Multiple GPUs-Based Distributed Learning for Classification of Breast Cancer Images

Tengfei Luo [a]
Computer Science (Artificial Intelligence), University of Nottingham, Nottingham, U.K.

Keywords: Distributed Machine Learning, GPUs, Deep Learning, Breast Cancer.

Abstract: The study focuses on the classification of breast cancer images using deep learning techniques, particularly emphasizing the role of multi-GPU setups to handle the demanding computational needs of this task. Breast cancer, notorious for its high misdiagnosis rate, poses a significant challenge in medical diagnostics, where Computer-aided Diagnosis (CAD) systems can play a pivotal role. This study experiments with various convolutional neural network models including ResNet and MobileNet. These models are tested on a dataset divided into three categories—normal, benign, and malignant images—sourced from ultrasound scans. The dataset used comprises a substantial number of images, which are then processed and augmented to fit the model requirements. The study evaluates the models' performance based on accuracy and efficiency metrics, revealing that while multiple Graphics Processing Units (GPUs) theoretically increase computational speed, they do not always correspond to better model performance due to potential issues in data synchronization and parallel processing inefficiencies.

1 INTRODUCTION

With the recent advances in deep learning, computers are able to take a step forward from humans to perform complex tasks that have a high misdiagnosis rate. In medicine, Computer-aided Diagnosis (CAD) is becoming increasingly important, especially in diseases with high misdiagnosis rates, such as breast cancer, nonspinal fractures, and spinal fractures (Jeremy, 2013; Qiu, 2019; Qiu, 2022).

Among these diseases, breast cancer has the highest misdiagnosis rate due to its difficult diagnosis and detection (Ma, 2020), posing a serious threat to women's health as the second leading cause of death among women. Early diagnosis can significantly reduce the mortality rate (40 per cent or more). There are two ways to detect breast cancer. The first is mammography (Gøtzsche, 2013), which has a high resolution and is highly standardised, but it is costly, can lead to overdiagnosis and carries a risk of radiation exposure. The second method is ultrasound (Guo, 2018), which is radiation-free and has high sensitivity in detecting solid masses, but it has low resolution and is dependent on the experience and skill of the operator. Ultrasound images do not have any distinctive features compared to other medical images. Ultrasound features of breast cancer may include irregular shape, blurred borders, uneven internal echoes, etc. However, in benign lesions, there are no distinctive features that can be detected. Benign lesions (such as cysts or fibroadenomas) can sometimes show similar features, which can lead to diagnostic uncertainty. Ultrasound is healthier for people who have regular testing, but it is a condition that relies heavily on the experience and skill of the doctor, and therefore has a higher rate of misdiagnosis. Given deep learning's proficiency in learning from past experiences, the application of it can be considered in reducing the rate of misdiagnosis of breast cancer.

The technique being used in breast cancer detection is commonly known as machine learning. Machine learning is characterised by the researcher finding a filter or feature that makes the results clearer and then learning to find the values of the relevant features in an image to make a final judgement. Yali proposed the use of H-Scan image, in this ultrasound image there is a big difference in the result of benign and malignant tumours, benign breast tumours have more red areas and malignant tumours have more

[a] https://orcid.org/0009-0000-2766-8519

Luo, T.
Multiple GPUs-Based Distributed Learning for Classification of Breast Cancer Images.
DOI: 10.5220/0012937700004508

In *Proceedings of the 1st International Conference on Engineering Management, Information Technology and Intelligence (EMITI 2024)*, pages 333-336
ISBN: 978-989-758-713-9

blue areas, so it is good to classify them (Yali, 2019). But deep learning doesn't need humans to do feature extraction, the machine will analyse the image from a higher dimensional perspective. Because deep learning can learn some abstract features that are difficult for humans to understand. These features may be more useful for image classification. Zhantao proposed a supervised learning method that uses trained textures to classify breast tissue into different categories (Zhantao, 2019). Some researchers have tried to use deep learning models for training, Boukaache tried to use pre-trained convolutional neural network models such as VGG16, ResNet18 and ResNet50 for breast cancer image classification. (Boukaache, 2024) And got good results (97.8%). But the general training is done on a high-performance Graphics Processing Unit (GPU). The trained texture is not very easy to get in practical situation, ultrasound image is a more three-dimensional image, the picture is the same for different angles at the same position, in some angles the picture is very noisy and thus a good texture cannot be obtained.

So the main purpose of this research is to attempt to classify breast cancer images using multiple GPUs. This paper will use ResNet 18 and MobileNetV2, which are simpler models that can be trained on smaller machines. This paper also tried to compensate for the lack of performance of a single GPU by using multiple GPUs.

2 METHOD

2.1 Dataset Preparation and Preprocessing

This data utilized medical images of breast cancer based on ultrasound scans. The breast ultrasound dataset is divided into three categories: namely normal, benign and malignant images. Firstly, the dataset is medical images of breast cancer scanned with ultrasound (Al-Dhabyani W). The source of the dataset is downloaded from Kaggle (Kaggle, 2021). The original dataset classifies the images into three categories, benign tumours, malignant tumours and normal (no tumour) and is accompanied by images of the tumour location and its shape. There are 891 images of benign tumours, 421 images of malignant tumours and 266 images of normal. Figure 1 provides sample images on the dataset.

The raw data can be used for target recognition, so each image has not only the classification information but also the coordinates of the target. In this experiment, this paper only focuses on classification, so the location information contained in the data will be removed. After processing the images, the data is divided into a training set and a test set. The data is then enhanced. For different sizes of images, the images are firstly resized to 256x256 and then cropped in the centre by taking a square of length 224x224 with the centre as the origin. Then, the data is normalised. This paper also converted the pixel values from 0 to 255 to around 0 to 1. Finally, the image data was divided into a training set and a test set using 5-fold cross validation to evaluate and improve the generalisation of the model.

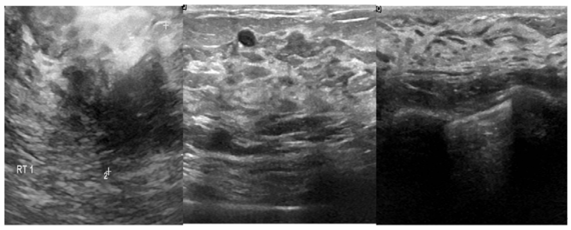

Figure 1: Breast image Malignant (Left) benign(middle) normal (right) (Photo/Picture credit: Original).

2.2 Model Establishment

In this study, MobileNet V2 is used to deal with an image classification task that is prone to overfitting (Sandler, 2018). MobileNet V2 is a simple deep learning model for very simple image classification tasks. If a complex model is prone to overfitting, it is common to choose a simple model in addition to tuning the parameters.

MobileNetV2 is a lightweight deep learning model optimised for mobile devices, with the core advantage of Depthwise Separable Convolution. This technique reduces the number of parameters and computational cost of the model, while providing efficient performance. The model also introduces Inverted Residual Blocks and Linear Bottlenecks to further improve efficiency. The last layer in these blocks usually does not use the ReLu activation function to prevent information loss.

The model was implemented using the PyTorch framework, and the accelerate library was used to support training in a multi-GPU environment. The accelerate library allows for parallel processing of data and training of the model in a multi-GPU environment, which theoretically improves the efficiency of training. At the end of the epoch, the study evaluates the model's calibration by its accuracy and saves the model when it finds a higher accuracy in a test.

This study used Resnet 34. Firstly, pretrained was set to false. To reduce overfitting, dropout

Table 1: The performance of various models based on different parameters.

	Batch size	Time	GPU number	Train accuracy	Test accuracy
ResNet34	64	771	2	0.99	0.42
ResNet34	128	840	2	0.99	0.46
ResNet34	128	1588.14	4	0.93	0.53
ResNet34(Dropout= 0.8)	64	771	2	0.88	0.43
ResNet 18	128	799	2	0.82	0.45
ResNet 50	64	770.45	2	0.7	0.45
ResNet 50	128	865	2	0.7	0.42
MobileNet V2	128	801	2	0.99	0.4
MobileNet V2(cross-validation)	128	801	2	0.99	0.83
MobileNet V2(cross-validation)	128	2200	4	0.99	0.63

regularisation was added by setting the dropout rate to 0.5. The optimiser was defined as stochastic gradient descent with a learning rate of 0.001 and momentum of 0.9. The model then uses a learning rate scheduler, which stops at the validation loss, and a learning rate scheduler, which stops at the validation loss, and a learning rate scheduler, which stops at the validation loss. The model then uses a learning rate scheduler to reduce the learning rate when the verification loss stops improving.

For this paper, ResNet34 which is a deep residual network that belongs to the ResNet of deep learning architectures was chose, proposed by He et al. proposed in 2015. The main idea of ResNet is to solve the degradation problem in the training process of deep neural networks through residual learning. Before the emergence of ResNet, if the number of layers increases there may be problems such as gradient disappearance, and the performance of the network will be saturated or even decline, rather than continue to improve.

ResNet34 is composed of 34 convolutional layers. The innovation of this network is the introduction of residual modules, each of which consists of two or three convolutional layers and is connected by jumps. The minnow connection allows the gradient to flow directly through multiple layers, thus increasing the efficiency of gradient propagation during training and allowing the network to learn deeper features.

A 3080x4 GPU was used for this study. After configuring the distributed model, they used the same dataset with batch sizes set to 64 and 128 to compare the results. The epoch is also set to 10, 50, 100 to compare the correctness rate. The highest accuracy of the training batch is then used as the final accuracy of the training.

The methodology and dataset used in this study are described above, and the experiments are conducted while controlling all other variables, in order to investigate the difference between running on a single GPU and running on multiple GPUs at the same time. The experiments are then quantified by two metrics: runtime and accuracy.

3 RESULTS AND DISCUSSIONS

As shown in Table 1, the best accuracy of this experiment is 0.83, using MobilNet V2 with a batch size of 128. In this experiment, four models, Resnet 18, Resnet 34, Resnet 50 and MobileNet V2, were tested. The batch size was set to 64 and 128, and the number of GPUs was set to 2 or 4.

From the experiments, it can be concluded that the batch size has little effect on the correctness rate in the case of overfitting, and cannot change the status quo of overfitting. In terms of time, as the depth of the model increases, the training time increases slightly, and when the number of GPUs is set from 2 to 4, the time increases to 3 times of the original one, which may be due to the inefficiency of data transmission and synchronisation. This suggests that the communication between GPUs may cause the training parallelisation to be inefficient.

Since the initial training with ResNet 34 resulted in a high rate of correct model training, the next experiments were conducted to make the model more generalisable by other methods. First, this paper tried using Dropout. All models were given a default Dropout value equal to 0.5, and then the paper tried expanding the value to 0.8 or 0.9. This did reduce the correctness of the training ensemble, but there was no significant increase in the generalisation ability.

When using ResNet50, the training results became even worse and did not increase the correctness of the test set, so the only way to solve the overfitting problem is to make the model as simple as possible. Next, the MobileNet V2 model was used in this study. Firstly, the amount of data is not very large, so this study uses cross-validation for random segmentation, which allows the model to learn more features. The model is then trained on different numbers of GPUs. When the number of GPUs is equal to 4, the experiment doesn't have excellent performance, which may be due to the improper configuration of the model parallelisation or the uneven distribution of the data and other problems. The best results are obtained on two GPUs.

However, a problem common to all models is that the best test set results tend to occur within 10 training runs, and as the number of runs increases, the correctness of the test set decreases. The training set basically reaches 99% around 80 times. The test set does not increase with training. This may also require reducing the complexity of the model.

In conclusion, the use of cross-validation significantly improves the test accuracy of MobileNetV2, demonstrating its importance for improving generalisation. ResNet18 is not yet proficient enough, and a cleaner model is needed to improve the accuracy. Furthermore, increasing the number of GPUs did not always reduce training time or improve accuracy, suggesting the need to optimise multi-GPU training strategies.

4 CONCLUSION

The purpose of this study is to investigate whether parallel computing on GPUs increases the performance of the trained model. From the results, it does not, because GPU training also needs to take into account the transfer of data between GPUs, and the integration time of the weights across GPUs increases as the number of GPUs increases. The results do not get better as the number of GPUs increases, and overfitting reappears as the number of training sessions increases. This model is characterised by the fact that the data is very easy for the model to overfit, and the increasing complexity of the model is not friendly to the extraction of features from simple images. Currently, there is no good classification for data that is overfitted because of the simplicity of the images. In the future, further study will try to find out which part of the model is slowing down the training process and try to improve the accuracy of the model by using libraries that allow multi-GPU training or algorithms that integrate the parameters of different GPUs. Further study will also try to get a model that can solve the problem of overfitting images easily.

REFERENCES

Boukaache, A., Benhassıne, N. E., & Boudjehem, D. 2019. Breast cancer image classification using convolutional neural networks (CNN) models. *International Journal of Informatics and Applied Mathematics, 6*(2), 20-34.

Cao, Z., et al. 2019. An experimental study on breast lesion detection and classification from ultrasound images using deep learning architectures. *BMC Medical Imaging, 19*, 1-9.

Guo, R., Lu, G., Qin, B., & Fei, B. 2018. Ultrasound imaging technologies for breast cancer detection and management: a review. *Ultrasound in medicine & biology*, 44(1), 37-70.

Gøtzsche, P. C., & Jørgensen, K. J. 2013. Screening for breast cancer with mammography. *Cochrane database of systematic reviews*, (6).

He, K., et al. 2016. Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Kaggle. 2021. Breast ultrasound images dataset. Retrieved from https://www.kaggle.com/datasets/aryashah2k/breast-ultrasound-images-dataset/code, last accessed time: April 13, 2024

Ma, Y. 2020. Diagnosis of Benign and Malignant Breast Lesions in Rats by MRI Plain Scan Combined with Diffusion-Weighted Imaging. *Revista Científica de la Facultad de Ciencias Veterinarias*, 30(5), 2464-2473.

Ouyang, Y., et al. 2019. Classification of benign and malignant breast tumors using h-scan ultrasound imaging. *Diagnostics, 9*(4), 182.

Qiu, Y., Chang, C. S., Yan, J. L., Ko, L., & Chang, T. S. 2019. Semantic segmentation of intracranial hemorrhages in head CT scans. *In 2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS) (pp. 112-115). IEEE.*

Qiu, Y., Wang, J., Jin, Z., Chen, H., Zhang, M., & Guo, L. 2022. Pose-guided matching based on deep learning for assessing quality of action on rehabilitation training. *Biomedical Signal Processing and Control*, 72, 103323.

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* (pp. 4510-4520).

Whang, J. S., et al. 2013. The causes of medical malpractice suits against radiologists in the United States. *Radiology, 266*(2), 548-554.