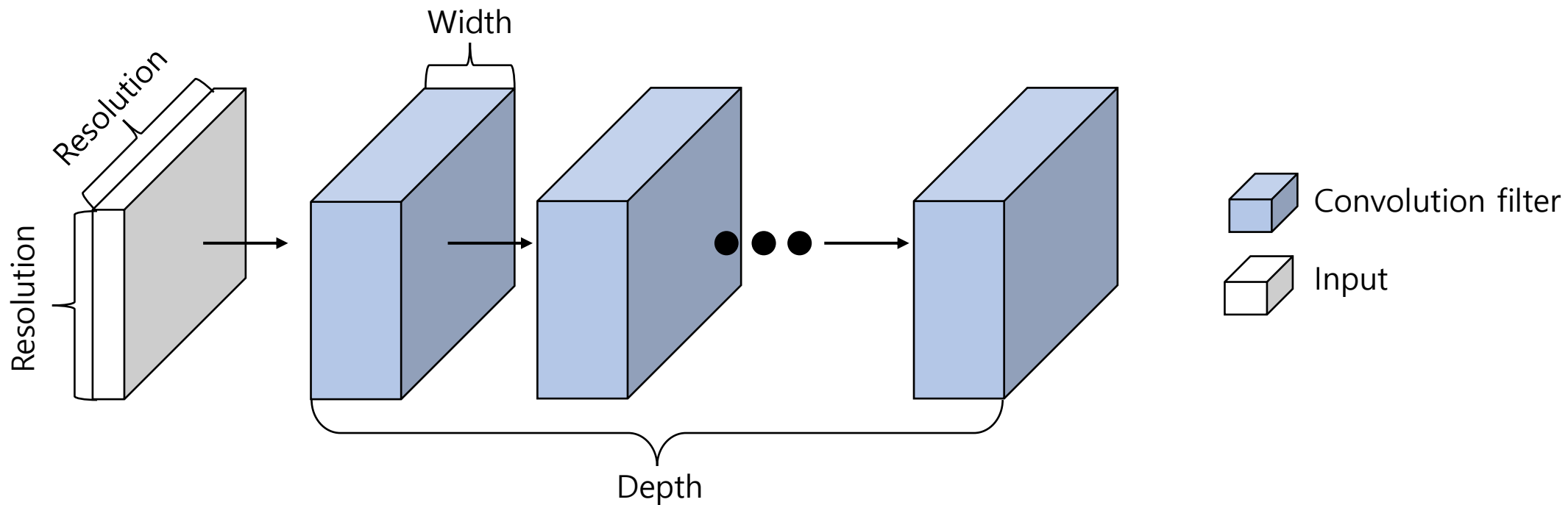


EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks

Written in 2022.08.23

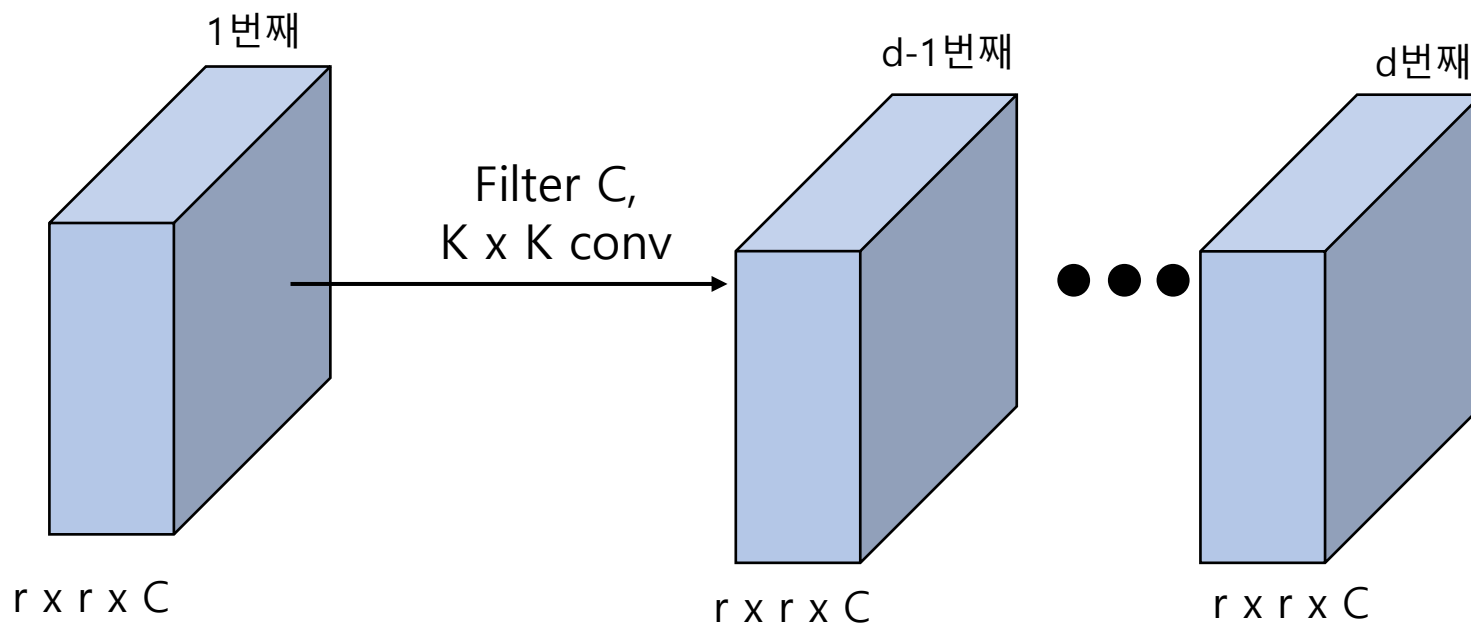
Research Background

- 모델의 성능을 높일 때 일반적으로 모델의 depth(레이어가 쌓인 정도), width(필터의 개수), 입력 이미지의 resolution를 조절함
- 기존에는 이 세 가지를 수동으로 조절하였기 때문에, 최적의 성능과 효율을 얻지 못함
- 최적의 depth, width, resolution을 찾는 compound scaling 을 제안함



Model

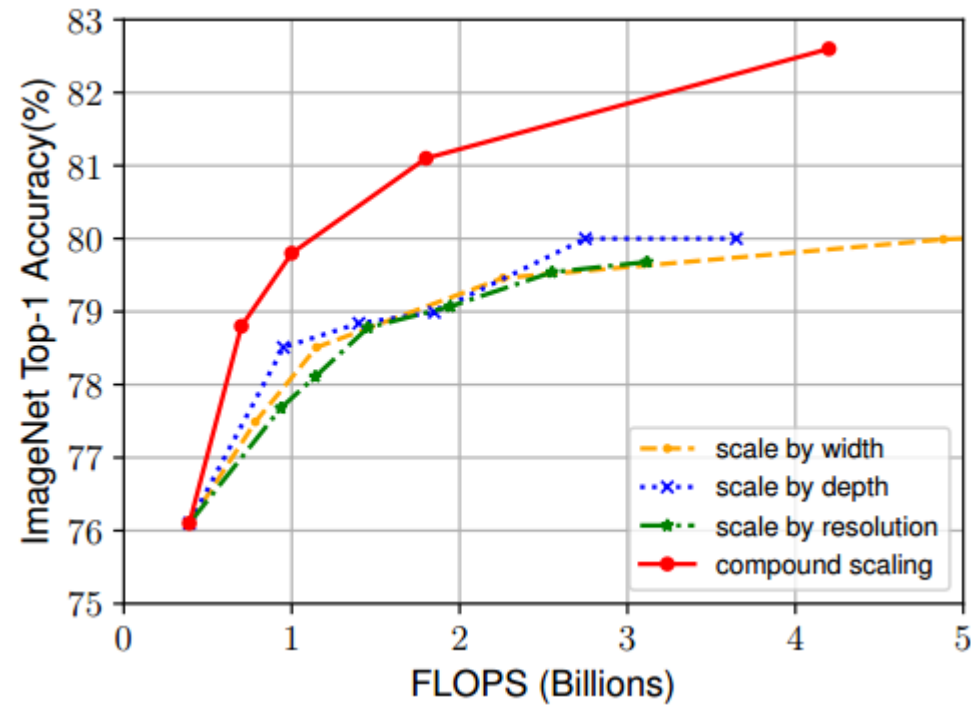
- 아래 계산식에서 depth가 d 이고, 연산량은 d 가 증가하는 것에 비례하여 증가함
- 아래 계산식에서 width는 c 이고, 연산량은 c 가 증가하는 것의 제곱 배만큼 증가함
- 아래 계산식에서 resolution은 r 이고, 연산량은 r 이 증가하는 것의 제곱 배만큼 증가함
- $\text{depth} = \alpha^\phi$, $\text{width} = \beta^\phi$, $\text{resolution} = \gamma^\phi$, ϕ 는 주어진 연산량에 따라 사용자가 결정하는 상수
- $\alpha \cdot \beta^2 \cdot \gamma^2 = 2$ 로 제한하여 제한된 범위에서 최적의 값을 찾음
- 연산량은 $(\alpha \cdot \beta^2 \cdot \gamma^2)$ 배 만큼 늘어남



연산량: $(r * r * C) * (K * K * C) * d$

Result

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



- 모델의 성능을 높이기 위해 구조를 바꾸거나 새로운 알고리즘을 제안하는 것이 아닌 모델 자체를 최적화하는 방법이 있다는 걸 깨달았다.