MP.PB.2

胡一鸣 (中科院自动化所) 朱佳刚 (中科院自动化所)
Yiming Hu (CASIA) Jiagang Zhu (CASIA)
孙思洋 (中科院自动化所) 王欣刚 (中科院自动化所)
Siyang Sun (CASIA) Xingang Wang (CASIA)
李建权 (中科院自动化所) 顾庆毅 (中科院自动化所)
Jianquan Li (CASIA) Qingyi Gu (CASIA)

基于多损失感知的神经网络通道剪枝方法

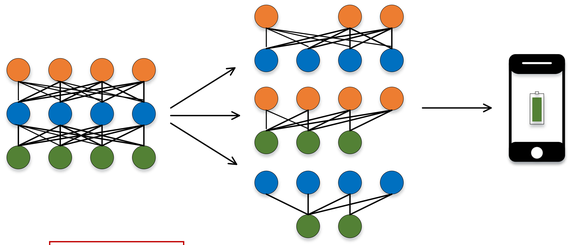
Multi-Loss-Aware Channel Pruning of Deep Networks

○胡一鸣, 孙思洋, 李建权, 朱佳刚, 王欣刚, 顾庆毅（中国科学院自动化研究所）
○ Yiming Hu, Siyang Sun, Jianquan Li, Jiagang Zhu, Xingang Wang, Qingyi Gu (CASIA)

ICIP 2019

CASIA

Background



Objective

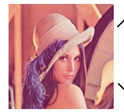
Accelerate the existing neural networks by pruning for efficient deployment.

Multi-Loss-Aware Channel Pruning Algorithm

Train a large model from scratch

Select the most important channels based on the final loss

Fine-tune pruned Network



Baseline Model

Pruned Model

Feature Maps

Before Pruning:

After Pruning:

Reconstruction Error

Feature & Semantic Correlation Loss

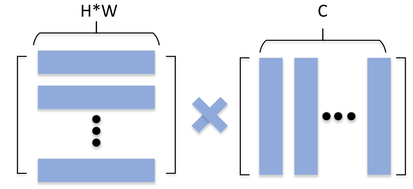
Classification Loss

⊕ Final Loss

Convolution Fully connected

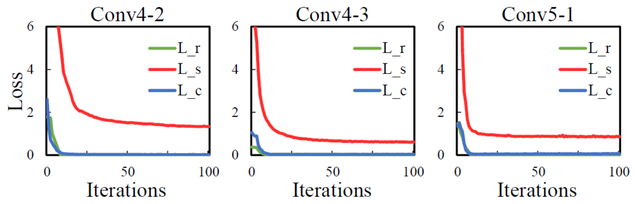
Max pooling Softmax

Feature Correlation



Semantic Correlation

Experimental Result



Methods	Training Err. (%)	Test Err. (%)
$\mathcal{L}_r$	3.75	9.74
$\mathcal{L}_s$	6.53	10.95
$\mathcal{L}_c$	0.85	8.35
$\mathcal{L}_r + \mathcal{L}_s$	1.31	8.27
$\mathcal{L}_r + \mathcal{L}_c$	0.73	8.14
$\mathcal{L}_s + \mathcal{L}_c$	1.03	8.20
$\mathcal{L}_r + \mathcal{L}_s + \mathcal{L}_c$	1.09	8.00

Models	Methods	Baseline(%)	Err. Gap(%)	#Param.	#Param↓	#Flops	#Flops↓
VGG	CP	6.01	+0.32	7.70M	1.92×	155.80M	2.00×
	ThiNet		+0.14				
	Slimming		+0.19				
	DCP		-0.17				
	Ours	6.08	**-0.13**	5.50M	2.26×	140.30M	2.23×
ResNet56	CP	6.20	+1.00	0.42M	1.97×	63.20M	1.99×
	ThiNet		+0.82				
	WM		+0.56				
	DCP		+0.31				
	Ours	6.26	**-0.24**				