Few-shot Thangka image classification based on improved DenseNet

Guoyuan He¹²*, Panpan Xue¹³, Jiahao Meng¹²
¹Key Laboratory of China’s Ethnic Languages and Information Technology of Ministry of Education, Northwest Minzu University, Lanzhou, Gansu 730000, China
²National Languages Information Technology, Boudler, Northwest MinZu University, Lanzhou 730000, China
³School of Mathematics and Computer Science, Northwest Minzu University, Lanzhou 730000, China
*Corresponding author’s e-mail: y192730592@stu.xbmu.edu.cn, 1441146936@qq.com

Abstract. Thangka is an important part of Tibetan culture, and the classification of Thangka image is one of the basic works of Thangka research. DenseNet (Densely connected convolutional networks) has achieved a very good effect in the field of image classification. Considering that the DenseNet adopts ReLU function which loses the negative feature of the image in the feature propagation process, this paper proposed an improved DenseNet, called L-DenseNet that Leaky ReLU replaces ReLU function to increase the negative feature of propagation. In order to solve the problem of insufficient Thangka image sample, we adopted the method of based fine-tuned network. Experimental results show that L-DenseNet obtains an outstanding performance, which improves 1.1% performance compared with DenseNet. Compared with other CNNs, such as VGG16, ResNet50 and InceptionV3, L-DenseNet obtains state-of-the-art performance in the classification of Thangka image.

1. Introduction
Thangka is a unique art form of painting in Tibetan culture. Its subject is related to the history, politics, culture, social life and many other fields of Tibetan culture. It has distinctive national characteristics, strong religious color and unique artistic style. Thangka Depicting the world of the sacred Buddha with bright colors. In the background of cultural globalization, the inheritance and protection of Thangka art have become the consensus of people. The Thangka image of collection, excavation, collation and protection are the need of both the era and the social development. Digitized protection of Thangka contributes to the inheritance and development of Thangka culture. Effectively distinguishing different categories of Thangka images is the important work of Thangka field knowledge excavation and Tibetan culture learning. It also provides a strong foundation for digitized protection of Thangka. It is particularly important to classify Thangka images due to the fact that there are many different types of Buddha, Bodhisattva, and Vajra in Thangka, some of them have multiple incarnations, and the images worshipped by different denominations are also different, this caused very difficult to recognize them.

In recent years, the study of image classification has made great progress, and its main methods include traditional ways and image classification approaches based on deep learning. Traditional classification methods, such as random forest (RF), decision tree, and support vector machine (SVM), which obtain a good classification effect for ordinary images. At present, classification method based on deep learning has become
the hot point of image classification research, and has been applied for various fields, such as remote sensing images[1], fire images[2] and medical images[3]. So far, image classification methods based on CNNs have made good development, among which GoogLeNet, VGGNet, ResNet and DenseNet are uppermost. GoogLeNet[4] obtains the champion of the ImageNet visual recognition in 2014, and it promotes utilization efficiency of parameters and reduces the dimensionality of input. Although GoogLeNet has greatly improved the accuracy of image classification. With the increase of the number of layers, and the gradient vanish problem becomes more and more obvious. VGGNet[6] was the runner-up of ImageNet visual recognition challenge in 2014. It used deeper network structure, smaller convolution kernel and pooled sampling domain, which makes it obtain more image features and controls more parameters, to avoid excessive computation and very complicated networks structure. ResNet, proposed by Kaiming He et al.[7], successfully trained a 152-layer neural network through using the ResNet Unit and obtained the champion of ILSVRC in 2015. It solved the gradient vanish problem while deepening the network. DenseNet, proposed by Gao Huang et al.[8], is superior to ResNet in performance.

In the study of Thangka image classification, Wanpin Gao et al proposed a different Thangka image classification approach based on sub-block color histogram. First, they divide the Thangka images into subblocks, extract the histogram features of every image sub-block, then identify and classify the image using histogram intersection method. Tiejun Wang et al.[5] constructs a multi-core SVM based on the information entropy feature-weighted radial basis kernel function, which effective classification of the icon image and the mandala image in Thangka is realized. Due to the Thangka has complicated features, the biggest problem of Thangka image classification is that it is difficult to obtain a good classification effect through the traditional ways. Hence, this paper uses CNN model to classify Thangka images for the first time. We proposed an improved DenseNet, called L-DenseNet that Leaky ReLU replaces ReLU function to increase the negative feature of propagation, and it is used to classify three types of Thangka images(See Figure.1). In order to solve the problem of insufficient sample, we used the method of few-shot learning.

This paper has used the technology based on fine-tuned network to classify Thangka images.

![Figure 1](image.png)

Figure 1 The example of Thangka image

2. DenseNet

DenseNet, proposed by Gao Huang et al.[8] in the 2017 CVPR, was rated as the best paper of CVPR 2017. DenseNet break away thinking that deepens the number of network layers and widens the network structure to improve the performance of network. From the perspective of features, DenseNet reduces the number of network parameters and alleviates the gradient vanish problem by reuse of feature and setting of bypass. It enhances the reuse of features and is easier to train through reducing the number of parameters, and has the regular effect. The layout is shown in Figure.2.

As can be seen from Figure 2, the output of each layer is seen as input of the subsequent layer, which reinforces the reuse of features. Between layer and layer of network adds three continuous operation, Batch Normalization(BN)[11], ReLU, pooling, or Convolution The $i^{th}$ obtained before all the feature map $x_0, x_1, ..., x_{i-1}$ as input:

$$x_i = H_i([x_0, x_1, ..., x_{i-1}])$$

(1)
where \([x_0, x_1, \ldots, x_{i-1}]\) refers to the concatenation of the feature maps produced in layers \(0, \ldots, i-1\) [8].

![Figure 2 DenseNet Architecture](image)

**Figure 2** DenseNet architecture

### 3. Improved DenseNet

The ReLU[10] function is an activation function widely used by CNN. ReLU have solved the problem that the absolute value of the derivative of the activation function is less than 1 and vanishes to 0 in the process of continuous multiplication by back propagation, but it is particularly sensitive to outliers because its derivative is always zero at the segment whose input is negative. This outlier may make ReLU shut, and kill the neuron. Considering that DenseNet adopts ReLU function that loses the negative feature of the image in the feature propagation process, this paper proposed an improved DenseNet, called L-DenseNet that Leaky ReLU replaces ReLU function to increase the negative feature of propagation.

#### 3.1. Leaky ReLU function

The Leaky ReLU function is the variant of the ReLU, whose output has the negative features of image. The derivative is always non-zero, which reduces the occurrence of silent neurons and allows gradient learning. Therefore, the Leaky ReLU solves the problem that neurons do not learn after the ReLU function enters the negative interval. The mathematical expression of the Leaky ReLU is Eq(2).

\[
f(x) = \begin{cases} 
  x & x \geq 0 \\
  \frac{x}{a} & x < 0 
\end{cases}
\]  

(2)

When the input characteristic information is less than zero, Leaky ReLU gets a small slope (See Figure 3). Hence, it retains the negative information of feature graph and increases the learning of effective feature information. In this paper, we take \(a = 5\).
3.2. L-DenseNet architecture

In order to improve the classification performance, we proposed L-DenseNet model for Thangka image classification. The architecture of our proposed L-DenseNet (Figure 4) contains four dense blocks and four transition layers, and each dense block contains BN, Leaky ReLU and convolution.

4. Experiments and analysis

4.1. Data augmentation

Due to the amount of dataset is limited, data augmentation is a good method that increases the diversity of training samples, which improves the robustness of the model and avoids overfitting. In computer vision, typical data augmentation methods include flip, rotate, scale, random crop or pad, color jittering and noise, and it is generally divided into offline expansion and online augmentation. Offline expansion is to perform all necessary transformations in advance, radically increase the size of the dataset. Online augmentation performs these transformations on small batches before sending them to the machine learning model. In this paper, we chose the method of online augmentation to expand dataset.
4.2. Dataset
Due to the scarcity of Thangka images and the large number of types, through screening, three types of Thangka images were obtained and classified, including 150 Buddha, 200 Sakyamuni, 250 Vajra. Considering the type samples quantity is limited, we take the image rotation and flip method to expand data set(Figure.5), and get 450 Buddha, 600 Sakyamuni, 750 Vajra.

Figure.5 The example of dataset

4.3. Overfitting analysis
We adopted the method of flipping and rotating, which not only expands dataset, also reduces the phenomenon of overfitting and improves the robustness of the model. Although they look similar, it is a different image for CNN. We used the technology of network fine-tuning to train our model, which further reduces the phenomenon of overfitting.

4.4. Experimental results and analysis

4.4.1. Evaluations.
To evaluate the performance of the algorithm, we use a train-validation-test scheme, with 70% of the data as a training set, 20% of the data as a validation set, 10% for the test set. The test accuracy is used to evaluate the performance of the model (See Eq.(7)).

\[
\text{precision}_x = \frac{\text{samples exactly classified as } x}{\text{samples of classified as } x}
\]  
(3)

Where \( x \) is a class of classified samples.

4.4.2. Experiments results.
In order to validate the correctness of this paper thought, we used different batch size and normalization experiment comparison(Table 1). Experiments results show that the accuracy increased with the increase of batch size, and batch size was chosen as the highest value of 32. However, regardless of the value of the Batch size and BN or GN, the accuracy rate of adopt Leaky ReLU activation function always more than ReLU activation function.

| Batch_size | Normalization | Activation function | Precision |
|------------|---------------|---------------------|-----------|
| 4          | GN            | ReLU                | 0.844     |
| 4          | GN            | Leaky ReLU          | 0.871     |
| 8          | GN            | ReLU                | 0.872     |
| 8          | GN            | Leaky ReLU          | 0.880     |
| 8          | BN            | ReLU                | 0.911     |
| 8          | BN            | Leaky ReLU          | 0.915     |
We made an experimental comparison with L-DenseNet and DenseNet (Table 2). Experiments results show that the performance of L-DenseNet improved 1.1% relative to DenseNet.

| Method      | Batch_size | Activation function | Precision |
|-------------|------------|---------------------|-----------|
| DenseNet    | 32         | ReLU                | 0.945     |
| L-DenseNet  | 32         | Leaky ReLU          | 0.956     |

We also compared with other CNNs models, like VGG16 [5] and ResNet50 [6] (Table 3). Experiment results show that compared with other CNNs model, our proposed L-DenseNet has the state-of-the-art performance for classification of Thangka images, and it is 4.7% higher than ResNet50. Based on the above experiments, it is feasible that Leaky ReLU can improve the performance of DenseNet.

| Method     | Batch_size | Precision |
|------------|------------|-----------|
| VGG16      | 32         | 0.868     |
| ResNet50   | 32         | 0.909     |
| InceptionV3| 32         | 0.931     |
| L-DenseNet | 32         | 0.956     |

5. Conclusion
In this paper, we studied the classification of Thangka images under the condition of few-shot by means of data augmentation, improved the DenseNet structure and called it L-DenseNet. Then we used the L-DenseNet model to classify the three types of Thangka images of Sakyamuni, Vajra and Buddha. The L-DenseNet contains four dense block and four transition layers, and the activation function of each dense block is Leaky ReLU function. Experimental results show that L-DenseNet performs better than DenseNet, VGG16, InceptionV3 and ResNet50 for the classification of Thangka images. In the future, our study will focus on the few-shot learning, such as the technology based on the distance distribution between samples and the method based on meta-learning, to further improve the accuracy and efficiency of Thangka image classification.

Acknowledgement
This work was supported in part by The Nature Science Foundation of China under Grant No.61862057, Fundamental Research Funds for the Central Universities under Grant No.31920200066, program for innovative research for innovative research team of SEAC (No.2018[98]) and Key Laboratory of China's Ethnic Languages and Information Technology of Ministry of Education and by the special fund for talent introduction of northwestern nationalities university.

References
[1] Tuia D, Ratle F, Pacifici F, et al. Active learning methods for remote sensing image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2009, 47(7): 2218-2232.
[2] Celik T, Demirel H, Ozkaramanli H, et al. Fire detection using statistical color model in video sequences[J]. Journal of Visual Communication and Image Representation, 2007, 18(2): 176-185.
[3] Kumar A, Kim J, Lyndon D, et al. An ensemble of fine-tuned convolutional neural networks for medical image classification[J]. IEEE journal of biomedical and health informatics, 2016, 21(1): 31-40.
[4] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1-9.
[5] Wang T, Wang W. Research on a Thangka Image Classification Method Based on Support Vector Machine[J]. International Journal of Pattern Recognition and Artificial Intelligence, 2019.
[6] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.
[7] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
[8] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4700-4708.
[9] Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks[C]//Proceedings of the fourteenth international conference on artificial intelligence and statistics. 2011: 315-323.
[10] Maas A L, Hannun A Y, Ng A Y. Rectifier nonlinearities improve neural network acoustic models[C]//Proc. icml. 2013, 30(1): 3.
[11] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift[C]// International Conference on International Conference on Machine Learning. JMLR.org, 2015.
[12] Wu Y, He K. Group normalization[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 3-19.
[13] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2818-2826.