

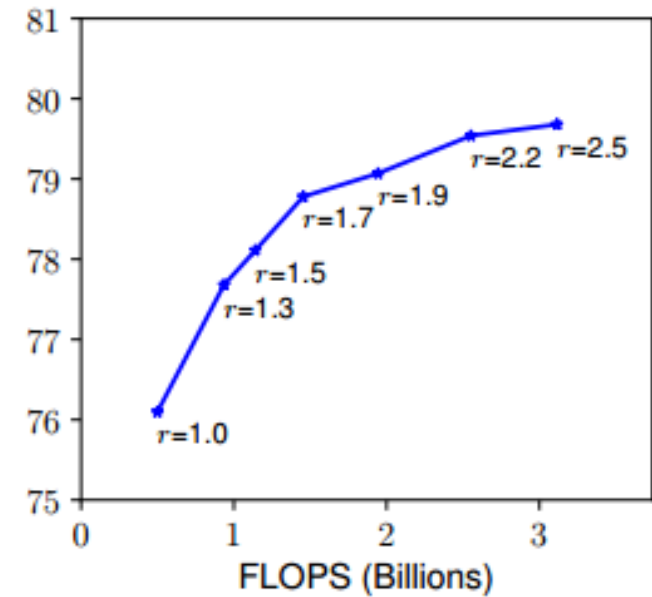
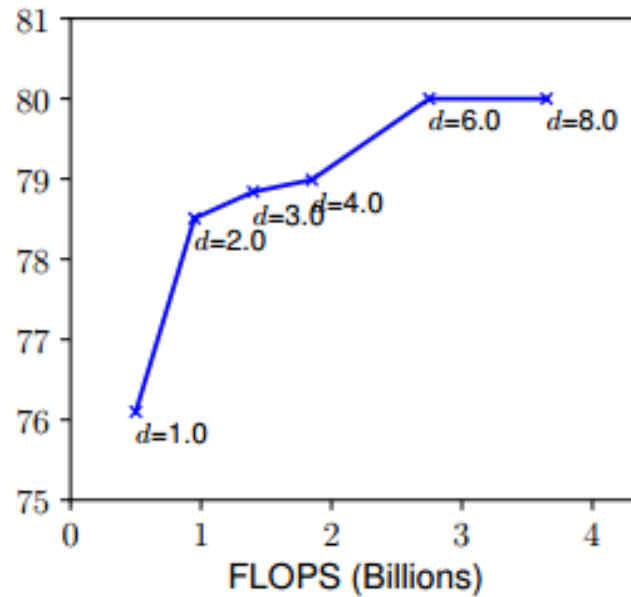
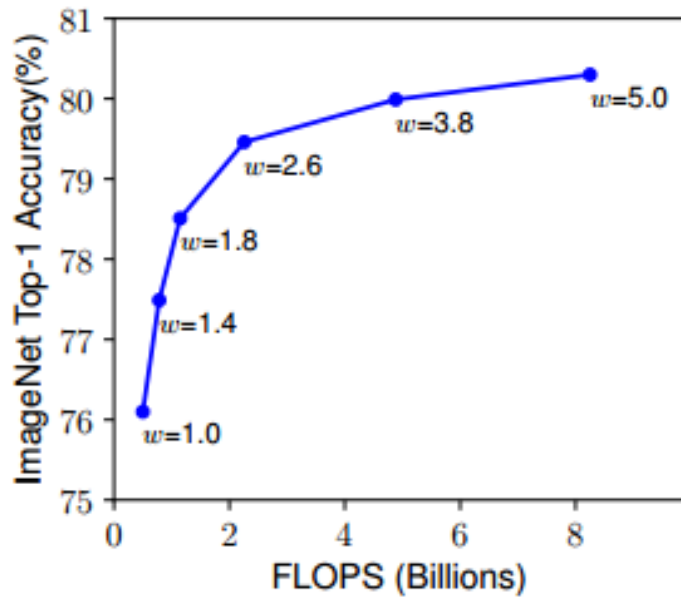
EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks

Background

- 모델의 성능을 높일 때 일반적으로 **모델의 깊이, 너비, 입력** 이미지의 크기를 조절한다. 기존에는 이 세 가지를 수동으로 조절하였기 때문에, 최적의 성능과 효율을 얻지 못한다.

Motivation

- 그림을 보면 w , d , r 이 일정 값 이상이 되면 정확도가 빠르게 포화한다. 그리고, w , d , r 이 낮을 때는 약간만 값을 조절해도 효과가 크네요.



<순서대로 네트워크 width, depth, resolution을 scaling 했을 때 연산량과 정확도>

Model

- EfficientNet은 width, depth, resolution을 조절하여 최고의 효율을 찾는 compound scaling 을 제안한다.
- α, β, γ 는 small grid search로 결정되는 상수입니다. 그리고, ϕ 는 주어진 연산량에 따라 사용자가 결정하는 상수입니다. **FLOPs는 너비와 해상도에 따라 제공배가 상승합니다. 따라서 FLOPs는 $(\alpha \cdot \beta^2 \cdot \gamma^2)^2$ 배 만큼 증가합니다. 논문에서는 $\alpha \cdot \beta^2 \cdot \gamma^2 = 2$ 로 제한합니다. 제한된 범위에서 α, β, γ 를 찾는 것입니다. 총 FLOPs는 2^ϕ 만큼 증가합니다.**

$$\text{depth: } d = \alpha^\phi$$

$$\text{width: } w = \beta^\phi$$

$$\text{resolution: } r = \gamma^\phi$$

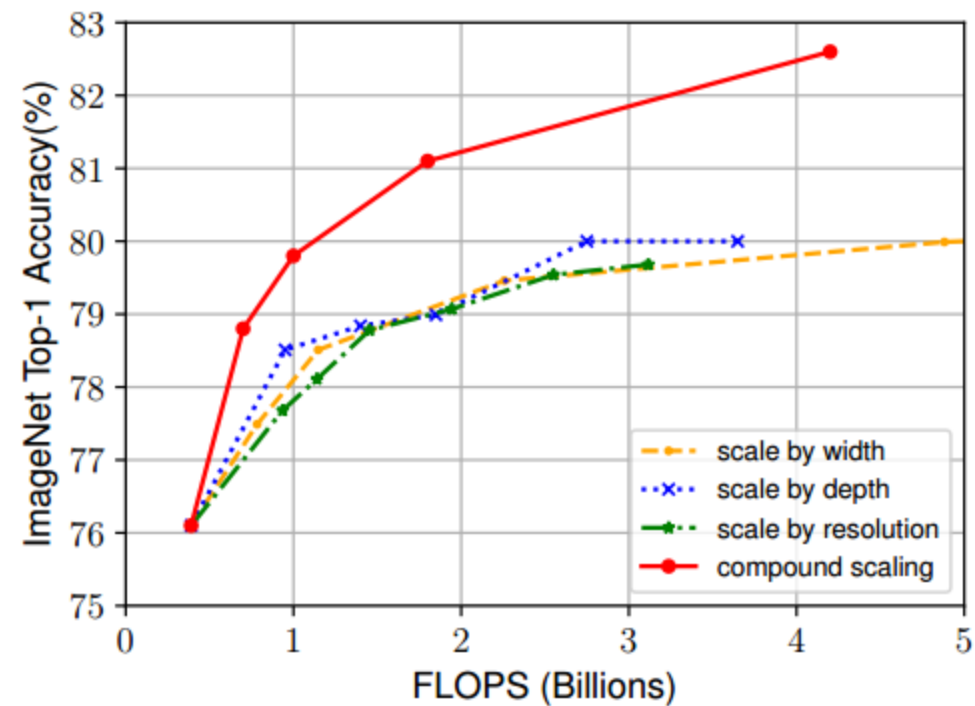
$$\text{s.t. } \alpha \cdot \beta^2 \cdot \gamma^2 \approx 2$$

$$\alpha \geq 1, \beta \geq 1, \gamma \geq 1$$

<변수와 조건>

Result

- Compound scaling 할 때 depth, width, resolution 중 하나만을 scaling하였을 때보다 같은 연산량에서 더 좋은 성능을 보였다.



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%

<한가지만 scaling 했을 때와 compound scaling 비교>