**COMP9444 Project Summary**

Image recognition and classification on Cifar-10

1st Yifan SHUAI   
z5254204  
z5254204@ad.unsw.edu.au

2nd Haotian Lyu  
z5283856  
z5283856@ad.unsw.edu.au

3rd Qingrong Xie  
z5263333  
z5263333@ad.unsw.edu.au

4th Jiaxi Liu  
z5320711  
z5320711@ad.unsw.edu.au

5th Buyi Teng  
z5324985  
z5324985@ad.unsw.edu.au

1. **Introduction**

It is well-known that there are plenty of different items in the real life and they are the basis for the normal operation of the world. With the constant renewal of human needs, such as vehicle access identification management, security check or multiple object tracking [1], image recognition and classification technology have already played an important role in daily life. Compared to traditional CNN models, a classic deep learning network “ResidualNet” [2] which created by He et al. has already been widely used in image classification and other fields now. It is usually used as default architectures or baselines in daily research nowadays. In this report, we are trying to build a deep residual learning network Resnet18 to solve a universal problem in the world - Object classification (identify and classify many different kinds of items in real life based on Cifar10 dataset), with hope that this model could achieve a higher accuracy so that it could play a role in related classification issues in future. For instance, with a series of adjustments to the model, it could reach 85.8% top-1and 83.6% top-5 accuracy at resolution 224224 without any other extra dataset.

1. **Methods**

*1.Resnet18*

Resnet18 is a deep convolutional network which has pretty good performance in image classification with 18 convolutional layers deep. It is an excellent model to solve the gradient vanishing issue when the number of hidden layers greater than a certain limit, since Resnet could overlay some shortcut connection inside different residual blocks without reduce the accuracy after training the residual tend to 0.

*2. ReLU Convolutional layers*

Compare with traditional CNN models, Resnet does not need to do the feature extraction manually, and the ReLU activation function has lower computational complexity since it does not have exponential level operation during calculate the value of each nerve cell and has better convergence rate compared to algorithm such as sigmoid or tanh which makes the gradient would be saturated.

*3. Batch normalization*

Since the cifar10 dataset are all in a lower resolution which means there will be many noise data would influence the model performance. After use batch normalization could help to reduce the internal data sample shift and eliminate the adverse effects of singular samples in cifar10.

*4. Data augmentation*

For the data augmentation, we have used some image transformation algorithm such as random rotation in 15 degrees, horizontal random flips and random resize crop (RRC) which are general algorithms used in GoogleNet[3] to avoid overfitting and improve the generalization ability of our model. Consider about the oversampling[4], we just apply these augmentations to the train set split out from the Cifar-10.

1. **Experimental Setup**

The dataset we use for this project is Cifar10 (<URL:https://www.cs.toronto.edu/~kriz/cifar.html> ) which has 60000 3232 resolution RGB color images in total. It has 10 different classes which include airplane, automobile and so on. The number of training set and validation set split from the Cifar10 in ration 8:2. As mentioned above, we have to resize these images since this dataset images are all in a low resolution and the number of images is pretty large with complete independent images (completely different object ratio and feature) in 10 classes which makes a bit hard to classify images through some linear model such as Softmax. Hence, we need to adjust some hyperparameter we used in our model. We choose 0.01 learning rate with 36 batch size to train our model in 70 epochs after a series of accuracy tests. For the optimizer, Adam is used along with 0.9 momentum since it is a commonly used value in Residual network.

1. **Results**

|  |  |  |  |
| --- | --- | --- | --- |
| Method | Epoch | Loss | Test Accuracu Top5 |
| Resnet18 - 1(Lr=0.01) | 50 | 0.0051 | 0.836 |
| Resnet18 - 2(Lr=0.1) | 50 | 0.0436 | 0.779 |
| Resnet18 (Diff data aumentation) | 70 | 0.0031 | 0.749 |
| AlexNet | 50 | 0.7734 | 0.691 |

As we can see, for the image classification problem running on Cifar10, we use Resnet18 model to achieve 0.0051 of loss value, 83.6% of Top5 accuracy and 0.8782 of F1-score under 0.01 learning rate. For this program, we perform some experiments to adjust the learning rate to make the performance better. As what the tables shows above, when the Resnet18 model was trained in a higher value of learning rate, the accucacy of model decreased on the contrary which means model which means the model might become overfit. And when try it on different data augmentation such as Colorjitter and RandomRatation, the model gives a lower accuracy while doing image classification since the new augmention lead to the data distribution offset which cause a worse performance.

After a serise adjustment of the Resnet18 model, we use a traditional CNN network model Alexnet as the comparison of our model. The result of AlexNet only gives 69.1% and 0.7 loss which is much lower than Resnet18. Hence, we found that with the problem of gradient exploding and vanishing, a better performance and results would be presented if a higher and deeper convolutional layer used while the model is used is the real-word application. The neural network strcture of Alexnet only has 5 convolutional layers with 3 fully connected layers(maxpool) processed after dropout algorithm. Compare with the Resnet18 we built, a deeper and complex convolutinal layers applied, the dropout has not been choosen since we have used the batch nominalization, use dropout and BN at a same network layer will cause variance shift while network switching. [5]

From the benchmark of CIFAR-10 image classification, the Top1 accuracy produced by resnet18 model we built can be ranked in about TOP180, which is a good performance since we cannot apply a deeper convolutional layer into our model under the limitaion of hashrate we got.

1. **Conclusions**

This project was mainly focus on solving the image identification and classifications, which are playing more and more important role in our society, and we were trying to find a good solution on the cifar10 dataset. The cifar10 dataset consists of 60000 colour images in 10 classes with 10 common things in daily life, we all thought this is a good dataset for us to compare and find out the better way to classify the images. After comparing with Alexnet and Resnet18, we found out that when we keep increasing the number of layers, we will meet the degradation problem, which is that instead of continuously increasing, the accuracy of the neutal network tends to saturate and eventually decline. And the residual neural network uses the residual connection to skip the training of few layers to avoid the degradation problem to build the deeper network, so that we can get the better accuracy on the dataset.

For the limitation, because of the limit of hash rate, we unable to try the bigger batch size, the max batch size we have tried is 32, we can try bigger batch size such as 128 to see if we can get a better result. Also, the epoch size is another limitation, since even we use cuda, the time we cost for each epoch would be around 5 minutes, this made us have no enough time to try the training with more epochs. Since the limitaion of hash rate and time, we could not try the deeper residual neural network such as resnet50 or optimize a better network structure of Resnet18. To improve our study, if we have no time limit and higher hash rate, we can try the deeper residual neural network, more batch size and more epoch size. Trying more different ways of data augmentation is another way to improve our study such as Mixup algorithm, we only tried several ways of data augmentation, there may be another combination that can make the study better.

Reference:

1. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Deep Residual Learning for Image Recognition, Cornell University, Dec 10, 2015 https://doi.org/10.48550/arXiv.1512.03385
2. D. Marr, “Vision: A Computational Investigation into the Human Representation and Processing of Visual Information’’, Freeman, San Francisco (1982).
3. C. Szegedy, Wei Liu, Yangqing Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Van- houcke, and A. Rabinovich. Going deeper with convolutions. In *Conference on Computer Vision and Pattern Recognition*, 2015.
4. Barandela, R., Sanchez, J, Garcia, V, Rangel, E. Strategies for learning in class imbalance problems, Pattern Recognition 36, pp. 849-851, 2003.
5. X. Li, S. Chen, X. Hu and J. Yang, "Understanding the Disharmony Between Dropout and Batch Normalization by Variance Shift," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 2677-2685, doi: 10.1109/CVPR.2019.00279.