Selecting Useful Knowledge from Previous Tasks for Future Learning in a Single Network

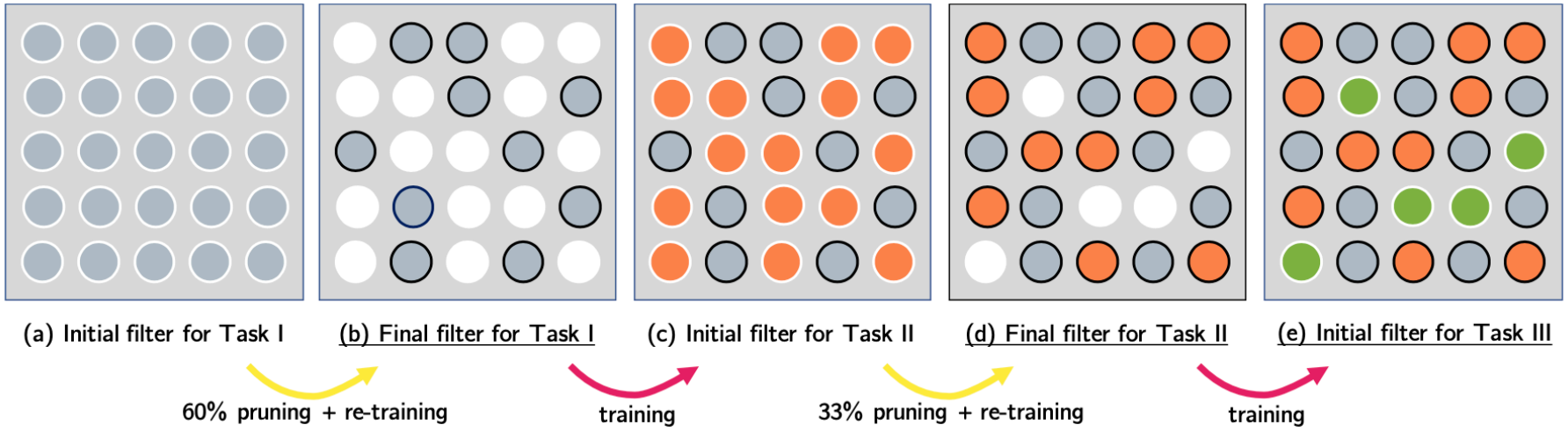
Feifei Shi, Peng Wang, Zhongchao Shi, Yong Rui
Institute of Computing Technology, Chinese Academy of Sciences
University of Chinese Academy of Sciences
AI Lab, Lenovo Research
Lenovo Research

Contents

Introduction

Previous Work

- **PackNet**
  - Iterative pruning and re-training network
  - **Prune** the unimportant weights
  - **Freeze** all the previous weights
  - **Transfer** all the previous weights



Motivation

- **Problem :** PackNet transfers all the previous weights. It does not consider whether weights are helpful for current learning.

- **Our idea:**
  - Important weights for transfer
  - Unimportant weights are masked

Methods

How to Select the Useful Knowledge

- **Distinguish whether the frozen units are important**
  - Units with large gradients hinder current learning
  - Units with small gradients are easy to reuse

- **Gradient-based threshold method**
  - Gradient pruning

$$W_0 = k * (|g|_{max} - |g|_{min}) + b$$

    - k is a coefficient, b is a constant, g represents the gradient
  - Prevent excessive pruning

$$W = max(W_0, W_1)$$

    - W_1 represents a upper bound

Algorithm

Algorithm 1: Training details on the current task

while *training the current task* **do**
Load previous weights into the network F.
$F \leftarrow$ data(X, Y).
Knowledge-Selective mask M_0 is generated according to the calculated threshold W.
Reset network.
Load previous weights into the network F.
F is masked by M_0, which is represented F_{new}.
$F_{new} \leftarrow$ data(X, Y).
Update network.
end

Experiments

Results

CLASSIFICATION ACCURACY ON VGG16.

Dataset	PackNet	Ours	Individual network
CUBS	78.44	78.58	79.78
Stanford Cars	83.09	83.14	86.99
Flowers	89.06	89.04	91.78
Average	83.53	83.59	86.18

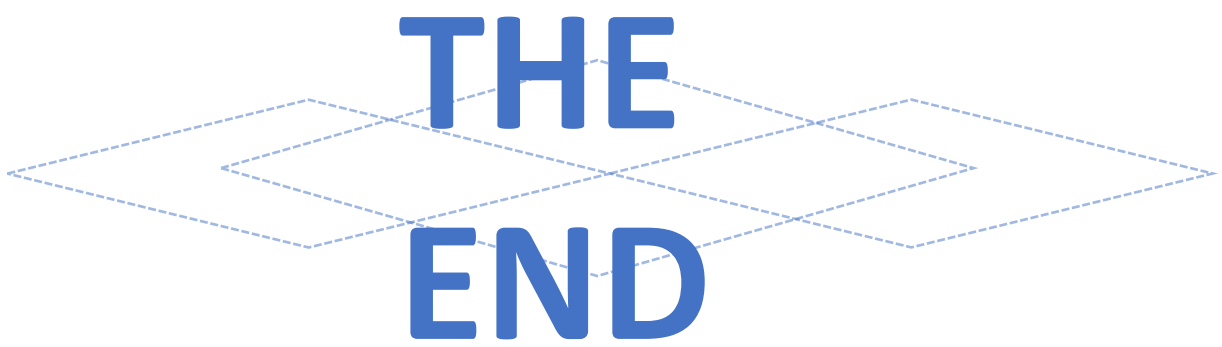
CLASSIFICATION ACCURACY ON SUB DATASETS OF CUBS.

Dataset	PackNet	Ours	Individual network
CUB2	84.01	84.2	84.67
CUB3	83.54	83.64	84.29
CUB4	83.66	83.85	84.95
CUB5	82.8	82.84	83.42
Average	83.5	83.63	84.32

Conclusions

Conclusions

- **Main Conclusions:**
  - **Not all previous weights are helpful for current learning**
  - **Knowledge-selective mask picks suitable knowledge for transferring**
  - **Gradient-based threshold method can make great use of the gradient in current Network**


THE
END