**Supervised and unsupervised training report**

This report is divided into 4 parts as below.

1. Task

This part describes the target of task and database used.

1. Solution.

This part describes solution for the task. And also it is divided into 3 parts further.

2-1. Network Structure, 2-2. Autoencoder Part, 2-3. Classification Part.

3.Performance&Analysis

This part describes whole precision and precision of each classification of network .

1. **Task**

Design a network that combines supervised and unsupervised architectures in one model to classify Cifar-10 data set under the condition that 50% of the following classes(bird, deer, truck) and any percentage of the other classes.

Cifar-10 is a database with 50000 train data, 10000 test data for recognizing following classes: airplane, automobile, bird, cat, deer, dog, frog, horse, ship, trunk. Each data is a 32\*32 pixel RGB picture.

1. **Solution**

2-1.Net work structure

According to the requirement that the network combines autoencoder and CNN, following structure is designed.

Autoencoder consist of encoder part and decoder part, doing the job of pre-training of CNN by linking encoder result to CNN input. Encoder part consist 2 convolution layers and 1 pooling layer, decoder part consist of 1 unpooling layer and 2 unconvolution layers to reconstruct image..

CNN consist of 3 convolution layers and 3 full-connection layer. Each convolution layer is consist of convolution, pooling and local response normalization. Each full-connection layer is consist of l2 normalization and dropout.

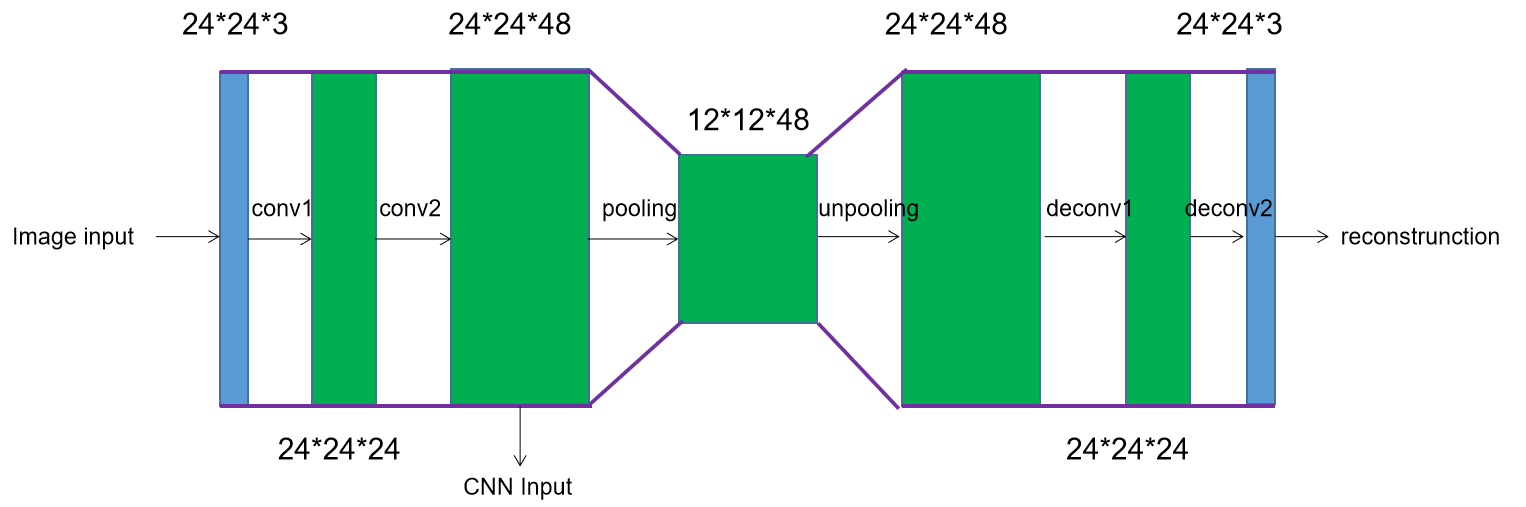


Figure 1. Autoencode structure

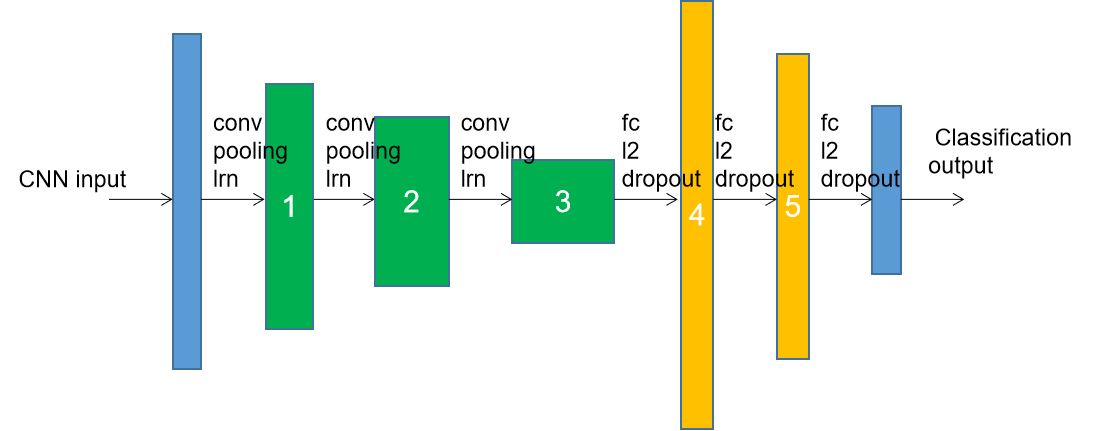


Figure 2. CNN structure

2-2. Autocoder Part

Two kinds of autoencoder is considered, and figure 3 and figure 4 shows that Autoencoder with CNN is clearly perform better than the one without CNN.

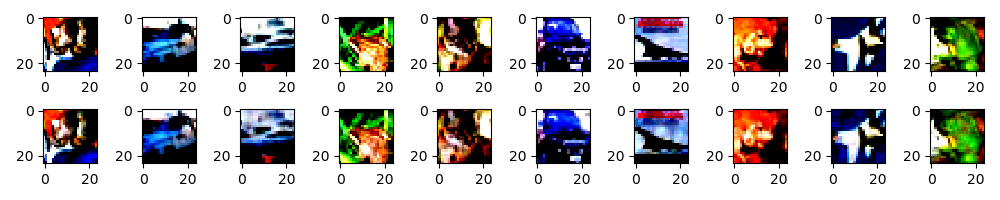


Figure 3. Autoencoder with CNN(1st row: original image, 2nd row: Reconstruction image)



Figure 4. Autoencoder without CNN(1st row: original image, 2nd row: Reconstruction image)

Parameter setting:

(1).Learning rate: 0.01. (Even though, training loss is much smaller while setting to 0.001, but there is no significant difference found between 0.01 and 0.001 on final classification precision.)

(2).Convolution cell: 3x3. (3x3 perform better than 5\*5, which give 4%-precision up)

(3).Training step: 3000.(Training loss shows no significant restrain after 2500 step, while batch size is set to 200)

2-3.Classification Part

Classification part is designed to consist of 3 convolution hidden layer and 2 full connection hidden layer. Local response normalization in convolution layer, L2 normalization and dropout in full connection layer is designed to overcome overfitting.

Parameter setting:

(1).Learning rate:0.001.(a 5%-precision-down result is showed while setting to 0.0001 than 0.001. )

(2).Convolution cell: 3x3.

(3).Training step: 10000.(a 3%-precision-up result is showed while setting to 10000 than 5000, while batch size is set to 200.)

(4).Hidden layer number:(As common sense that adding layer could perform better, however, no significant improve is showed by setting 3 or 4 convolution layers )

**3.Performance & Analysis**

Pre-condition:

Training data: 50% of all classes is used. Test data: 100% of all classes is used.

Data augmentation by randomly crop, flip, brightness, contrast is used.

Precision:

(1) Precision for all classes : 82.4%

(2) Precision of each class :

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Airplane | automobile | bird | cat | deer | dog | frog | horse | ship | trunk |
| 85.7% | 90.1% | 69.4% | 69.9% | 83.9% | 71.1% | 90.5% | 83.7% | 90.4% | 89.3% |

Graph 1. Precision of each class.

Analysis:

About autoencoder part, even though changing parameter(for example, learning rate, training step), no significant precision up could see.

About classification part, increasing batch size and training step could obviously improve precision, changing convolution cell from 5\*5 to smaller one 3\*3 could improve precision, which share the same concept with advanced complicated network(e.g. VGG, resnet).

Actually a single CNN classification network without autoencoder part also performs 77% precision while learning steps set to 3000 in the test for comparison, however this network with autoencoder part perform even worse to 74% while learning steps is same to 3000, which probably means autoencoder pre-training has little impact on improving precision on Cifar10 classification according to the above result.