

EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks

- * 모델 스케일링을 통해 network depth, width, resolution이 더 좋은 성능을 낸다는 것을 확인
- * 논문의 저자들은 뉴럴 아키텍처를 사용해 새로운 베이스라인 네트워크를 디자인하고, 그것을 스케일링해 EfficientNets라고 불리는 family of models를 얻음
- * ConvNets를 스케일링하는 것은 더 좋은 성능을 이루기 위해 널리 사용됨
- * 그렇지만 ConvNets를 스케일링하는 과정은 잘 이해되지 않았고 최근에는 하기 위한 다양한 방법들이 있음
 - ConvNets를 스케일링하는 가장 흔한 방법은 그것들의 depth나 width를 사용하는 방법이 있음
 - 덜 흔한 방법, 그렇지만 점점 인기를 얻고 있는 방법으로 image resolution을 이용해 모델을 스케일링하는 방법이 있음
- * 논문에서는 “더 나은 성능과 효율성을 얻기 위해 ConvNets를 scale up을 하는 주요한 방법일까?”라는 질문을 던짐
- * Input image가 커질수록, 네트워크는 receptive field를 증가시키기 위해 더 많은 레이어들이 필요하고, 더 큰 이미지에서 fine-grained pattern을 캡처하기 위해 더 많은 채널들이 필요함
- 복합적인 스케일링 메소드가 필요

* 관련 연구

1) ConvNet Accuracy

- 최근 연구들은 더 나은 ImageNet 모델들이 다양한 전이 학습 데이터셋과 객체 탐지와 같은 다른 컴퓨터 비전 task에서 좋은 성능을 보여줌
- 비록 더 높은 성능이 많은 애플리케이션에서 중요하지만, 하드웨어 메모리 제한에 이미 걸리고, accuracy gain은 더 좋은 효율성을 필요로 함

2) ConvNet Efficiency

- Model Compression은 모델 사이즈를 줄이기 위한 흔한 방법, 효율성을 위해 성능을 교환함
- 최근에 뉴럴 아키텍처 search는 효율적인 mobile-size ConvNets를 디자인하는 것이 굉장히 인기가 많고, network width, depth, convolution kernel type과 size를 광범위하게 튜닝함으로써 hand-crafted mobile ConvNets보다 좋은 효율성을 이룸
- 그렇지만, 이 기술들을 훨씬 큰 design space와 비싼 tuning cost를 가진 더 큰 모델에 적용하는 방법이 불분명함

3) Model Scaling

- 이전의 연구들이 ConvNet들의 expressive power를 위해 network depth와 width 모두 중요하다는 것을 보여줬지만, ConvNet이 더 나은 효율성과 성능을 이루기 위해 효율적으로 스케일링하는 것은 의문으로 남음

* Compound Model Scaling

1) Problem Formulation

- ConvNet

$$\mathcal{N} = \bigodot_{i=1 \dots s} \mathcal{F}_i^{L_i}(X_{\langle H_i, W_i, C_i \rangle})$$

- 최적화 문제

$$\max_{d, w, r} \text{Accuracy}(\mathcal{N}(d, w, r))$$

$$s.t. \quad \mathcal{N}(d, w, r) = \bigodot_{i=1 \dots s} \hat{\mathcal{F}}_i^{d \cdot \hat{L}_i}(X_{\langle r \cdot \hat{H}_i, r \cdot \hat{W}_i, w \cdot \hat{C}_i \rangle})$$

$$\text{Memory}(\mathcal{N}) \leq \text{target_memory}$$

$$\text{FLOPS}(\mathcal{N}) \leq \text{target_flops}$$

- 2) Scaling Dimension: 전통적인 메소드들은 ConvNet들은 다음과 같은 dimension들 중 하나 중에서 거의 스케일링함

- Depth: Deeper ConvNet은 더 풍부하고 더 복잡한 특징들은 캡처하고, 새로운 task에 더 일반화할 수 있음. 그렇지만 Deeper 네트워크는 그래디언트 소실 문제를 훈련시키기 위해 더 힘들

- Width: 네트워크 width를 스케일링하는 것은 작은 크기의 모델을 위해 자주 사용됨

- Resolution

→ 어느 network width, depth, or resolution의 차원을 스케일링 up하던지 성능을 향상시키고, 더 큰 모델에서는 성능 향상이 사라짐

3) Compound Scaling

- 더 큰 receptive field들이 더 큰 이미지에 있는 더 많은 픽셀들을 포함할 수 있도록 높은 해상도를 가진 이미지에서 네트워크 깊이를 늘려야 함

- 그것과 같이 해상도가 높을 때 높은 해상도의 이미지에서 더 많은 픽셀들을 가진 fine-grained pattern을 더 캡처할 수 있도록 네트워크 너비를 늘려야 함

→ 더 나은 성능과 효율성을 위해 ConvNet 스케일링 동안 네트워크 깊이, 넓이, 해상도가 모두 균형을 가지도록 하는 것이 중요함

- 이 논문에서는 compound coefficient ϕ 를 사용하는 새로운 복합적인 스케일링 메소드를 제안

함

→ network width, depth, resolution을 원칙에 기반하여 균등하게 스케일링함

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

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

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

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

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

* EfficientNet Architecture

- EfficientNet-B0 베이스라인에서 시작해서 compound scaling method를 scale up하기 위해 2단계로 적용함