**EfficientNet**: Rethinking Model Scaling for Convolutional Neural Networks

Mingxing Tan, Quoc V. Le
How to Scale Up A ConvNet?

- **(a) baseline**: 
  - #channels
  - layer_i
  - resolution HxW

- **(b) width scaling**: 
  - wider

- **(c) depth scaling**: 
  - deeper
  - higher resolution

- **(d) resolution scaling**: 
  - deeper
  - higher resolution

- **(e) compound scaling**: 
  - wider
Compound Scaling

depth: \( d = \alpha^\phi \)

width: \( w = \beta^\phi \)

resolution: \( r = \gamma^\phi \)

s.t. \( \alpha \cdot \beta^2 \cdot \gamma^2 \approx 2 \)

\( \alpha \geq 1, \beta \geq 1, \gamma \geq 1 \)

Step1:
- First fix \( \phi = 2 \), and find \( \alpha, \beta, \gamma \) with local search.

Step2:
- Then fix \( \alpha, \beta, \gamma \), and scale the network with different \( \phi \).

EfficientNet-B0: A New Baseline Network Found by AutoML

Simple, clean, no branches irregularity in layer types

MBConv represents “mobile inverted bottleneck” [See MnasNet for more details]
Scaling the Same Baseline EfficientNet-B0

![Diagram showing the comparison between different scaling methods in terms of FLOPs and ImageNet Top-1 Accuracy. The graph illustrates the performance gain of Compound Scaling over other scaling methods.]
ImageNet Results

Our AutoML + Compound Scaling

Previous AutoML

Hand Crafted
Transfer Learning Results

EfficientNets

Graphs showing Accuracy (%) vs. Number of Parameters (Millions, log-scale) for different datasets:
- CIFAR10
- CIFAR100
- Birdsnap
- Stanford Cars
- Flowers
- FGVC Aircraft
- Oxford-IIIT Pets
- Food-101

Key:
- DenseNet-201
- ResNet-50
- Inception-v1
- ResNet-152
- NASNet-A
- GPIPE
- ResNet-101
- Inception-v3
- DenseNet-121
- EfficientNet
- Inception-ResNet-v2
- DenseNet-169
- Inception-v4