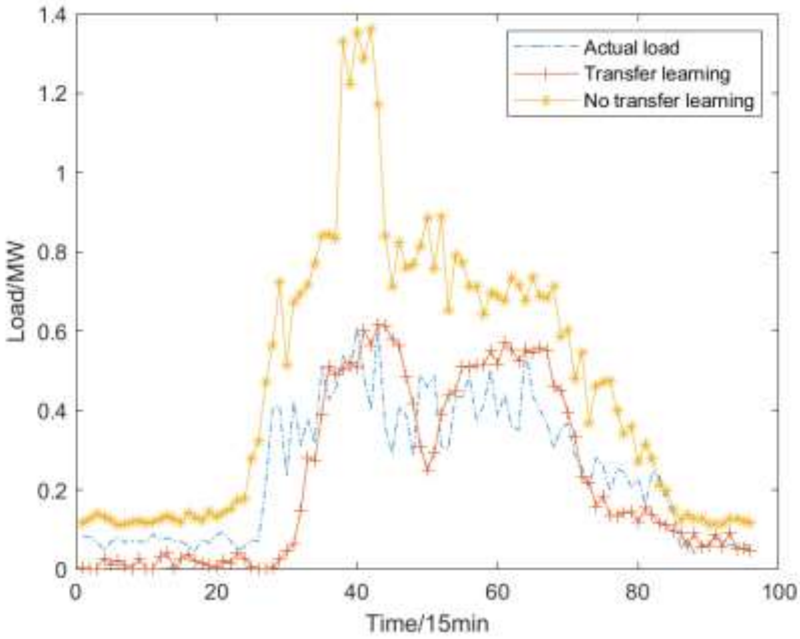



Fig 6: Model prediction results in two cases

**TABLE II. Model prediction error in two cases**

| | MAPE | RMSE | MSE |
|---|---|---|---|
| Transfer learning | 46.27% | 0.1171 | 0.0137 |
| No transfer learning | 96.60% | 0.3008 | 0.0905 |

It can be seen from the experimental results that without the transfer of source domain load data, the GRU prediction model is limited by the lack of data and does not show good prediction effect. Between 30 and 80 sampling points, the transfer learning model captures the characteristics of the industrial bus, basically showing a ' two peaks and one valley ' trend.

(2)Further analysis

After training the prediction model by feature transfer learning, the existing models can be reused to solve similar prediction tasks. Shih[21] combined the idea of the shared hidden layer with LSTM for load forecasting and achieved good prediction results. This shows that for the load forecasting task with a large number of source domain data, sharing the hidden layer is an effective way to improve the prediction accuracy.

In order to further explore the impact of data scarcity and shared hidden layer on prediction tasks, this

*Forest Chemicals Review*
*www.forestchemicalsreview.com*
*ISSN: 1520-0191*
*Sept-Oct 2021 Page No. 1774-1785*
*Article History: Received: 10 August 2021, Revised: 25 August 2021, Accepted: 05 September 2021, Publication: 31 October 2021*

paper uses the same three-layer neural network as above and selects the following three methods for simulation analysis.

Method 1: Transfering source domain prediction model, fixing the first layer GRU layer parameters and using only one bus load data training network.

Method 2: Same as Method 1, but the parameters of the first two GRU layers are fixed.

Method 3: Same as Method 1, but all neural network parameters are fixed, that is, the model does not participate in training.

The mean square error is used as the loss function, and the loss values of the three methods in the training process are shown in Figure 7, Figure 8, and Figure 9.

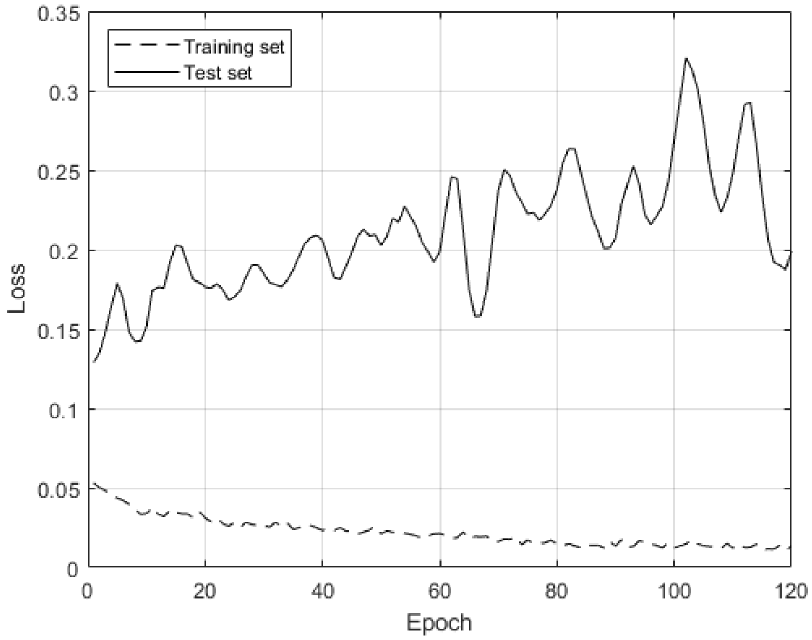



Fig 7: Method 1 training process

*Forest Chemicals Review*
*www.forestchemicalsreview.com*
*ISSN: 1520-0191*
*Sept-Oct 2021 Page No. 1774-1785*
*Article History: Received: 10 August 2021, Revised: 25 August 2021, Accepted: 05 September 2021, Publication: 31 October 2021*

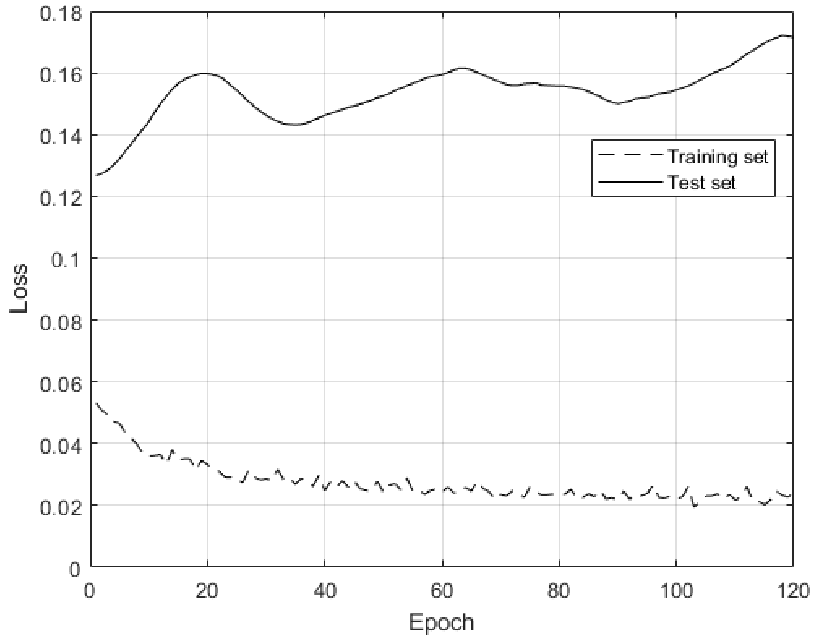



Fig 8: Method 2 training process

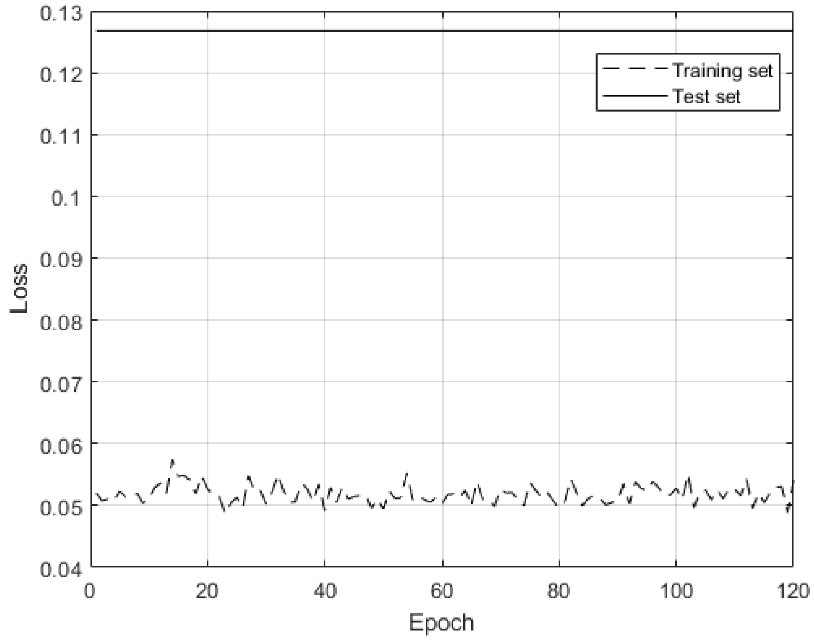



Fig 9: Method 2 training process

After reusing the existing prediction models, the training set and test set of the three methods initially

*Forest Chemicals Review*
*www.forestchemicalsreview.com*
*ISSN: 1520-0191*
*Sept-Oct 2021 Page No. 1774-1785*
*Article History: Received: 10 August 2021, Revised: 25 August 2021, Accepted: 05 September 2021, Publication: 31 October 2021*

have lower Loss values. Method 1 only fixes the parameters of the first layer network, and the parameters of the last two layers are updated with the increase of iterations. Due to the lack of data, the difference between the loss values of the training set and the test set continues to expand, and the model has a severe over-fitting phenomenon. Method 2 only allows the last layer of the network to update parameters, so the degree of over-fitting is slightly lower than Method 1. Method 3 model parameters are not updated, so the Loss value of the test set is unchanged.

Based on the analysis of the above simulation results, it can be seen that the short-term load forecasting of bus based on transfer learning can effectively improve the load forecasting accuracy of the bus with many source domains. Figure 1 and Figure 4 clearly show the behavior of bus electricity, giving people an intuitive and distinct feeling. In addition, Figure 5 and Table 1 show that the image classification method based on CNN is not weaker than the traditional load data classification method. The subsequent Fig. 6 and Fig. 7 show that it is reasonable to select the bus with similar electrical behavior for transfer learning because the superior nonlinear fitting ability of the neural network makes it prone to over-fitting in small sample prediction tasks. Transfer learning can effectively alleviate over-fitting and reduce prediction error.

## IV. CONCLUSION

Aiming at the problem that insufficient historical data affects the accuracy of the neural network to predict the bus load, the article proposes a short-term bus load forecasting method that combines visual source domain extraction and transfer learning. Based on load visualization, CNN is used to classify the bus load data and determine the source domain bus. Then, the source domain load data train the network model and realize the transfering feature. Finally, the influence of shared hidden layers and lack of training data on the prediction model is discussed. The above simulation results verify the effectiveness of the method proposed in the article, whether it is possible to perform secondary clustering on the same type of bus load data after classification, and to achieve more accurate source domain extraction is the next problem to be solved.

## ACKNOWLEDGEMENTS

This research was supported by National Natural Science Foundation of China (Grant No. 51877044).

## REFERENCES

[1] Chow J H, Wu F F, Momoh J A. Applied mathematics for restructured electric power systems//Applied mathematics for restructured electric power systems. Springer, Boston, MA, 2005: 1-9.
[2] Sun Qian. Research on forecasting technology for bus load and impacts of its characteristics on power grids. Hunan University, 2013.
[3] Zhu Heyan, Zhang Mingli, Guo Xiaoling, Yu Changyong, Li Yiran, Shen Chen. Net load prediction of bus

*Forest Chemicals Review*
*www.forestchemicalsreview.com*
*ISSN: 1520-0191*
*Sept-Oct 2021 Page No. 1774-1785*
*Article History: Received: 10 August 2021, Revised: 25 August 2021, Accepted: 05 September 2021, Publication: 31 October 2021*

considering influence of distributed photovoltaic power output. Electrical Measurement & Instrumentation, 2020,57(12): 69-74+104. DOI: 10.19753/j.issn1001-1390.2020.12.011.
[4] Wei Huadong, Tao Yuan, Cai Changchun, Hu Gang. Short-term load forecasting based on improved long short-term memory neural network. Electrical Measurement & Instrumentation, 2020,57(19):93-98. DOI: 10.19753/j.issn1001-1390.2020.19.015.
[5] Shi T , Mei F , Lu J , et al. Phase space reconstruction algorithm and deep learning-based very short-term bus load forecasting. Energies, 2019, 12.
[6] Wang Y , J Chen, Chen X , et al. Short-term load forecasting for industrial customers based on TCN-lightGBM. IEEE Transactions on Power Systems, 2020, PP(99).
[7] Zeng P, Sheng C, Jin M. A learning framework based on weighted knowledge transfer for holiday load forecasting. Journal of Modern Power Systems and Clean Energy, 2019, 7(2): 329-339.
[8] Zhang Yong, Tao Yifan, Gong Dunwei. Load forecasting of buildings using LSTM based on transfer learning. Control and Decision, 2021, 36(10): 2328-2338.
[9] Sun Xiaoyan, Li Jiazhao, Zeng Bo, Gong Dunwei, Lian Zhiyong. Small-sample day-ahead power load forecasting of integrated energy system based on feature transfer learning. Control Theory & Applications, 2021, 38(01): 63-72.
[10]Zhou D, Ma S, J Hao, et al. An electricity load forecasting model for integrated energy system based on BiGAN and transfer learning. Energy Reports, 2020, 6:3446-3461.
[11]Jung S M, Park S, Jung S W, et al. Monthly electric load forecasting using transfer learning for smart cities. Sustainability, 2020, 12.
[12]Bengio Y, Guyon G, Dror V, et al. Deep learning of representations for unsupervised and transfer learning. Workshop on unsupervised & transfer learning, 2011.
[13]Pan S J, Qiang Y. A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10):1345-1359.
[14]Lv X, Guan Y, Deng B. Transfer learning based clinical concept extraction on data from multiple sources. Journal of Biomedical Informatics, 2014, 52:55-64.
[15]Wang H, Qi L, Yan L, et al. Load p-hoto: a novel analysis method for load data. IEEE Transactions on Smart Grid, 2020, 12(2): 1394-1404.
[16]Wu Yanchen, Chenchang, Li Shaomei, Gao Chao. Person re-identification using attribute priori distribution. Acta Automatica Sinica, 2019, 45(05): 953-964.
[17]John Milyahilu,Jong Nam Kim. Generative probabilistic model with dirichlet prior distribution for similarity analysis of research topic. Journal of Korea Multimedia Society, 2020,23(4): 595-602.
[18]Sun Q, Liu Y, Chua T S, et al. Meta-transfer learning for few-shot learning//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 403-412.
[19]Deepak S, Ameer P M. Brain tumor classification using deep CNN features via transfer learning. Computers in biology and medicine, 2019, 111: 103345.
[20]Jiang Yuanyuan, Zhang Shuting. Research on soft fault diagnosis method of PV inverter based on improved VMD and CNN neural network. Electrical Measurement & Instrumentation, 2021, 58(02):158-163.
[21]Shih S Y, Sun F K, Lee H. Temporal pattern attention for multivariate time series forecasting. Machine Learning, 2019, 108(8): 1421-1441.