

Expediting Training Process of Rice Grading Neural Netowks Using Singular Value Decomposition

Na Wenqi(22018000164)* and Chang Chengxia(22018000163)*

*School of Information Science & Engineering, Yunnan
University

Abstract

The abstract goes here.

- 1 Introduction
- 2 Related works
- 3 Methods

In this section we will introduce the method for accelerating training of deep networks via using SVD in order to reduce the number of parameters of networks. According Figure [1](#), we will argue our method in three steps as follows.

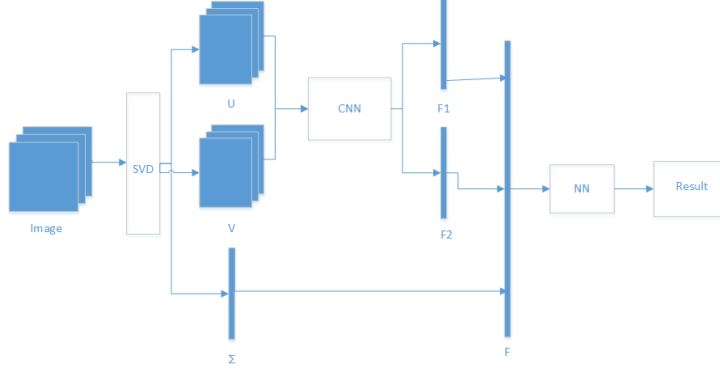


Figure 1: The neural network with SVD

First of all, we assume that the size of the input images is $h \times w \times c$. Namely the image height is h , the width is w , and the count of channel is c . Before inputting the images into the neural network, we first perform singular value decomposition on the image. The Result of decomposition is three tensors: U , V , and Σ , and the dimensions of them are: $h \times r \times c$, $w \times r \times c$, $r \times 1 \times c$ respectively, where $r \ll h$ and $r \ll w$.

We then stack the U and V before inputting them into our predefined convolutional neural network and obtain two vectors F_1 and F_2 through the convolutional layer and the flattened layer.

Finally, overlaying F_1 , F_2 , and Σ together to form a vector F , and then input F into the final fully connected layer of the convolutional neural network to get the output.

In this way, a large matrix input can be decomposed into two smaller matrices and a vector input. By adjusting the value of r , the number of parameters required by the neural networks can be controlled.

4 Experiments

4.1 Dataset

We conduct our experiments using the FIST-Rice dataset. FIST-Rice is a multi-sensor based open rice grading dataset that includes approximately 30,000 samples taken from three angles by three high speed cameras, where samples are divided into two categories: sound kernels and unsound kernels.

4.2 The Architecture

The introduction of SVD in the deep learning networks aims to reduce network parameters and shorten training time, thus improving training efficiency. How to design an effective deep learning network is not the main task of this paper. Therefore, in this paper, we implement the part of CNN and NN in the model based on VGG-16[1], and compare the plain VGG-16 network method and the VGG-16 method using SVD to illustrate the effect of the model.

During the actual experiment process, one thing needs special attention. Since there are many convolutional layers in the VGG model, and the matrix dimensions obtained by SVD are relatively small, these matrix inputs into the CNN model will converge to the fully connected layer faster than the original image. Therefore, we need to adjust the size of the convolution kernels or reduce the number of convolutional layers to ensure that the algorithm performs correctly.

The other issue that needs attention is that the singular value decomposition of the image consumes computational resources, which reduces computational efficiency and increases runtime. However, when we train a dataset, we usually perform dozens of epoches of training. At this time, if we cache the result of each image singular value decomposition, we do not need to perform singular value decomposition operation when repeating each epoch of calculation. Therefore, the operation time for singular value decomposition of an image can be regarded as a constant time, so that it is ignored when comparing computational efficiency.

4.3 Results

By comparing the results of the experiments from the VGG model and the model proposed in this paper, we found that the number of the network parameters is about 8 million, when using the SVD model and letting $r = 50$, which is less than the numbers of the parameters of the VGG-16 model(14 million) about 6 million. We can reduce by 57% parameters from the VGG model. From the running time point of view, the VGG model runs for about 4300 seconds, and after using SVD, the proposed model runs for about 2800 seconds and the running time is reduced by about 35%.

5 Conclusion

Our work proposed a method to apply singular value decomposition to convolutional neural networks in order to reduce network parameters and accelerate the training process. This method can be applied to the training process of the FIST-Rice dataset, which can effectively improve the training efficiency. At the same time, this method can also be applied to other datasets or convolutional neural networks.

There are currently two problems with this method. First of all, although this method can significantly reduce the network parameters and improve the efficiency from time to time, this method does not have any significant improvement on the classification accuracy of rice grading. In the case of improper training parameters, it will be counterproductive and the classification accuracy will be reduced. Another problem is that the convolutional neural networks are generally suitable for the case where the image height and width are approximated, and the two matrices U and V obtained by the singular value decomposition will have $r \ll h$ or $r \ll w$, so that U and V cannot be applied to deeper convolutional networks.

The next step we will be doing is to find an approach not only to reduce network parameters but also to be proper to apply to convolutional networks, while at the same time not affecting the classification accuracy of the networks.

References

- [1] Karen Simonyan and Andrew Zisserman, “Very deep convolutional networks for large-scale image recognition”, arXiv preprint arXiv:1409.1556, 2014. [3](#)