EfficientNet: Rethinking Model Scaling for ConvNets

(Mingxing Tan, Quoc V. Le)

Aaron Baruch (amb565) Mohammad Labadi (maa366) Katie Popova (kap272)
Justin Tien-Smith (jzt5) Peter Zheng (pcz3)

Paper Motivation & Background

Traditional ConvNet scaling approaches increase either depth, width, or input resolution individually, but doing so leads to diminishing returns and inefficient use of resources. EfficientNet introduces a compound scaling method that jointly scales these dimensions using a coefficient ϕ and constants α, β, and γ under the constraint:

$$\begin{aligned} \text{depth:} \quad & d = \alpha^{\phi} \\ \text{width:} \quad & w = \beta^{\phi} \\ \text{resolution:} \quad & r = \gamma^{\phi} \\ \text{s.t.} \quad & \alpha \cdot \beta^2 \cdot \gamma^2 \approx 2, \quad \alpha \geq 1, \beta \geq 1, \gamma \geq 1 \end{aligned}$$

Starting from a baseline architecture, EfficientNet-B0 (found via multi-objective neural architecture search), the authors produce a family of models (B1–B7) that achieve high accuracy on ImageNet with significantly fewer parameters and FLOPs. The central contribution is a unified, principled scaling method that consistently improves performance while being computationally efficient.

Chosen Result

We replicate a core experiment from the EfficientNet paper (Tan & Le, 2020), illustrated in Figure 1 and Table 2, which compares various model scaling strategies on EfficientNet-B0. These include single-dimension scaling of depth ($d = 4$), width ($w = 2$), and resolution ($r = 2$), as well as compound scaling ($d = 1.4$, $w = 1.2$, $r = 1.3$). This experiment supports the paper's central claim: that compound scaling achieves better accuracy-efficiency trade-offs than scaling any single axis. We aimed to reproduce this result under different conditions, using CIFAR-10 instead of ImageNet, and fewer training epochs, to assess whether the same qualitative trends hold. Our goal was to compare scaling strategies using relative performance, parameter count, and FLOPs.

Methodology

We re-implemented EfficientNet in PyTorch and evaluated how different scaling strategies affect performance on CIFAR-10, a dataset of 60,000 images across 10 classes. We trained five models: the EfficientNet-B0 baseline; depth-only scaling ($d = 4$); width-only scaling ($w = 2$); resolution-only scaling ($r = 2$); and compound scaling. All models use the same MBConv blocks with squeeze-and-excitation as described in the original B0 architecture. For resolution scaling, CIFAR-10 inputs

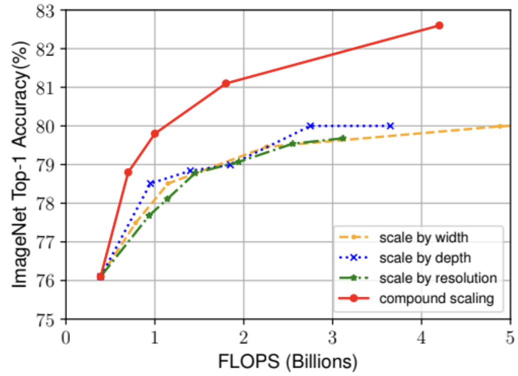

Figure 8. **Scaling Up EfficientNet-B0 with Different Methods.**

Figure 1: Tan & Le, Figure 8

Table 7. **Scaled Models Used in Figure 7.**

Model	FLOPS	Top-1 Acc.
Baseline model (EfficientNet-B0)	0.4B	77.3%
Scale model by depth (d=4)	1.8B	79.0%
Scale model by width (w=2)	1.8B	78.9%
Scale model by resolution (r=2)	1.9B	79.1%
Compound Scale (d=1.4, w=1.2, r=1.3)	**1.8B**	**81.1%**

Figure 2: Tan & Le, Table 7

were upsampled (e.g., to 64×64 for $r = 2$). Compound scaling follows the paper's $\phi = 2$ rule with $\alpha = 1.2$, $\beta = 1.1$, and $\gamma = 1.15$. Training was performed for 35 epochs using the RMSProp optimizer (momentum 0.9, $\alpha = 0.9$, $\epsilon = 0.1$), with a linear increase into exponential decay for our learning rate 3. Our loss function was cross-entropy. Although our compute budget was significantly smaller than in the original paper (which used 350 epochs on ImageNet), our setup was sufficient to examine relative trends in accuracy, parameter count, and FLOPs.

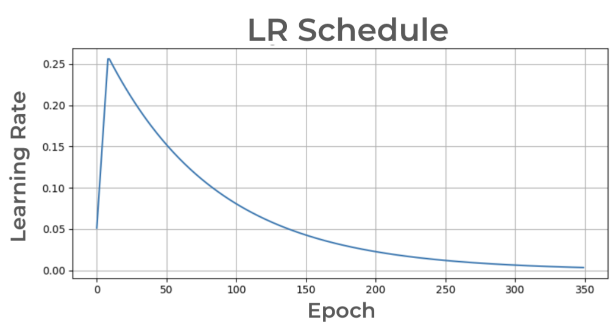

Figure 3: Our Learning Rate Schedule

Results & Analysis

Our re-implementation results are summarized in Figure ?? and Table ??, comparing top-1 accuracy and FLOPs across scaling strategies. While the original paper showed that compound scaling yielded clear accuracy gains over single-dimension scaling, our results struggle to show noticeable improvements. In fact, compound scaling offered only marginal gains over resolution or width-only scaling, and underperformed the baseline in top-1 accuracy, though it maintained a favorable FLOP count. Several factors likely contributed to these discrepancies. First, we trained for 35 epochs on CIFAR-10, rather than 350 epochs on ImageNet. Second, the paper's compound scaling strategy benefits from larger images and more classes, both of which are limited in CIFAR-10. Lastly, larger compound scaling steps (beyond B1) caused out-of-memory errors, limiting our ability to fully explore deeper scaling. Despite this, our models consistently reached high accuracy on CIFAR-10, with the B0 baseline achieving 89.45% and our compound scaling model achieving 85.21%. This suggests that EfficientNet's design principles are robust even with fewer resources, and that the FLOP-efficient B0 and B1 architectures are well-suited for smaller compute environments.

Reflections

Our reimplementation reaffirmed that EfficientNet architectures are highly effective even under limited resources, but also highlighted how sensitive scaling performance can be to dataset size, resolution, and training schedule. While the original paper emphasized compound scaling as a principled improvement over single-dimension scaling, our results did not show significant benefits from compounding under our constraints. In fact, the baseline model outperformed the other models, and compound scaling offered no clear accuracy advantage over any variant. A key takeaway is that the full benefits of compound scaling may only emerge with larger datasets, higher image resolutions, and longer training schedules. These are all conditions we could not fully replicate. Our inability to scale beyond B1 due to memory limitations further illustrates the gap in available compute compared to the original setup. For future work, we would extend our experiments to cover the full EfficientNet family (B0–B7), match the 350-epoch training regime, and evaluate on larger datasets such as CIFAR-100, mini-ImageNet, or the full ImageNet benchmark. These steps would allow a more faithful replication and deeper insight into the compound scaling framework's practical impact.

References

[1] Tan, M., & Le, Q. V. (2020). *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*. arXiv:1905.11946v5. `https://arxiv.org/abs/1905.11946`