Aggregated Residual Transformations for Deep Neural Networks-ResNext

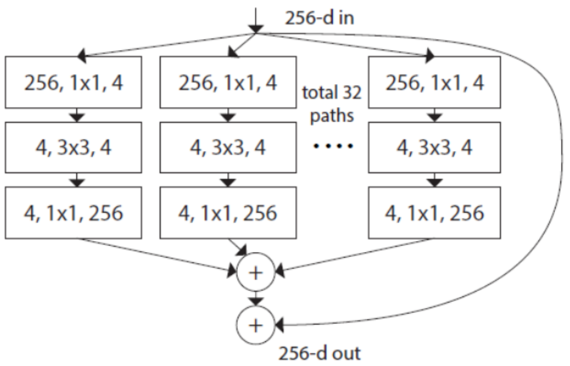
Motivation

We decided to review this architecture because it was an enhancement over the ground breaking ResNet-101/152 architecture. Furthermore, it was among those latest papers which was published by reputed organizations like [1]**UC San Diego** and [2]**Facebook AI Research**.

Introduction

The model name, **ResNeXt**, contains Next. It means the next dimension, on top of the ResNet. This next dimension is called the “cardinality” dimension. This network is constructed by repeating a building block that aggregates a set of transformation with the same topology. Moverover, the paper considers increasing cardinality as an effective way of gaining accuracy than going deeper or wider

Structure



$$y = x + \sum_{i=1}^{C} \mathcal{T}_i(\mathrm{x})$$

- ResNext has multi-branch CNN architecture similar to the Inception net with identical convolution branch instead of custom branch. For each path, **Conv1×1–Conv3×3–Conv1×1** are done at each convolution path. This design is called the bottleneck design. The internal dimension for each path is denoted as **d(d=4)**. The number of paths is the cardinality **C(C=32)**.
- The dimension is increased directly from 4 to 256, and then added together, and also added with the skip connection path.
- It has a highly modularize design following VGG/ResNets. The network consists of stack of residual blocks. These blocks have the same topology, and are subject to two simple rules **(i)** if producing spatial maps of the same size, the blocks share the same hyperparameter(width and filter sizes), and **(ii)** each time when the spatial map is downsampled by a factor of 2, the width of the blocks is multiplied by a factor of 2.

Analysis

- There are 3 main hyper-parameter, 1. Cardinality (group size), 2.Model Depth 3. Base Width
- The model follows the split-transform-merge paradigm.
- It work on splitting, tranforming and aggregating.
- The ResNext block is similar to the inception module except that the outputs of different paths are merged by adding them together, while in Inception they are depth-concatenated.

Experiments

1. **ImageNet-1K :**
   - In the 1000-class Imagenet classification task ResNeXt-50 obtains **22.2%** Top-1 validation error rate, which is **1.7%** less than ResNet baseline's **23.9%**. It also obtains a lower training rate which suggests that on increasing cardinality *the gains are not from regularization but from stronger representation*.
   - Comparision of increase in error (higher increase in ResNet-50 compared to ResNext-50) on removing residual connections suggests that the changes in ResNext have stronger representation as it performs consistently better than their counterparts irrespective of residual connections.
2. **ImageNet-5K :**
   - ResNext-50 obtains a drop of **3.2%** in the **5K-way top-1 error** over ResNet-50, and ResNext-101 obtains a drop of **2.3%** in the **5K-way top-1 error** over ResNet-101.

Strength

- ResNext achieved an increase in accuracy without increasing capacity (going deeper or wider).
- It has much simpler architectural design than all Inception models, and requires considerably fewer hyper-parameter to be set by hand.

Weakness

- Computationally hungry because of introduction of an extra dimension (Cardinality).
- On 8 GPUs of NVIDIA M40, it takes 0.95s per mini-batch vs 0.70s of ResNet-101 baseline that has similar FLOPs.

Scope of Improvement

- Multi-resolution blocks can be used to make it more accurate (idea from the Inception Net).
- Depthwise separable convolutions can be used to reduce the computing power needed for the network to improve its speed.

Paper Reviewed By:

Utkarsh Ranjan
200050147

Mahesh Bhupati
200260027