# Appendix

Appendix file documents network architectures and key parameters setting in Experiment part of our paper, which are referred as Preliminary Experiment, Experiment I, Experiment II, Experiment III respectively.

## 1. NETWORK ARCHITECTURE

This section is to show details of network architecture in experiment part.

**Table 1. Network Architecture of Experiment I**

| | input | filter | stride | padding | output | activation/loss function |
|---|---|---|---|---|---|---|
| conv0 | 1x28x28 | 32x1x3x3 | 1 | 1 | 32x28x28 | leakyRelu |
| pool0 | 32x28x28 | max,2x2 | 2 | 1 | 32x14x14 | / |
| conv1 | 32x14x14 | 32x32x3x3 | 1 | 1 | 32x14x14 | leakyRelu |
| pool1 | 32x14x14 | max,2x2 | 2 | 1 | 32x7x7 | / |
| conv2 | 32x7x7 | 16x32x3x3 | 1 | 1 | 16x7x7 | leakyRelu |
| pool2 | 16x7x7 | max,2x2 | 2 | 0 | 16x4x4 | / |
| | | | | | | |
| conv3 | 16x4x4 | 16x16x3x3 | 1 | 1 | 16x4x4 | leakyRelu |
| unsample0 | | | | | 16x8x8 | / |
| conv4 | 16x8x8 | 32x16x3x3 | 1 | 1 | 32x8x8 | leakyRelu |
| unsample1 | 32x8x8 | | | | 32x16x16 | / |
| conv5 | 32x16x16 | 32x32x3x3 | 1 | 0 | 32x14x14 | leakyRelu |
| unsample2 | 32x14x14 | | | | 32x28x28 | / |
| conv6 | 32x28x28 | 1x32x3x3 | 1 | 1 | 1x28x28 | softmax+binary cross entropy |

**Table 2. Network Architecture of Experiment II**

| | input | filter | stride | padding | output | activation/loss function |
|---|---|---|---|---|---|---|
| conv0 | 1x28x28 | 16x1x8x8 | 1 | 1 | 16x23x23 | leakyRelu |
| pool0 | 16x23x23 | max,2x2 | 1 | 0 | 16x22x22 | leakyRelu |
| conv1 | 16x22x22 | 32x16x4x4 | 1 | 1 | 32x21x21 | leakyRelu |
| pool1 | 32x21x21 | max,2x2 | 1 | 0 | 32x20x20 | leakyRelu |
| linear0 | 32x20x20 | | | | 32 | leakyRelu |
| linear1 | 32 | | | | 10 | softmax+cross entropy |

**Table 3. Network Architecture of Experiment III-Reuter5**

| | input | output | activation/loss function |
|---|---|---|---|
| linear0 | 1x1000 | 5 | leakyRelu |
| linear1 | 5 | 5 | softmax+cross entropy |

**Table 4. Network Architecture of Experiment III-Reuter12**

| | input | output | activation/loss function |
|---|---|---|---|
| linear0 | 1x1000 | 20 | leakyRelu |
| linear1 | 20 | 12 | leakyRelu |
| linear2 | 12 | 12 | softmax+cross entropy |

**Table 5. Network Architecture of Preliminary Experiment**

| | input | filter | stride | padding | output | activation/loss function |
|---|---|---|---|---|---|---|
| conv0 | 3x32x32 | 48x3x3x3 | 1 | 1 | 48x32x32 | leakyRelu |
| pool0 | 48x32x32 | max,2x2 | 2 | 1 | 48x16x16 | / |
| conv1 | 48x16x16 | 48x48x3x3 | 1 | 1 | 48x16x16 | leakyRelu |
| pool1 | 48x16x16 | max,2x2 | 2 | 1 | 48x8x8 | / |
| conv2 | 48x8x8 | 48x48x3x3 | 1 | 1 | 48x8x8 | leakyRelu |
| pool2 | 48x8x8 | max,2x2 | 2 | 1 | 48x4x4 | / |
| conv3 | 48x4x4 | 48x48x3x3 | 1 | 1 | 48x4x4 | leakyRelu |
| pool3 | 48x4x4 | max,2x2 | 2 | 1 | 48x2x2 | / |
| linear0 | 48x2x2 | | | | 10 | softmax+cross entropy |

## 2. TRAINING KEY PARAMETERS

This section is to show key parameters in meta-training and retraining process in our paper.

**Table 6. Meta-training Key Parameters**

| | Exp I | Exp II | Exp III-Reuter5 | Exp III-Reuter12 | Preliminary Exp |
|---|---|---|---|---|---|
| Inner Learning Rate | 0.01 | 0.1 | 0.1 | 0.1 | 0.01 |
| Outer Learning Rate | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| Batch Size | 4 | 8 | 8 | 8 | 2 |
| num_class_per_task | 10 | 10 | 5 | 12 | 10 |
| num_support_sample_per_class | 10 | 10 | 1 | 1 | 10 |
| num_query_sample_per_class | 1 | 1 | 1 | 1 | 15 |
| iters_per_epoch | 400 | 400 | 200 | 200 | 500 |
| Total Epochs | 100 | 800 | 100 | 100 | 100 |

**Table 7. Re-training Key Parameters for Experiment I**

| | $\rho = 0.5\%$ | $\rho = 5\%$ | $\rho = 10\%$ |
|---|---|---|---|
| Learning Rate | 0.05 | 0.05 | 0.005 |
| Batch Size | 32 | 32 | 32 |
| Total Epochs | 1 | 1 | 1 |
| Optimizer | Adam | Adam | Adam |

**Table 8. Re-training Key Parameters for Experiment II**

| | $\rho = 0.5\%$ | $\rho = 1\%$ | $\rho = 5\%$ |
|---|---|---|---|
| Learning Rate | 0.05 | 0.15 | 0.15 |
| Batch Size | 200 | 200 | 200 |
| Total Epochs | 30 | 10 | 10 |
| Optimizer | SGD | SGD | SGD |

**Table 9. Re-training Key Parameters for Experiment III and Preliminary**

| | Exp III-Reuter5 | Exp III-Reuter12 | Preliminary Exp |
|---|---|---|---|
| Learning Rate | 0.0001 | 0.005 | 0.001 |
| Batch Size | 16 | 32 | 64 |
| Total Epochs | 50 | 30 | 100 |
| Optimizer | Adam | Adam | SGD |