MATH6380P Project 2

Dissecting CNNs: A Scrutinisation Of Their Generalisation Ability

DENG, Yizhe, HUANG Yifei, SUN Jiaze, TAN Haiyi

Department of Mathematics, HKUST

Introduction

Convolutional Neural Networks (CNNs) are highly over-parametrised models, yet counter-intuitively, they do not suffer from the problem of overfitting. In this project, we aim to address the following problems:

I. **Are CNNs able to fit randomly permuted data?**

II. **Does more parameters mean worse overfitting?**

We will also provide our own explanation.

Data

We performed our experiments on the CIFAR-10, a 10-class data set of 32x32 RGB images, which consists of 50,000 training examples and 10,000 testing ones.

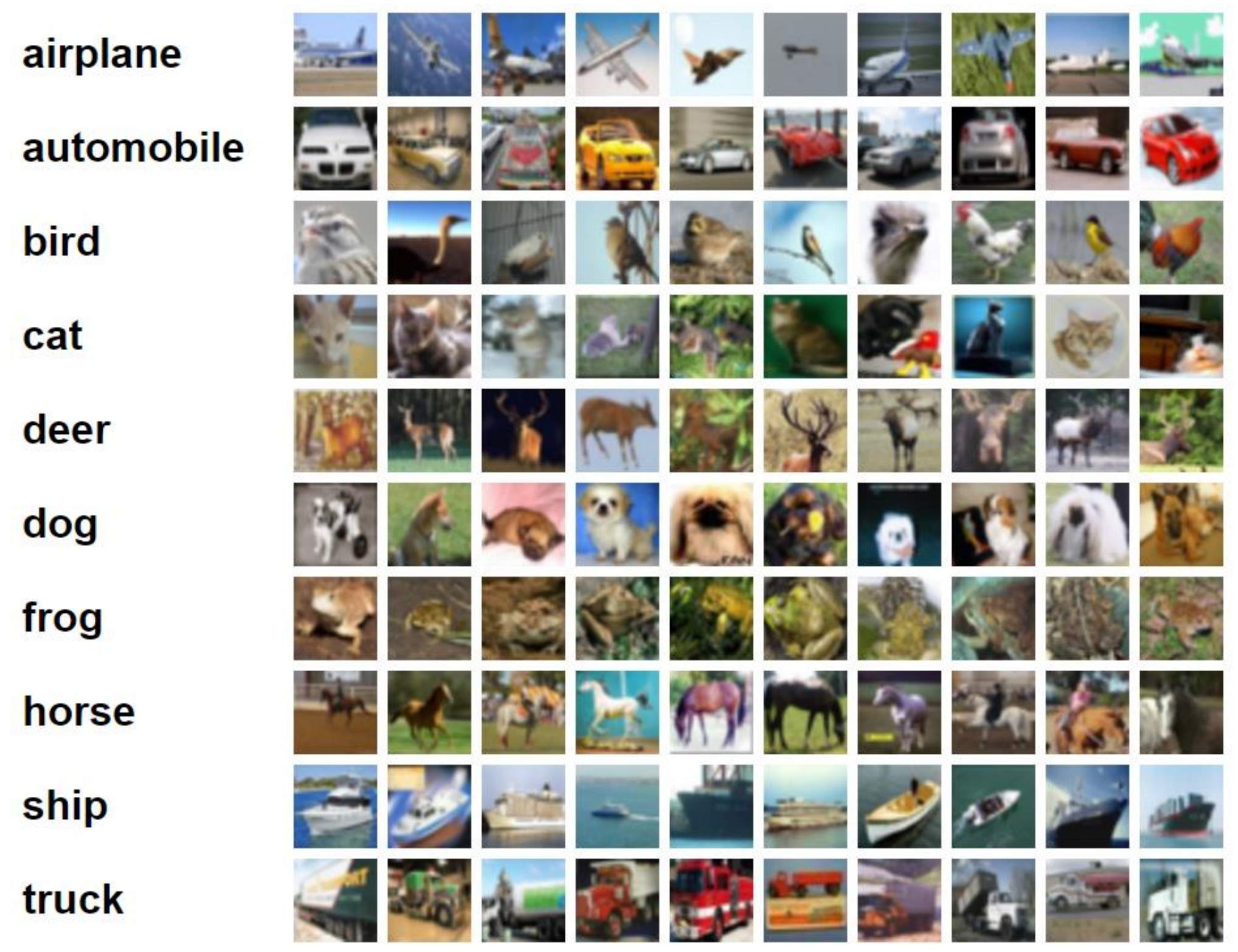

Figure 1. Examples from CIFAR-10.

Methodology

I. For the first task, we may permute the data by randomly shuffling:
   - the **labels** of the whole or part of the training set
   - the **pixels of each image** in the training set

For this part of the experiment, we tested well-known CNNs, including **AlexNet, VGG-16,** and **Resnet-18.**

II. For the second task, we used our own specially designed CNN. It has **5 convolutional layers,** where each one is equipped with
   - **3x3 filters of stride 2**
   - **Batch normalization + ReLU activation.**

We control the number of parameters by **increasing or decreasing the number of filters** in each layer.

I. Training CNN On Randomly Permuted Data

In this experiment, we trained the models using SGD with a learning rate of 0.1 and a mini-batch size of 128.

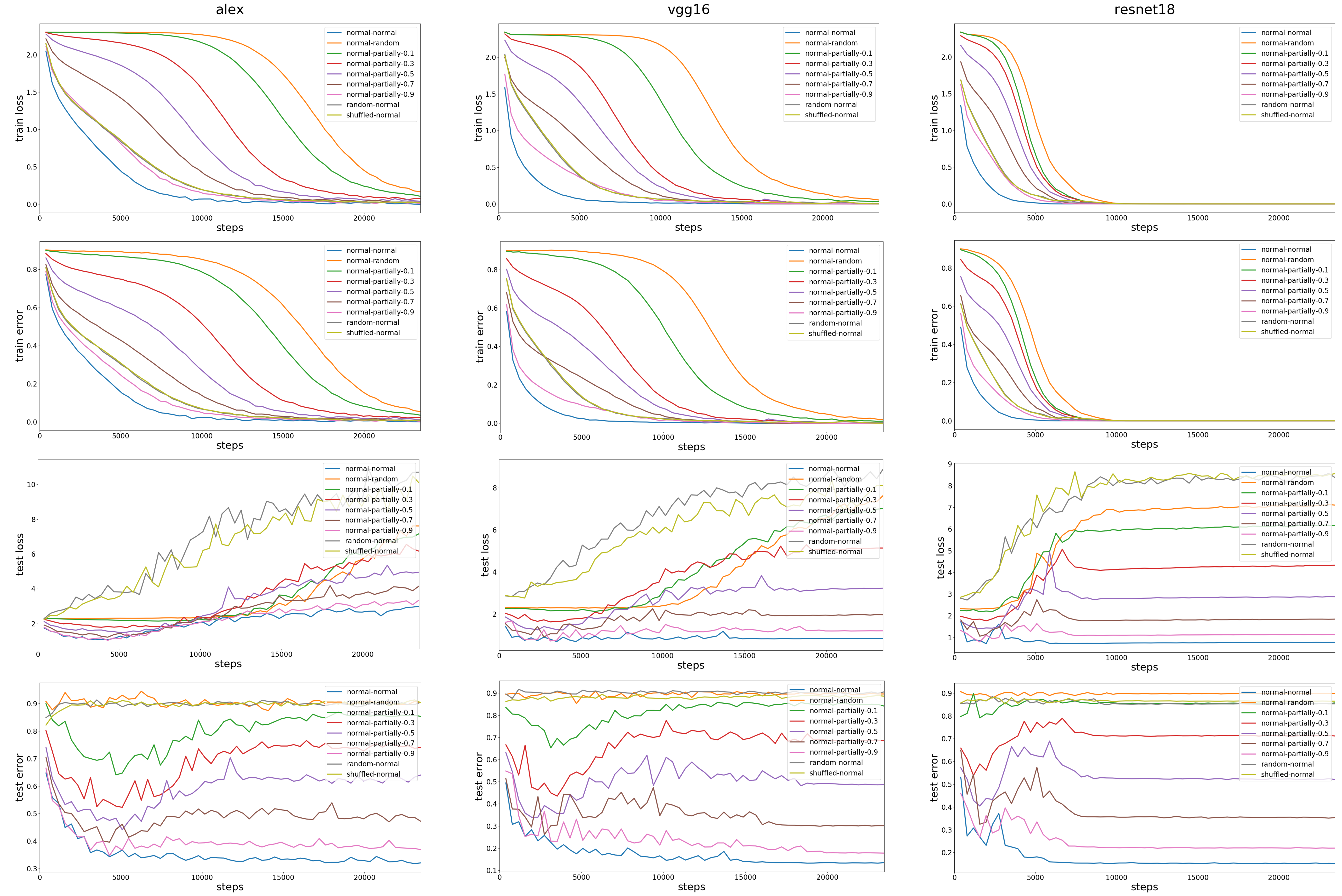

Figure 2. Results of the first task. The legends are in the '***A-B-C***' format, where '***A***' and '***B***' respectively indicate how the training **images** and **labels** are shuffled: '***normal***' means no shuffling; '***partially***' means only some training examples are shuffled, and the number '***C***' indicates the proportion of unshuffled examples. Lastly, ***A*** = '***shuffled***' or '***random***' both mean pixel-wise shuffling, respectively indicating the same or different permutation for different images. ***B*** = '***random***' means all the labels are randomly shuffled.

II. Number Of Parameters vs Overfitting

Using the same learning rate and batch size as in the first task, we trained our CNN 42 times, each time with a different number of filters in each layer. For instance, the first time we used only 3 filters for each layer, and the last time we used 350.

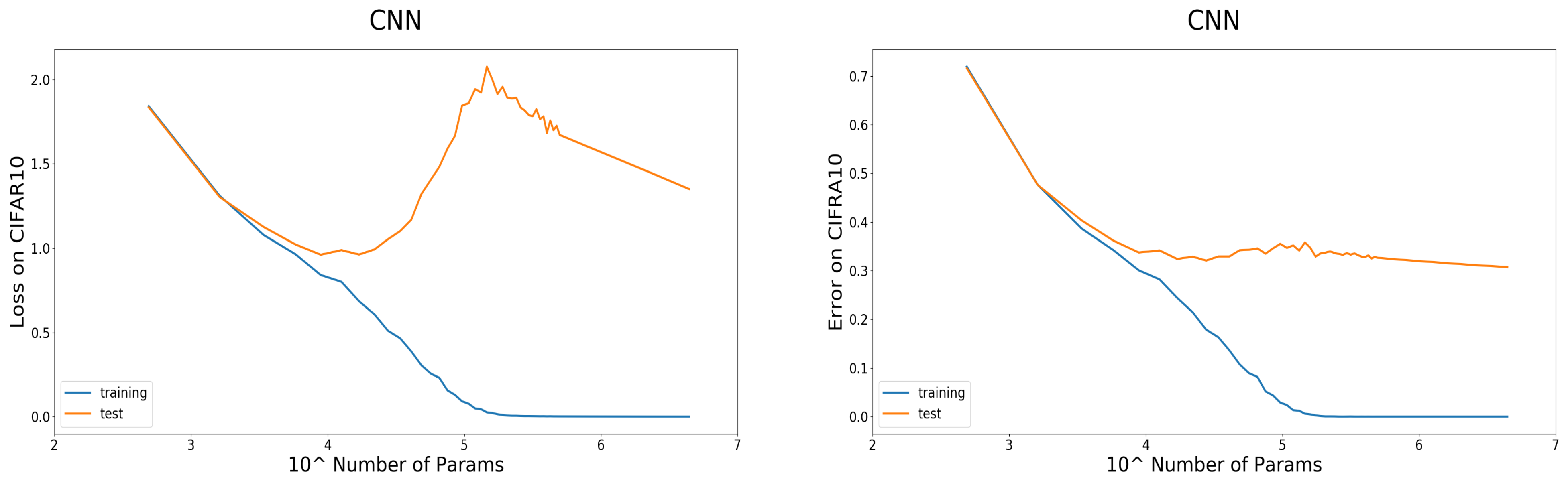

Figure 3. Results of the second objective.

Results Summary and Discussion

I. CNNs can easily fit random data.

Observation: CNNs were able to achieve zero training loss and error on random data, but as a result, performed badly at test time.

Possible explanation: Large CNNs indeed have sufficient capacities to 'memorise' the entire training set. A random transformation on the data merely complicates the landscape of the loss function, thus prolonging training. It is unsurprising that test accuracies were poor, as fitting random data certainly contributed nothing to testing.

B. More parameters causes overfitting, but ONLY in loss

Observation: As the number of parameters increased, test loss showed severe overfitting, whilst test error did not.

Possible explanation: It is not surprising that loss exhibited overfitting, the question is why the accuracy did not. Figure 4 shows that around 60% of test examples were consistently classified correctly regardless of the number of parameters in the model, which is why the accuracy does not drop significantly despite increasing the model complexity dramatically. However, it remans to be ascertained whether this is specific to the data set CIFAR-10 or a general phenomenon for CNNs, and this could be a potential direction for future research.

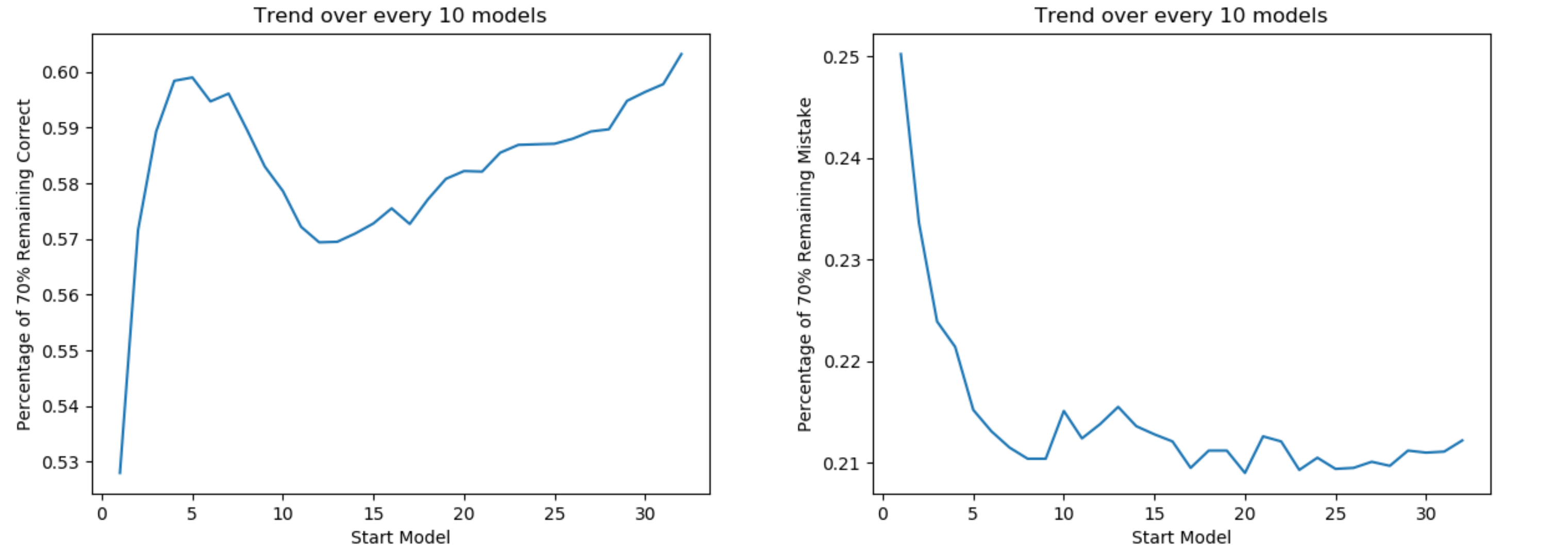

Figure 4. The proportion of examples that were classified correctly (left) or wrongly (right) 7 times out of 10 contiguous models of increasing complexity. In other words, we move a size 10 window over 42 models in the ascending order of complexity, which is why the horizontal axes run from 0 to 33.

Acknowledgement

- **DENG Yizhe**: Analyzing the results
- **HUANG Yifei**: Running the models and analyzing the results
- **SUN Jiaze**: Analyzing the results and writing the poster
- **TAN Haiyi:** Analyzing the results

References

- Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals. *Understanding deep learning requires rethinking generalization*. Nov 10, 2016.
- Tomaso Poggio, K. Kawaguchi, Q. Liao, B. Miranda, L. Rosasco, X. Biox, J. Hidary, and H. Mhaskar. *Theory of Deep Learning III: the non-overfitting puzzle*. Jan 30, 2018.