

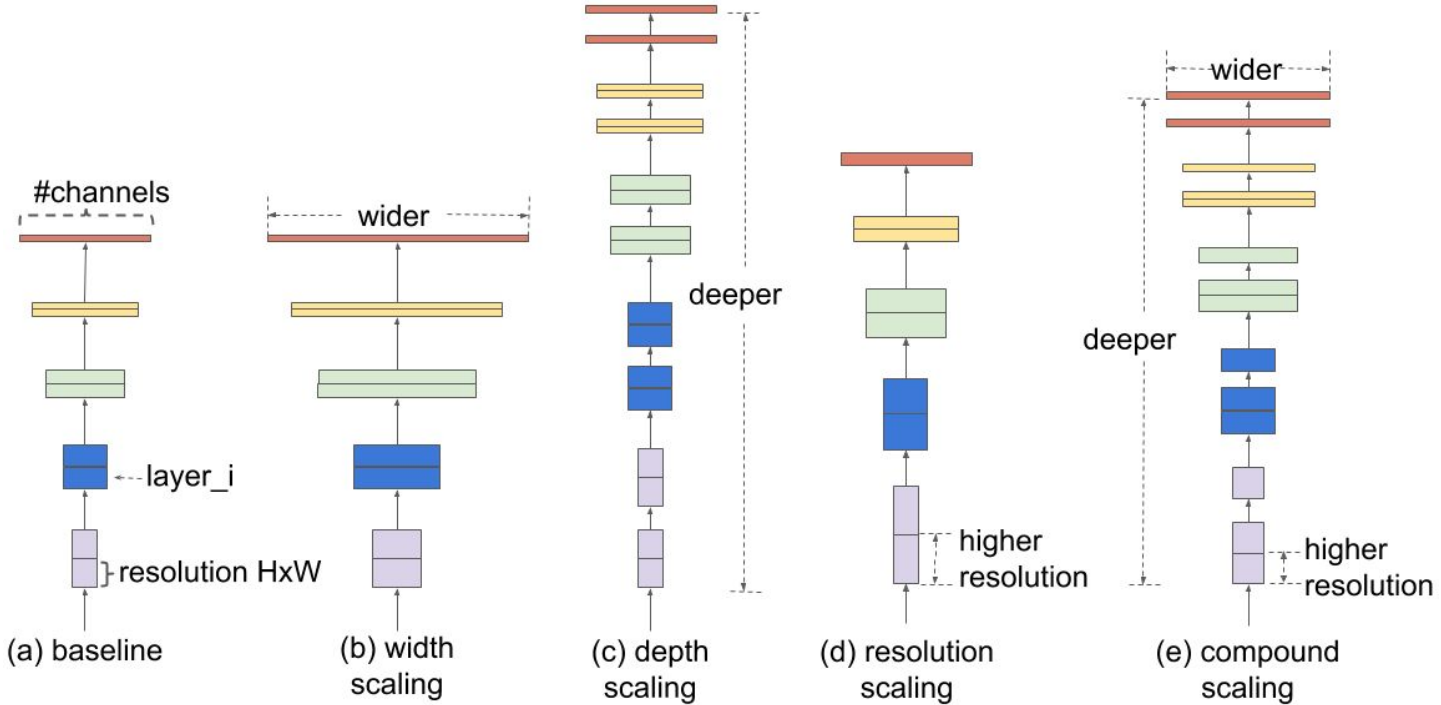
EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks



Mingxing Tan, Quoc V. Le



How to Scale Up A ConvNet?



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:

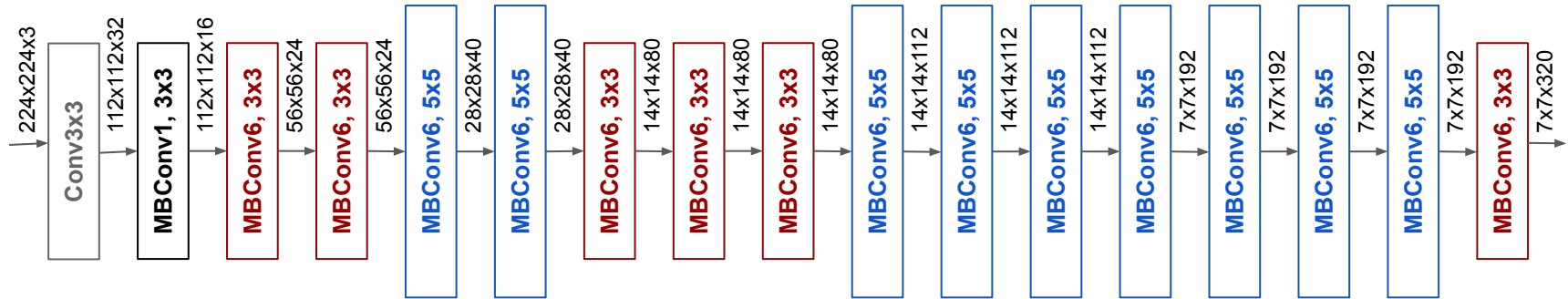
- First fix $\phi = 2$, and find α, β, γ with local search.

Step2:

- Then fix α, β, γ , and scale the network with different ϕ .

Compound scaling improves MobileNetV1, MobileNetV2, and ResNet-50.

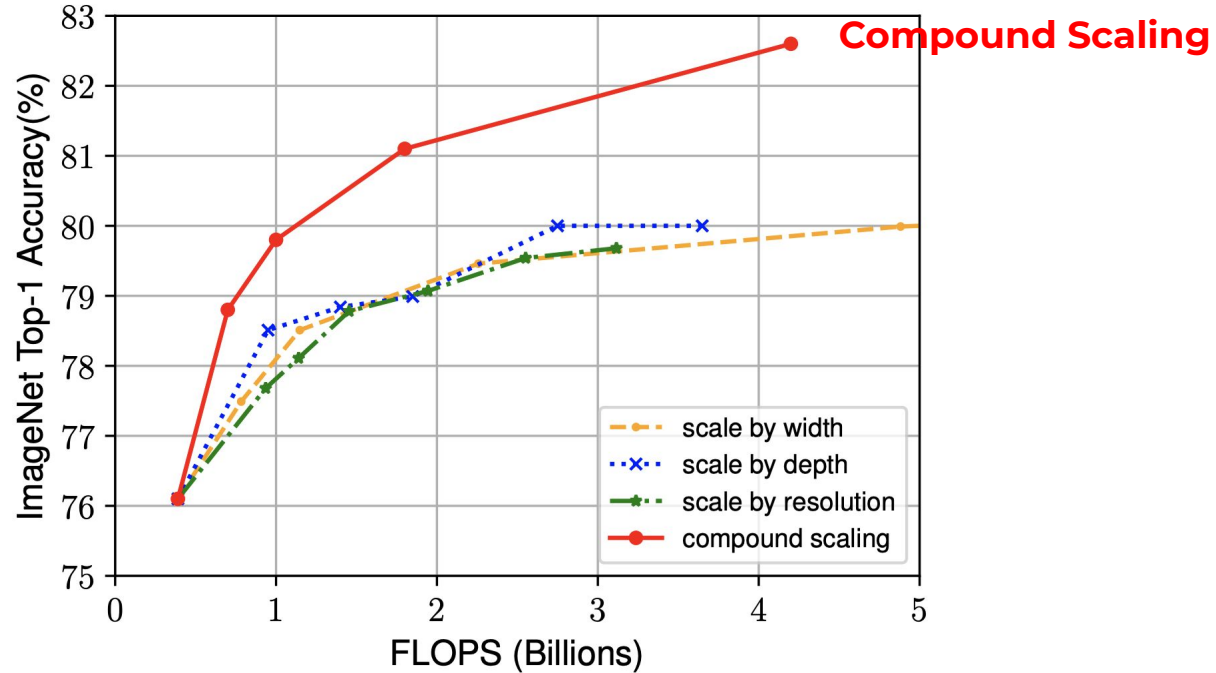
EfficientNet-B0: A New Baseline Network Found by AutoML



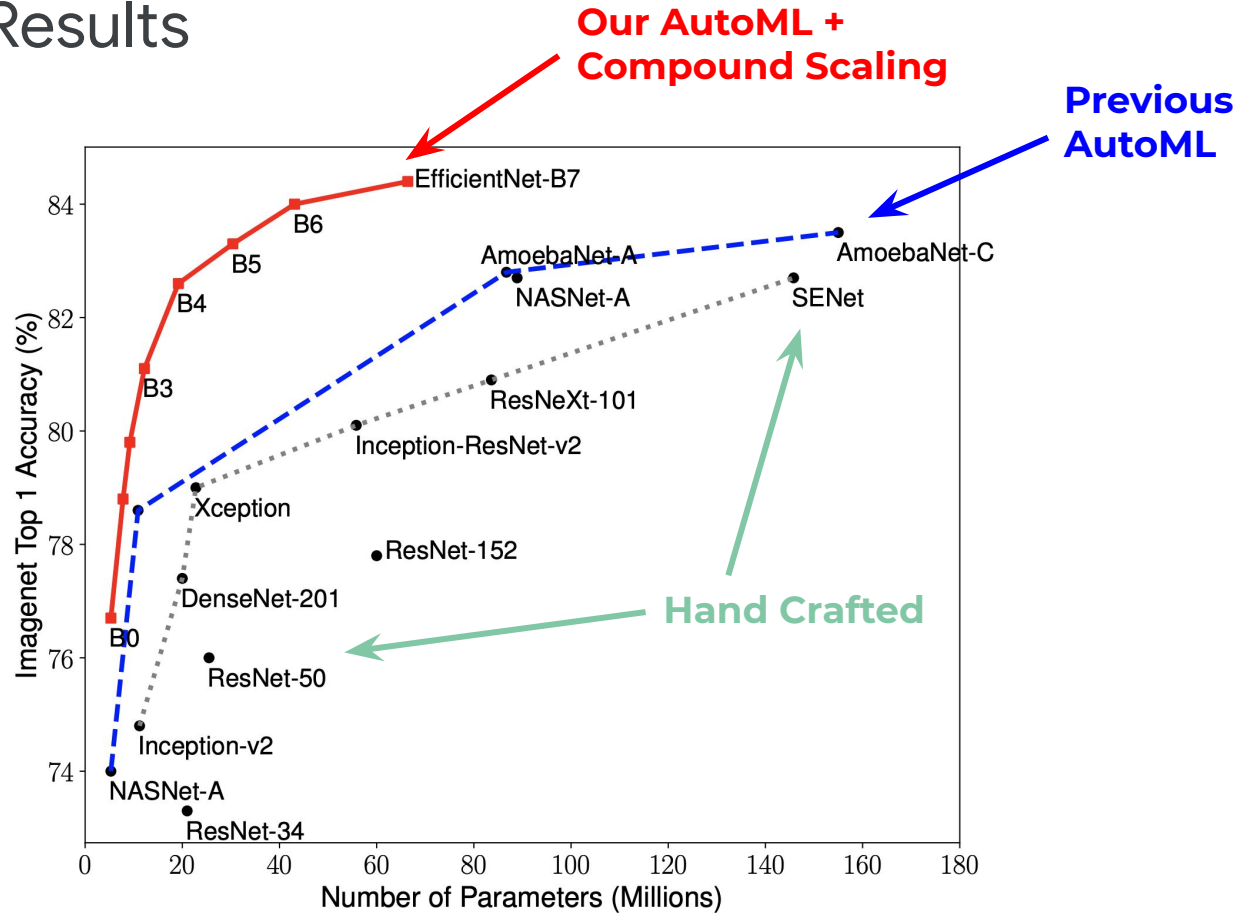
Simple, clean, no branches
irregularity in layer types

MBConv represents "mobile inverted bottleneck" [See [MnasNet](#) for more details]

Scaling the Same Baseline EfficientNet-B0

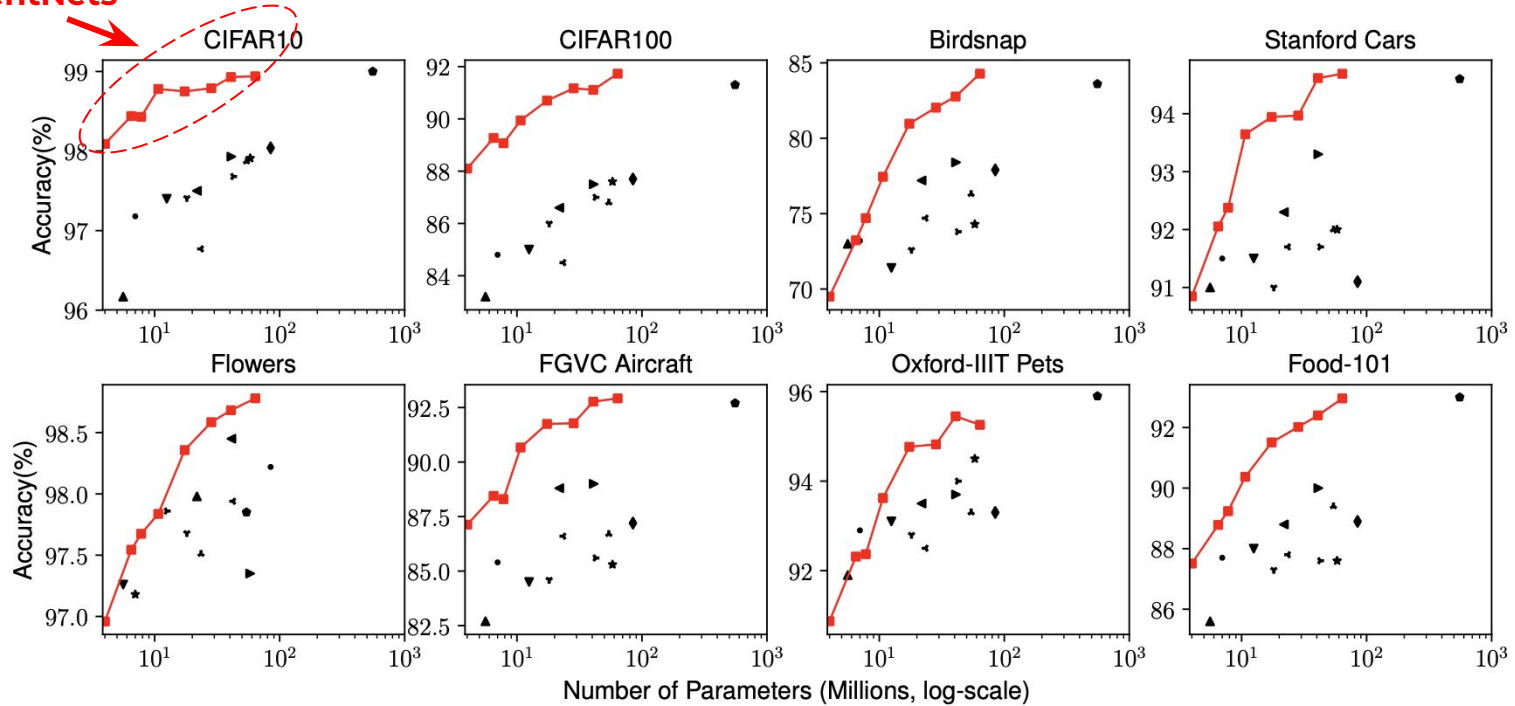


ImageNet Results



Transfer Learning Results

EfficientNets



▼ 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		