Journal of Artificial Intelligence Practice (2021) Vol. 4: 1-12
Clausius Scientific Press, Canada

DOI: 10.23977/jaip.2020.040101
ISSN 2371-8412

The Recognition of Tibetan Handwritten Numbers Based on Federated Learning

Zhuofan Mei

Jilin University Lambton College, Changchun, 130012, China
jovanmei@outlook.com

Keywords: Convolutional Neural Network, Federated Learning, Handwritten Recognition, Pattern Recognition, Privacy Protection

Abstract: Numeral recognition is closely related to individual lives, involving postal codes and bank checks. In recent years, despite many researches focusing on handwritten recognition, there were few on the identification of Tibetan handwritten numerals; and related to privacy protection. This paper proposed a recognition system based on lightweight CNN and federated learning, aiming to reduce the total calculating resource consumption and secure sensitive information. Besides, pre-processed TibetanMNIST dataset were adopted as the training samples for case study. This cases eventually obtained nearly 96% accuracy, and the expected time required to process a single image was approximately 0.017ms. The proposed system can be used for the recognition of Tibetan handwritten numbers.

1. Introduction

The recognition of handwritten numbers has attracted researchers' attention in recent years, which has been widely used in various industries, including cheque processing, automatic identification of postal code, tax intelligence system, etc. With the rapid growth of network informatization and Tibetan language information in Tibetan areas, the potential demand for Tibetan language processing technologies such as Tibetan language digital recognition is also increasing recently. However, due to the inconsistency of writing style, it is difficult to recognize the handwritten number accurately. Besides, individuals have handwritten signatures when they sign checks or contracts, but if people with wrong intentions use the signatures, it may cause property or other types of losses; that is why this kind of issues related to privacy leakage also need to be considered during the real-life application process as mentioned above.

In order to improve the performance of recognizing handwritten Tibetan numerals and protect the security of information, we will apply the Convolutional Neural Network (CNN), which could extract the discriminative feature to achieve accurate recognition of images. Furthermore, this technology will combine with federated learning [1], which can effectively protect private information.

During the experiment applied the CNN with the federated learning, we faced some technical problems, the difficulties as follows:

1. Pre-processing of Tibetan datasets is difficult because of the complexity of Tibetan digital images. It is challenging to get promising results with general neural networks.

2. Since Tibetan is a more pictographic script, there are similarities in the shapes of its different scripts. Specifically, the similarity of Tibetan handwritten numbers ‘4’ and ‘9’ is hard to tell the difference with the naked eye.

To address these challenges, the study applied some novel techniques, and the contribution of this paper can be summarized as follows:

1. The proposed system applies the lightweight convolutional neural network MobileNet V3 [2] to process large Tibetan digital data sets.
2. Federated learning is used in the proposed system to improve the speed of coverage while protecting information security.

The rest of this paper is organized as follows. Section 2 reviews the related works based on CNN and federated learning. Section 3 introduces the details about the proposed method. Section 4 describes the whole experimental process and reports the experimental results. The conclusions of this study and our future work are summarized in Section 5.

2. Related Work

This section consists of two parts, the first part will focus on handwritten text recognition based on CNN, and the studies related to privacy-preserving federated learning will be introduced in the second part.

2.1 Handwritten text recognition based on deep learning

Handwritten text recognition has been widely investigated in recent years; Akm Ashiquzzaman et al. [3] have proposed a model to finish the Arabic numeral recognition based on CNN. During the experiment, they introduced Rectified Linear Unit in the activation function to gradient disappearance, meanwhile preventing overfitting. In the study of Abu Sufian et al. [4], a new model called BDNet for recognizing Bengali handwritten numeral was proposed, and 99.78% experimental accuracy was achieved in the test. Besides, Vinay Uday Prabhu [5] release a new dataset based on MINIST called Kannada-MNIST and used it for training, which provides helpful experience in developing datasets for different languages. Duddela Sai Prashanth et al. [6] introduced the PR Tool in MATLAB to implement the Neural Network, which had been applied in recognition of handwritten Devanagari number, and eventually achieved an accuracy of above 95%. The hierarchical CNN was proposed by Zufar Kayumov et al. [7], which comprised of first level and second level neural network, and achieved a 99.79% test accuracy on the MNIST dataset. The studies referred have all improved on the basic CNN model, increasing handwriting recognition efficiency and achieving promising results based on the experimental dataset. The lightweight CNN MobileNet and federation learning will be applied in our experiment based on the above works.

2.2 Privacy protection based on federated learning

Federated learning is regarded as a flourishing and emerging technology that has attracted the interest of many researchers. Kunal Chandiramani et al. [8] applied the federated learning to the training process of the Fashion MINST dataset, which takes only 16.7 seconds to coverage. This study shows that federated learning can speed up the training process while maintaining privacy. Yu Chen et al. [9] projected a training-integrity protocol based on a trusted execution environment, ensuring the integrity of the deep learning process. Also, dishonest behavior can be quickly identified by the proposed system. Dong Yang et al. [10] proposed a novel framework that combined federated learning with semi-supervised learning and used it to detect the COVID-19 through the chest CT images from a multi-national database 1704 scans from three countries. This system achieved

promising results without sharing sensitive information and provided a promising direction for researchers. In general, federated learning is applied in many areas as mentioned above; by the same token, our study will also use federated learning for handwriting recognition.

3. Proposed Method

This section will introduce the technology and method that will be used in the experiment.

3.1 Lightweight Convolutional Neural Network

In order to process a large volume of Tibetan data, we applied the CNN model, which consists of four layers: Convolutional layer, ReLU layer, Pooling layer, and Fully-Connected layer. It is noteworthy that Jan LeCun [11] first introduced CNN's architecture in the early 1990s, the so-called LeNet. This architecture showed excellent performance in handwritten recognition and face detection. In recent years, it was only after AlexNet [12] was proposed in 2012 that CNN gained widespread attention.

Aiming to construct a light system, we chose the MobileNet V3, which is a combination of three models: the depthwise separable convolutions of MobileNet V1 [13], the inverted residual with a linear bottleneck of MobileNet V2 [14], and the lightweight attention model based on the squeeze and excitation structure of MnasNet [15].

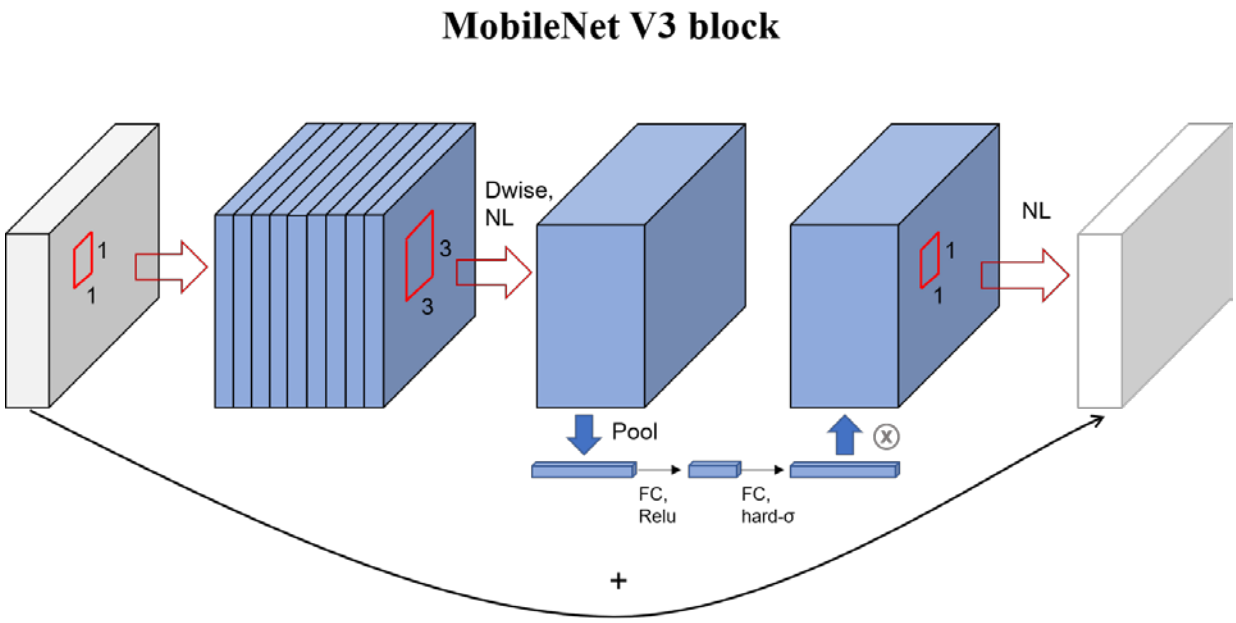

Figure 1. The architecture and block of MobileNet V3.

As shown in Figure 1, the 1 * 1 convolution is used for up-dimensioning firstly, and then the following operation is performed with residual edges. After inputting the 1 * 1 convolution for up-dimensioning, a 3 * 3 depth separable convolution is performed. The work modes of the attention mechanism are to adjust the weights of each channel. Consequently, the design reduces the amount of computing to increase the performance and speed of the processing.

3.2 Federated Learning

Since Tibetan recognition may involve users' sensitive information in real-life applications, this experiment applied federated learning in handwriting recognition to train the model while protecting privacy.

The primary attitude of federated learning is distributing the dataset across multiple devices called client nodes, training the model individually, then aggregating the models from clients, and eventually constructing the deep learning model at the server node [16]. The original idea was first introduced by Google [1] in 2016: devices are regarded as computational nodes that perform computations on local datasets to update the global model during the federated optimization. Also, each device only represents a portion of the total dataset. As we did in the experiment, the initial dataset was split into two parts averagely for each of the two client nodes, which created the necessary environment for the federated learning process.

Federated learning can be broadly divided into two categories: Horizontal federated learning and Vertical federated learning [16]. Since there are many samples with similar characteristics in the Tibetan handwritten dataset, this system uses Vertical federated learning. Specifically, participants train the model locally initially; subsequently, encryption algorithms are utilized to transfer the model without personal information to the server. The server will send back the federated model to participants after the secure aggregation, which allows the participants to update their previous model respectively.

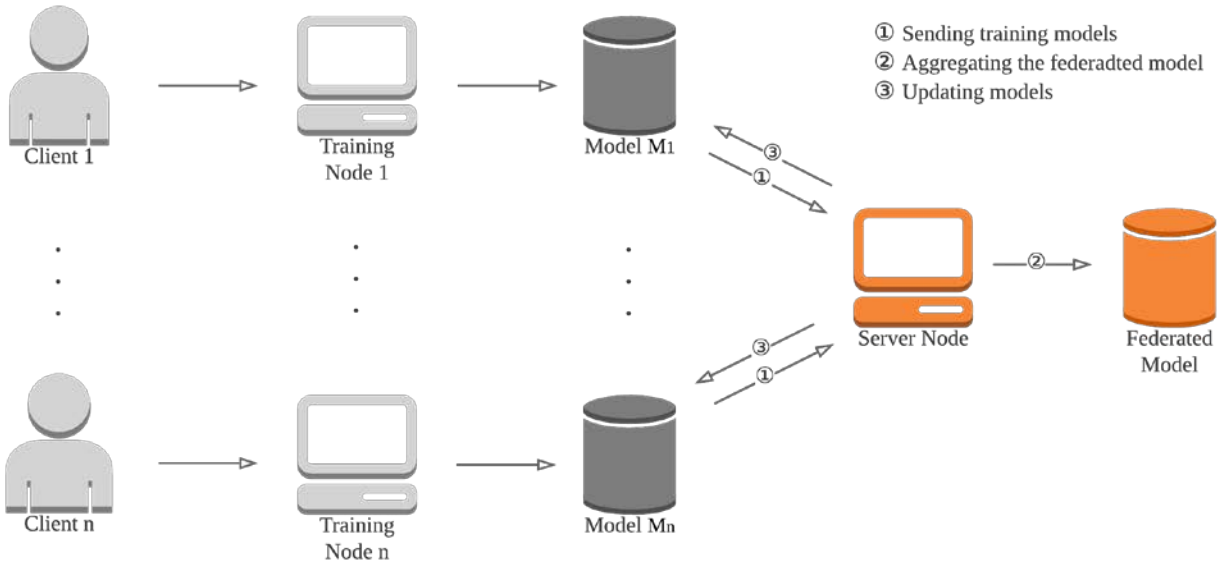

Figure 2. The process of federated learning.

The whole process is shown in Figure 2; as mentioned before, locally trained models are uploaded to the server and aggregated to build a federated model, send back to each client. In general, this procedure increases efficiency, and there is no spread of sensitive information.

3.3 Model Compression

We trained and tested around 18,000 sets of data during the experimentation; however, the amount of data and information is more tremendous than what was used in the experiment when users use the program in real life. In this case, the model is becoming increasingly important; this is because a

lightweight model can reduce the amount of computation and the cost of memory. In turn, the entire time of the process will be decreased. Besides, the system with the lightweight model can also be deployed on devices with low memory resources. Model compression is the crucial technique to achieve this purpose, and this methodology provides a promising direction for our future research.

In general, model compression can be classified into three categories: weights regularization, weights pruning, and low-precision [17]. Since each technique has its corresponding and suitable application, this passage will focus on weights pruning. To be more specific, a novel method called energy-aware pruning was proposed by Tien-Ju Yang et al. [18]. As shown in Figure 3, this method starts the pruning process based on the neural network layers' energy consumption. Then weights are adjusted locally to maintain the accuracy through estimating consumption, restoring, and fine-tuning weights. During the last procedure, the entire network will be globally fine-tuned using back-propagation after pruning all layers.

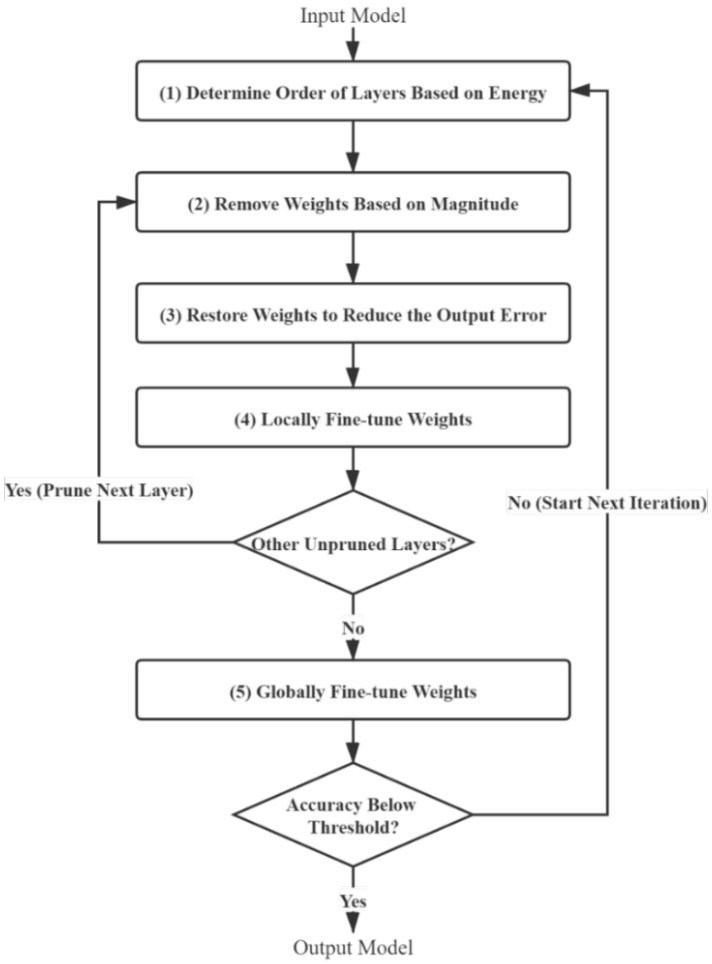

Figure 3. Process of energy-aware pruning.

The energy-aware pruning algorithm minimizes the overall energy consumption by pruning each layer's weights and lowering the consumption of energy. Even though the existing MobileNet V3 model is adequate for completing our experiment, we will include using this algorithm in the future study of Tibetan handwritten recognition to achieve model compression and maximize the energy reduction.

4. System Implementation

As mentioned earlier, the primary purpose of the experiment is to present a recognition system for handwritten numerals of Tibetan. For the implementation of the CNN, a Python-based library called Tensorflow [19] is used. All of the experiments have been done on the Google Colaboratory with Intel(R) Xeon(R) CPU, 12GB RAM, and TeslaK80.

4.1 Experimental dataset

To evaluate the performance of the proposed method, this experiment applied the TibetanMNIST dataset [20], which was created by the Institute of Tibetology, Minzu University of China. The researchers repeated scanning more than 300 times of repeated screening to get 17768 high-definition Tibetan handwritten digital images, as shown in Figure 4. In general, Tibetan is mainly composed of regular scripts and figures, and the TibetanMNIST used in the experiment is precisely the number in Figure Tibetan.

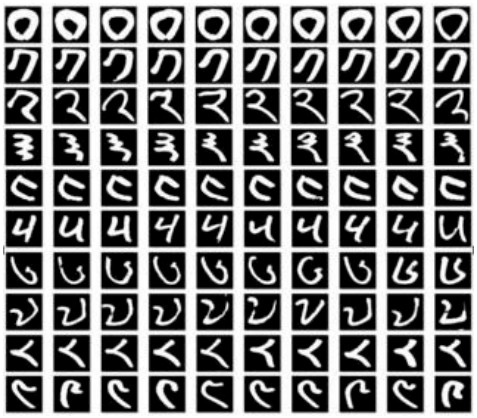

Figure 4. Sample of the Tibetan handwritten number.

During the training process, all the datasets have been divided into three independent datasets: training dataset, validation dataset, and testing dataset with the ratio of 70:20:10, respectively.

4.2 Training Process

The current dataset should be pre-processed to uniform picture size and picture type, making network training easier than the previous iteration while improving the recognition accuracy. After that, we define the MobileNet V3 CNN model, which will help the recognition system extract features from image samples.

The single machine is used to simulate the whole process of federated learning. The entire training process achieves the purpose of federal learning. To be more specific, the clients upload the local training model to the server. Subsequently, clients would update the model after the federated model is aggregated on the server-side. N training nodes are in our system, so the whole dataset is divided into N copies and assigned to each node. A folder called server node provides a space for clients to upload models and the server to reload models.

The accurate rate is the most intuitive indicator to show the experimental results to record the experiment's accuracy and loss values. Besides, the experiment applies the previously mentioned Tibetan handwritten number dataset, which was input to the neural network at the corresponding node for training. The experiment set every 32 images to each iteration size and configured the gradient

descent optimizer's initial learning rate to 0.01. The two client nodes will also swap the training model at the 1000th iteration, and the federated model was formed after 5000 iterations. The results demonstrate in Figure 5; the accuracy's value of the two client nodes increased alternately, eventually reaching the value of about 0.95 after 3500 iterations, even though the accuracy of the first set of data trained after each model exchange was low. As a result, we can use this method for handwritten number recognition due to the promising results.

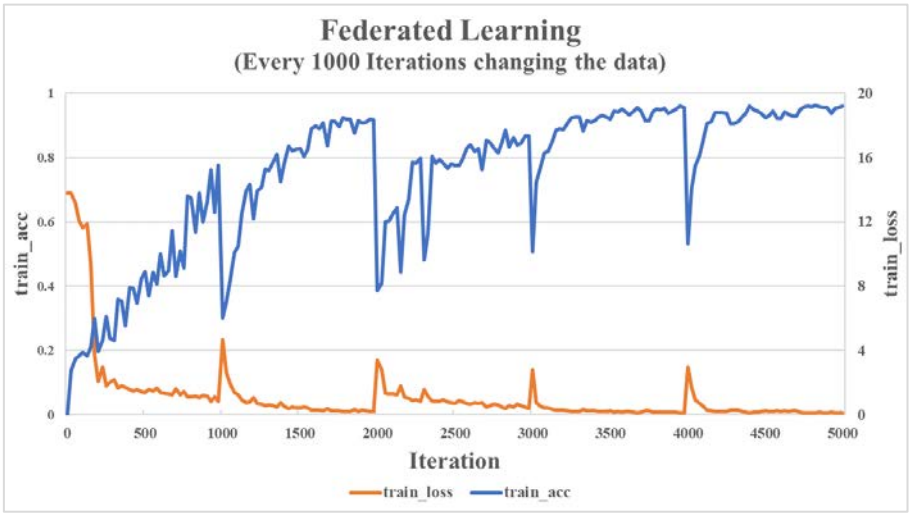

Figure 5. Accuracy and loss variation of when swapping models every 1000 iterations.

In the experiment, we applied the stochastic gradient descent method for the optimizer to increase convergence speed. Therefore, we carried out a control experiment on the optimizer's learning rate, which is used in our neural network, aiming to find the best experimental parameters and use them for the training model. Subsequently, we set the learning rate to 0.01 and 0.001 separately to compare the two models' training results.

The results are shown in Figure 6 when the learning rate is 0.01. The value of train-loss plummeted to about 2.5 after a few dozen iterations; the value of train-accuracy reached about 0.95 at 3500 iterations and converged at the end of 5000 iterations.

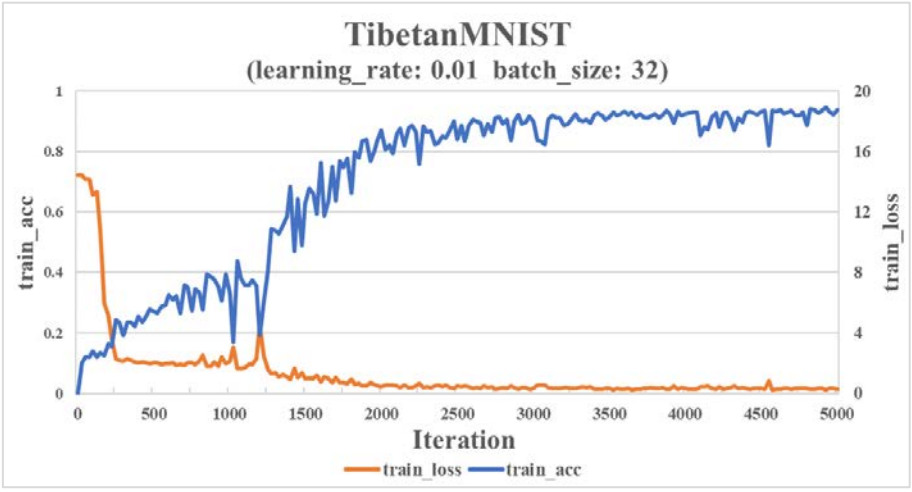

Figure 6. Accuracy and loss variation of the TibetanMNIST dataset when the learning rate is 0.01.

The results are shown in Figure 7 when the learning rate is 0.001. It shows the loss's value dropped sharply in the first few dozen iterations, but the training set failed to converge after 5000 iterations. As a result, the parameter of this experiment is not suitable for performing training models.

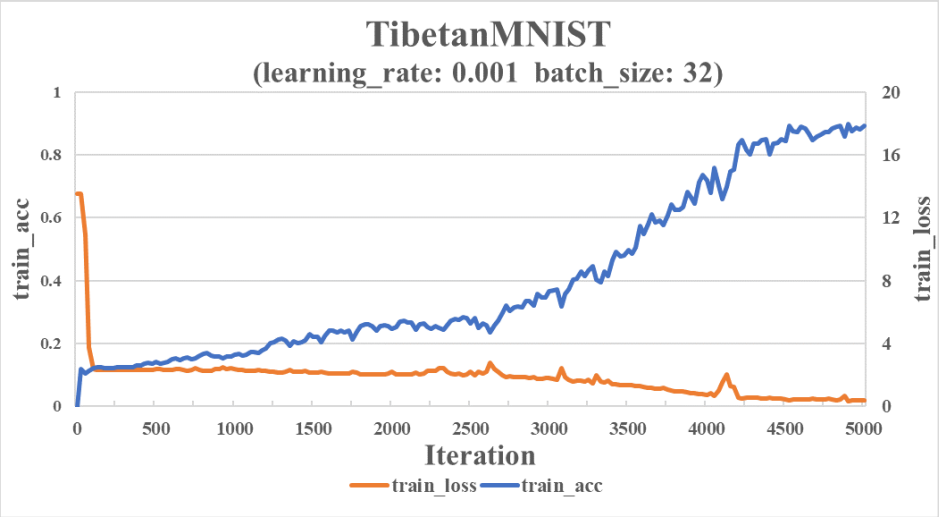

Figure 7. Accuracy and loss variation of the TibetanMNIST dataset when the learning rate is 0.001.

Afterward, we would follow the training results in Figure 5 for federated learning after finishing the controlled experiment mentioned above. The server node needed to send the training models to other client nodes during the federated learning simulation; because this process would ascertain whether the training models from different client nodes can be combined successfully; the value of accuracy and coverage speed can be influenced. Hence, the selection of the time point for the exchange model is truly essential.

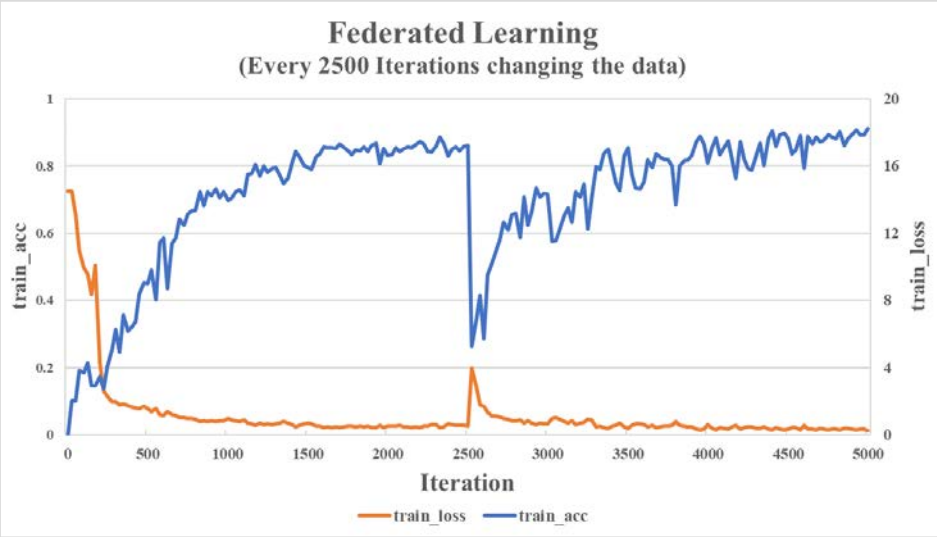

Figure 8. Accuracy and loss variation of when swapping models every 2500 iterations.

As shown in Figure 5, we set up a model exchange every 1000 iterations in the main experiment and finally obtained a well converged federated model based on the dataset distributed to each client. Besides, we conducted a group of control experiments to find the most suitable time point for

swapping the model. In detail, this control experiment would exchange the model every 2500 iterations; however, as Figure 8 shows, the value of train-accuracy has become stable after 4500 iterations. The main experiment's training results were better than this control experiment; thus, we will not apply this parameter in the test.

In another control experiment, we set the model to swap every 500 iterations. The result, as shown in Figure 9, although this training model converged after the 4000th iteration and achieved an accuracy rate of about 94, the cost was higher than the main experiment; therefore, we would not use this parameter either.

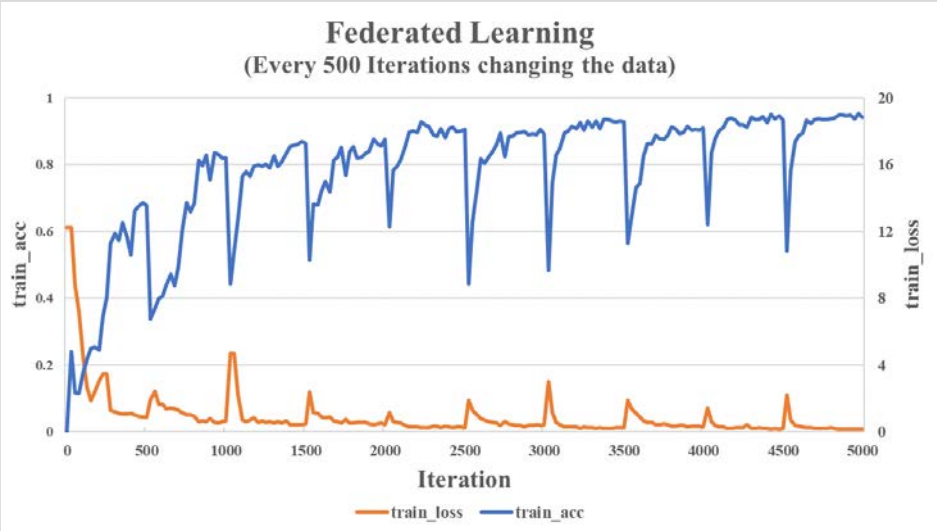

Figure 9. Accuracy and loss variation of when swapping models every 500 iterations.

Table 1 Comparative study for handwritten numeral recognition.

Method	Experimental Dataset	Accuracy
NN using PRTool [6]	Handwritten Devanagari Number	90.12%
AlexNet [21]	Handwritten Arabic Number	92.33%
Pretrained CNN using SVM [5]	Kannada-MNIST	96.84%
MobileNet V3	Tibetan-MNIST	95.94%

We have compared our experiment with other works concentrating on digit recognition in different languages, similar to Tibetan numbers. As shown in Table 1, even though our experiment's accuracy is slightly lower than Kannada-MNIST, we used federated learning to protect information security. It also inspires us to apply the SVM classifier in our future work.

4.3 Results

This study conducted a prediction test to verify the training effect of the experiment. Initially, we need to pre-process the input image for the model to recognize the input objects better. Specifically,

the jaggies would be removed firstly from the input image, then converting it to grayscale, transformed into a matrix meanwhile.

Since the character shapes of ‘4’ and ‘9’ are very similar in Tibetan, we input the handwritten images of these two items respectively to the prediction test. The output then displayed in a single image which consists of the original type of input number, the recognition result, and the processed image.

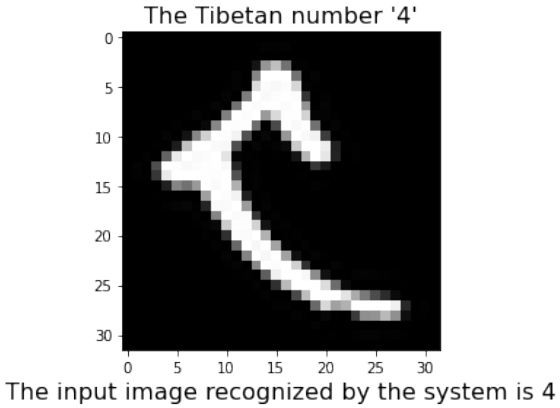

Figure 10. The predicted result of Tibetan handwritten number ‘4’.

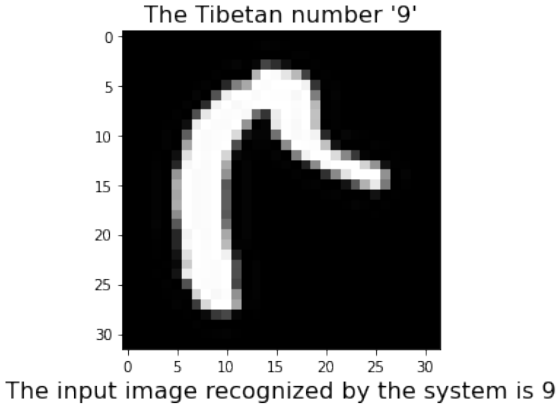

Figure 11. The predicted result of Tibetan handwritten number ‘9’.

The recognition results are shown in Figure 9 and Figure 10; the system has successfully identified the input Tibetan handwritten numbers. Overall, the prediction test results achieved a promising outcome, which indicates that our system can be used for handwriting recognition based on Tibetan handwritten numbers.

4.4 Discussion

The primary attitude of this experimentation is to construct a recognition system based on Tibetan handwritten numbers. The MobileNet V3 network combined with federated learning was applied to identify the handwritten numbers based on prior works. MobileNet V3 network reduces the weight and size of the model, and for federated learning, which protects the experimental dataset’s privacy and enhances the speed of model convergence simultaneously. As mentioned previously, we propose

to introduce the model compression method based on this experiment to simplify the model further and reduce consumption.

Even though the proposed system is capable of recognizing handwritten numbers, further improvements are needed. Our future research will use the weight pruning technique to achieve model compression; apart from it, we project to expand the experimental dataset size. The proposed recognition system will be more suitable for training models in simulated realistic scenarios with the above enhancements.

5. Conclusions

In this paper, we proposed a recognition system based on lightweight CNN MobileNet V3 and federated learning to protect data security and accelerate convergence. Also, we made a detailed analysis of the proposed system by comparing the results acquired by adjusting the network's parameters. With the inclusion of the method discussed above, we have obtained a nearly 96% recognition accuracy for Tibetan handwritten numbers. In our future work, we will focus on applying model compression techniques while scaling up the experimental dataset.

References

[1] Konečný, J., McMahan, H. B., Ramage, D., & Richtárik, P. (2016). Federated optimization: Distributed machine learning for on-device intelligence. arXiv:1610.02527.

[2] Howard, A., Sandler, M., Chu, G., Chen, L. C., Chen, B., Tan, M., ... & Adam, H. (2019). Searching for mobilenetv3. In Proceedings of the IEEE/CVF International Conference on Computer Vision, 1314-1324.

[3] Ashiquzzaman, A., & Tushar, A. K. (2017). Handwritten Arabic numeral recognition using deep learning neural networks. In 2017 IEEE International Conference on Imaging, Vision & Pattern Recognition (icIVPR), 1-4. IEEE.

[4] Sufian, A., Ghosh, A., Naskar, A., Sultana, F., Sil, J., & Rahman, M. H. (2020). Bdnet: Bengali handwritten numeral digit recognition based on densely connected convolutional neural networks. Journal of King Saud University-Computer and Information Sciences.

[5] Prabhu, V. U. (2019). Kannada-MNIST: A new handwritten digits dataset for the Kannada language. arXiv: 1908.01242.

[6] Prashanth, D. S., Mehta, R. V. K., & Sharma, N. (2020). Classification of Handwritten Devanagari Number–An analysis of Pattern Recognition Tool using Neural Network and CNN. Procedia Computer Science, 167, 2445-2457.

[7] Kayumov, Z., Tumakov, D., & Mosin, S. (2020). Hierarchical convolutional neural network for handwritten digits recognition. Procedia Computer Science, 171, 1927-1934.

[8] Chandiramani, K., Garg, D., & Maheswari, N. (2019). Performance analysis of distributed and federated learning models on private data. Procedia Computer Science, 165, 349-355.

[9] Chen, Y., Luo, F., Li, T., Xiang, T., Liu, Z., & Li, J. (2020). A training-integrity privacy-preserving federated learning scheme with trusted execution environment. Information Sciences, 522, 69-79.

[10] Yang, D., Xu, Z., Li, W., Myronenko, A., Roth, H. R., Harmon, S., ... & Xu, D. (2021). Federated semi-supervised learning for COVID region segmentation in chest CT using multi-national data from China, Italy, Japan. Medical image analysis, 70, 101992.

[11] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.

[12] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25, 1097-1105.

[13] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv: 1704.04861.

[14] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition, 4510-4520.

[15] Tan, M., Chen, B., Pang, R., Vasudevan, V., Sandler, M., Howard, A., & Le, Q. V. (2019). Mnasnet: Platform-aware neural architecture search for mobile. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2820-2828.

[16] Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning: Concept and applications. ACM Transactions on Intelligent Systems and Technology (TIST), 10(2), 1-19.

[17] Cheng, Y., Wang, D., Zhou, P., & Zhang, T. (2017). A survey of model compression and acceleration for deep neural networks. arXiv: 1710.09282.
[18] Yang, T. J., Chen, Y. H., & Sze, V. (2017). Designing energy-efficient convolutional neural networks using energy-aware pruning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 5687-5695.
[19] Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... & Zheng, X. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv: 1603.04467.
[20] Yuan, M. Q., Cai, R. X. M., & Tang, J. A. (2018). TibetanMNIST - Tibetan handwritten digital dataset. https://www.kesci.com/mw/dataset/5bfe734a954d6e0010683839/content
[21] Gupta, D., & Bag, S. (2021). CNN-based multilingual handwritten numeral recognition: A fusion-free approach. Expert Systems with Applications, 165, 113784.