简介

作者其实提出了一种训练方法，和一种slimmable的模式，可以切换预先设置好的几个网络模型参数，比如`Mobilenet 0.25 0.5 0.75 1.0`。
作者发现如果直接以共享的方式去训练一个不同宽度的网络，虽然训练集的表现比较正常，但是在验证集上会直接垮掉。作者认为主要原因在于不同网络的mean和varience不同，所以导致BN层失效。
作者还试了一种方法，先训练模型A(MobileNet 0.35x)，然后又加上一些参数B变成MobileNet 0.5x，这个时候将A的 参数固定去finetune B的参数，发现精度上涨的十分有限，因为这种训练模式没有考虑A对B的影响等因素。

方法

作者的解决方案是采用`Swichable Batch Normalization(S-BN)`，其实就是训练模式还是共享训练，但是不同训练参数下BN层采用独立的设置，这样就解决了mean和varience的问题，而且由于可以先缓存grad，所以训练效率也是比较高的。

训练步骤

Algorithm 1 Training slimmable neural network M.

Require: Define *switchable width list* for slimmable network M, for example, $[0.25, 0.5, 0.75, 1.0]\times$.

Initialize shared convolutions and fully-connected layers for slimmable network M.
Initialize independent batch normalization parameters for each *width* in *switchable width list*.
for $i = 1, ..., n_{iters}$ **do**
    Get next mini-batch of data x and label y.
    Clear gradients of weights, $optimizer.zero\_grad()$.
    for *width* in *switchable width list* **do**
        Switch the batch normalization parameters of current width on network M.
        Execute sub-network at current width, $\hat{y} = M'(x)$.
        Compute loss, $loss = criterion(\hat{y}, y)$.
        Compute gradients, $loss.backward()$.
    end for
    Update weights, $optimizer.step()$.
end for

结果分析

S-BN的效果解决了验证机不收敛的问题，而且在MobilnetV1上其slimmable的模型准确率还优于单独训练的模型，作者分析可能是因为这种训练方式会起到模型蒸馏的作用，但是在其他网络上却没有出现这种情况，值得关注

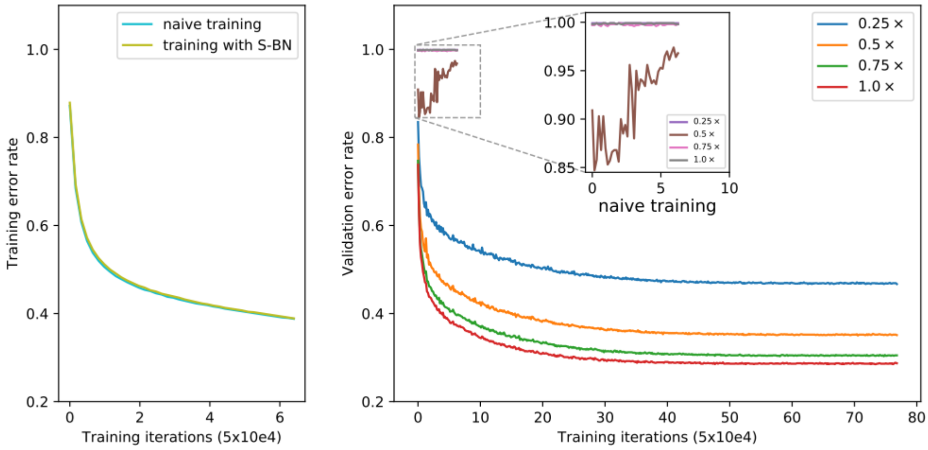

Figure 2: Training and validation curves of slimmable networks. Left shows the training error of the largest switch. Right shows testing errors on validation set with different switches. For naive approach, the training is stable (left) but testing error is high (right, zoomed). Slimmable networks trained with *S-BN* have stable and rank-preserved testing accuracy across all training iterations.

Table 3: Results of ImageNet classification. We show top-1 error rates of individually trained networks and slimmable networks given same width configurations and FLOPs. We use S- to indicate slimmable models, † to denote our reproduced result.

Individual Networks			**Slimmable Networks**			FLOPs
Name	Params	Top-1 Err.	Name	Params	Top-1 Err.	
MobileNet v1 1.0×	4.2M	29.1	S-MobileNet v1 [0.25, 0.5, 0.75, 1.0]×	4.3M	28.5 (0.6)	569M
MobileNet v1 0.75×	2.6M	31.6			30.5 (1.1)	317M
MobileNet v1 0.5×	1.3M	36.7			35.2 (1.5)	150M
MobileNet v1 0.25×	0.5M	50.2			46.9 (3.3)	41M
MobileNet v2 1.0×	3.5M	28.2	S-MobileNet v2 [0.35, 0.5, 0.75, 1.0]×	3.6M	29.5 (-1.3)	301M
MobileNet v2 0.75×	2.6M	30.2			31.1 (-0.9)	209M
MobileNet v2 0.5×	2.0M	34.6			35.6 (-1.0)	97M
MobileNet v2 0.35×	1.7M	39.7			40.3 (-0.6)	59M
ShuffleNet 2.0×	5.4M	26.3	S-ShuffleNet [0.5, 1.0, 2.0]×	5.5M	28.7 (-2.4)	524M
ShuffleNet 1.0×	1.8M	32.6			34.5 (-0.9)	138M
ShuffleNet 0.5×	0.7M	43.2			42.7 (0.5)	38M
ResNet-50 1.0×	25.5M	23.9	S-ResNet-50 [0.25, 0.5, 0.75, 1.0]×	25.6M	24.0 (-0.1)	4.1G
ResNet-50 0.75×†	14.7M	25.3			25.1 (0.2)	2.3G
ResNet-50 0.5×†	6.9M	28.0			27.9 (0.1)	1.1G
ResNet-50 0.25×†	2.0M	36.2			35.0 (1.2)	278M