Squeeze and Excitation Networks

- This paper focused on how to adaptively recalibrate
- channel-wise feature responses to enhance the representational power.
- They proposed a new block which they termed as ‘Squeeze and Excitation block’ or SEblock which can be incorporated in any existing SOTA architecture and can improve performance of that.

Mechanism:

- Let Ftr be the mapping from the input X to the feature maps U.
- Squeeze operation envolves aggregating feature maps across their spatial dimension.
- The function of this descriptor is to produce an embedding of the global distribution of channel-wise feature responses, allowing information from the global receptive field of the network to be used by all its layers.
- The aggregation is followed by an excitation operation, which takes the form of a simple self-gating mechanism that takes the embedding as input and produces a collection of per-channel modulation weights.
- These weights are applied to the feature maps U to generate the output of the SE block which can be fed directly into subsequent layers of the network.
- While the template for the building block is generic, the role it performs at different depths differs throughout the network. Either in a class agnostic manner or class specific manner.

Computational Efficiency

- The SE block adds only a small computational overhead (~0.26% in ResNet-50), making it a lightweight yet effective improvement.
- Despite this minor increase in complexity, SE blocks significantly enhance network performance across various tasks and datasets.

Experimentations

- Extensive experiments on ImageNet demonstrate that SE blocks improve accuracy across all tested models:
  - SE-ResNet-50 reduces top-5 error from 7.48% to 6.62%, while adding a negligible computational burden.
  - SE blocks are shown to outperform deeper networks without SE blocks.
- SE blocks also generalize well to other datasets such as CIFAR-10, CIFAR-100, and the Places365 scene classification dataset.

- The paper tested SE blocks on the Places365-Challenge dataset for scene classification and COCO dataset for object detection.
- SE blocks improved both scene classification and object detection, showing their effectiveness beyond image classification.
- Various experiments were conducted to analyze the effect of different components of the SE block.**Reduction ratio (r)**: A hyperparameter to control model complexity. The default value of r=16 was found to balance performance and complexity.
- **Pooling method**: Global average pooling was preferred over max pooling for the squeeze operation.
- **Non-linear activation**: Sigmoid activation was superior to alternatives like ReLU or Tanh for the excitation step.
- SE blocks consistently improve performance across a range of tasks and architectures, with minimal additional computational cost.SE-ResNet-50 achieves nearly the same performance as ResNet-101 while being computationally much lighter.These results establish SE blocks as a general method to enhance the representational power of neural networks by modeling channel-wise dependencies.