EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks

Aarav Khanna & Ashley Liu
github.com/Aarav-Khanna/EfficientNet-Reimplementation

1. Introduction

This project reimplements the key contributions of the paper *"EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks"* by Tan and Le (2019). The paper proposes a compound scaling method for CNNs, efficiently scaling depth, width, and resolution. Our goal was to reproduce the paper's scaling behavior using CIFAR-100 due to resource constraints, and compare it to ResNet-50, DenseNet-201, and ResNet-152. We also examined model interpretability using Class Activation Maps (CAMs).

2. Chosen Result

We reimplemented the core contributions of the EfficientNet paper. First, we reproduced the grid search for the optimal compound scaling coefficients (α, β, γ) under the constraint $\alpha \cdot \beta^2 \cdot \gamma^2 \approx 2$, as described in the original work. Using these coefficients, we constructed EfficientNet models B0 through B7 and trained them on CIFAR-100. We then replicated the key result of the paper: the accuracy improvements as model size increases, comparing our EfficientNet models to ResNet-50, DenseNet-201, and ResNet-152 on CIFAR-100, all under the same computational constraints—mirroring the analysis in Figure 1 of the original paper. Additionally, we reproduced the validation and test accuracy results (Table 8 in the paper) for our models. Finally, we generated Class Activation Map (CAM) visualizations to replicate the interpretability analysis shown in Figure 7 of the paper.

3. Methodology

Re-implementation Approach:

- **Architecture:** Built EfficientNet B0–B7 using PyTorch, with MBConv blocks, squeeze-and-excitation, and Swish activation. Baselines included ResNet-50, DenseNet-201, and ResNet-152.
- **Scaling Search:** Ran grid search for (α, β, γ) values under the compound constraint.
- **Dataset:** Used CIFAR-100 with aggressive augmentation (crop, flip, rotation, jitter).
- **Training:** Trained under a fixed computational budget (FLOPs/params) with mixed precision and multi-GPU.
- **Evaluation:** Measured top-1 accuracy, parameters, FLOPs. Used CAMs for interpretability.
- **Modifications:** Input/image size adapted to CIFAR-100. Used A100 GPU optimizations.

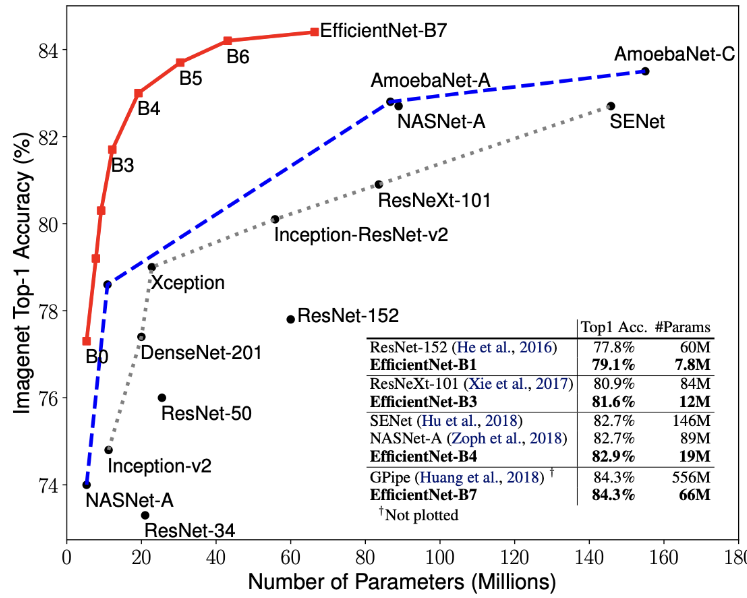

Figure 1. **Model Size vs. ImageNet Accuracy.** All numbers are for single-crop, single-model. Our EfficientNets significantly outperform other ConvNets. In particular, EfficientNet-B7 achieves new state-of-the-art 84.3% top-1 accuracy but being 8.4x smaller and 6.1x faster than GPipe. EfficientNet-B1 is 7.6x smaller and 5.7x faster than ResNet-152. Details are in Table 2 and 4.

Figure 1: Original: Model Size vs. ImageNet Accuracy

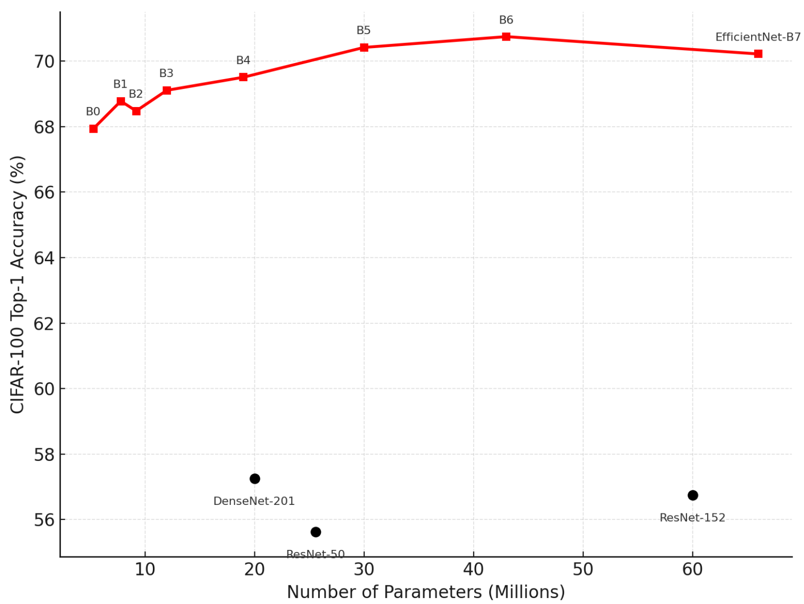

Figure 2: Reproduced: Model Size vs. CIFAR-100 Accuracy

4. Results & Analysis

Scaling Results: CIFAR-100 accuracy improves consistently from B0 to B6, validating compound scaling's benefits. EfficientNet outperforms DenseNet-201 and ResNet-152 at similar parameter levels.

Table 8. **ImageNet Validation vs. Test Top-1/5 Accuracy**.

	B0	B1	B2	B3	B4	B5	B6	B7
Val top1	77.11	79.13	80.07	81.59	82.89	83.60	83.95	84.26
Test top1	77.23	79.17	80.16	81.72	82.94	83.69	84.04	84.33

Figure 3: Original: ImageNet Validation vs. Test Accuracy

Table 1: **CIFAR-100 Validation vs. Test Top-1 Accuracy**

	B0	B1	B2	B3	B4	B5	B6	B7
Val top-1	63.14	64.26	63.94	65.98	67.04	66.44	66.56	67.48
Test top-1	67.94	68.78	68.48	69.11	69.51	70.42	70.75	70.22

Figure 4: Reproduced: CIFAR-100 Validation vs. Test Accuracy

Interpretability: CAM visualizations show that larger, compound-scaled models highlight more relevant image regions, supporting the paper's claims about attention.

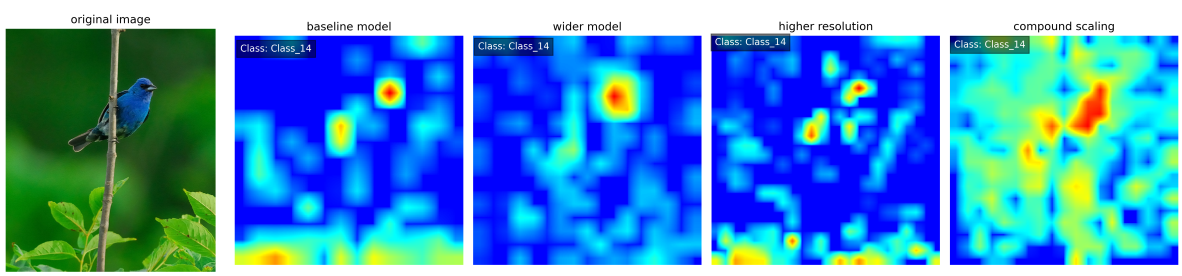

Figure 5: CAM visualizations across scaling strategies

Challenges: CIFAR-100's smaller image size required adapting resolution-based scaling. Matching ImageNet results was not feasible, but trends remained consistent. Some scaling levels (e.g., B7) overfit due to CIFAR-100's limited resolution.

5. Reflections

Motivation:

- As computational costs and environmental concerns rise, building efficient models is more important than ever.

- EfficientNet's compound scaling provides a principled method to scale networks while maintaining strong performance, aligning with industry trends in efficient deep learning.

Approach and Limitations:

- The original EfficientNet results were obtained using ImageNet and extensive compute resources, which were not practical for our setting.
- We adapted the approach to CIFAR-100 and scaled down input resolution to suit available compute.
- While we could not match ImageNet-level results, our reproduced trends still validate the method's effectiveness.
- With more time and resources, we could extend the work to include transfer learning and test EfficientNet's performance on larger or more diverse datasets.

Investigation and Insights:

- We carefully replicated the compound scaling formula, matching the hyperparameter search logic from the original work.
- Reproduced performance vs. model size trends with clear visualizations (figures/tables).
- Visual inspection via CAMs confirmed performance gains corresponded to better focus in feature maps.
- Assumptions like fixed FLOP budget and limited training time were documented and used in comparative evaluations.

Key Takeaways:

- Compound scaling is applicable beyond ImageNet, with consistent accuracy benefits on CIFAR-100.
- Interpretability improves with model scaling, likely due to better capacity and focus.

Future Directions:

- Use pretraining or transfer learning for faster convergence on small datasets.
- Evaluate scaling on medical imaging or low-light datasets.
- Replace Swish with newer activations like GELU or Mish to test efficiency.

6. References

- Tan, M., & Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the 36th International Conference on Machine Learning (ICML).*
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).*
- Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *CVPR.*
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. In *CVPR.*

- Ahmed, T., & Sabab, N. (2022). Classification and understanding of cloud structures via satellite images with EfficientUNet. *SN Computer Science*, 3(2), 1–10.
- Krizhevsky, Alex. (2012). Learning Multiple Layers of Features from Tiny Images. University of Toronto.