**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 xxx% top-1 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 compare 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.

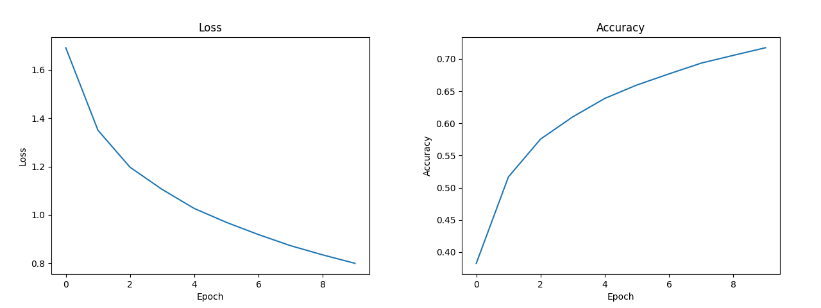
1. **Experimental Setup**

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 (complete 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 common used value in Residual network.

1. **Results**

<Provide key experimental results from your experiment(s). Describe the main findings from the evaluation. Provide analysis whether the findings are good enough to be deployed in real-world application(s). Describe how the proposed solution is different from other proposed solution(s) based on the literature. Where your proposed solution stands in terms of standard evaluation compared to existing state-of-the-art.>

Linear convolutional neural networks are the most common neural networks for artificial intelligence. It refers to the use of convolutional computation in the convolutional layers and subsampling or pooling layers to highlight the most representative features, followed by classification in the fully connected layer to achieve image recognition. This looks like a good solution to complete cifar10. In this case, we got some results with LeNet-5, which is appeared on the lecture notes.



Obviously, this is a very low accuracy rate, which is no more than 75%. Naturally, we want to add more convolutional layers and subsampling or pooling layers, because theoretically the deeper the network, the better the results should be. In fact, after our attempts, more convolutional layers added to LeNet-5 would make it even worse result, because of the gradient disappearance and gradient explosion.

To avoid weight disappearance and weight explosion, common methods are batch normalization, using ResNets, using DenseNets, Xavier and Pre-training. After discussion, we decided to use ResNet18, the batch normalization, and data augmentation to retraining the model.

First, (only use ResNet18)

Second, (plus Batch Normalization)

Third, (plus data augmentation)

(Discuss the latest algorithm)

1. **Conclusions**

<Describe contribution(s) in the project. What are the key strength(s) of the proposed solution? Describe any limitation(s) of your current study and how it can be improved given more time for the project.>

**Note:** Your project summary should be in at most 2 pages (excluding references). There is no limit on the number of references. Any main text beyond the 2-page limit will be ignored by the evaluator(s).

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.