Zheng Xinyi, 02441303

## 1. CNN introduction

**Task 1.1. Classification** In the classification task, varying CNN architectures yielded distinct accuracy levels, as depicted in the bar graph (Figure 1). The initial model, comprising solely fully connected layers, achieved a validation accuracy of approximately 0.47. Introducing a single convolution layer improved accuracy to around 0.56. Further enhancements involved constructing a complex network (structure detailed below), which significantly increased the accuracy to 0.704. This progression underscores the impact of network complexity on classification accuracy.

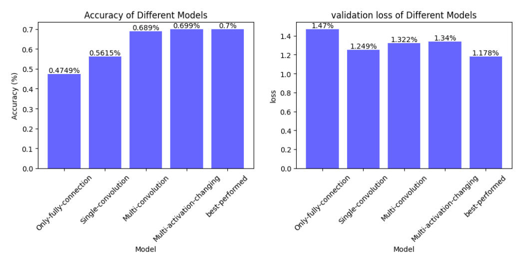


Figure 1. T1.1. Comparison of loss or performance vs iterations.

In the study, I evaluated the impact of various modifications on the best-performing CNN architecture, which initially achieved an accuracy of 0.70. Altering the activation function from ”relu” to ”softmax” in this network resulted in a marginal accuracy decrease to 0.69, as illustrated in the accompanying figure.

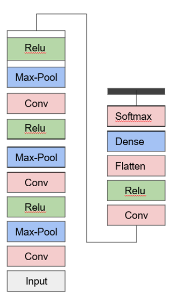


Figure 2. T1.1. Best preformed network structure for classification.

After adjusting the pooling size in the optimal CNN architecture, accuracy slightly decreased to 0.68, indicating a minor impact of this parameter. The experiments emphasize the convolution layer structure as the key determinant of performance. The most effective configuration, achieving a validation accuracy of 1.178, is shown in Figure 2.

**Task 1.2. Regression** When using the pictures from kitchen, the graphs of training and validation loss vs the number of training epochs are as follows. In these pictures, we can see that the training and validation loss both have a rapid decreasing at the beginning then slowly reach the minimum. The best preformed network structure and the result graph are as below:

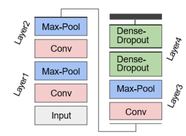


Figure 3. T1.2. Best preformed network structure for regression.

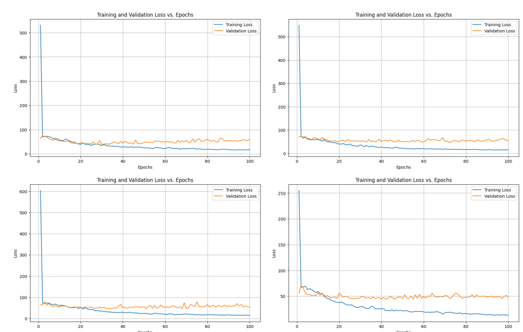

Figure 4. T1.2. Comparison of loss or performance vs iterations of different architectures.

As shown in the above pictures, the last architecture was the best performances one. The discrepancies between training and validation errors in our graphs indicate potential over-fitting, where the model performs better on familiar training data than on new validation data. This often stems from issues like limited dataset variety, complex model structures, and inadequate regularization. Nonetheless, the close and stabilizing training and validation losses suggest effective generalization. Techniques like cross-validation and data augmentation enhance this. Further tests with bedroom, bathroom, and kitchen datasets confirm the model’s generalization capabilities across diverse data types. Using the best architecture, the results are as follow table 1.

| Dataset | Estimation Error (%) |
|---|---|
| frontal | 53.7 |
| bedroom | 60.73 |
| bathroom | 55.01 |
| kitchen | 48.1 |

Table 1. T1.2 Estimation error of different dataset using the best network structure.