

EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks

≡ Author	Mingxing Tan, Quoc V. Le
≡ Field	CV
⌵ Journal	ICML
# Published Year	2019
≡ Speaker	나보영 조태완
≡ Summary	EfficientNet
⌵ status	Finished!
🔗 link	https://arxiv.org/pdf/1905.11946v5.pdf

1. 논문이 풀고자 하는 문제

ResNet과 Inception Networks가 등장한 이후, 깊이와 복잡성이 늘어난 모델들이 ImageNet 리더보드를 장악하게 되었습니다. 이러한 모델들은 많은 수의 파라미터와 데이터가 필요로 했습니다.

▼ ResNet

- **개념:** ResNet은 "Residual Networks"의 약자로, 깊은 네트워크에서 발생할 수 있는 기울기 소실(Vanishing Gradient) 또는 기울기 폭발(Exploding Gradient) 문제를 해결하기 위해 "Residual Block"이라는 구조를 사용합니다.
- **특징:** 이는 입력을 출력에 바로 더해주는 스킵 연결(skip connection)을 통해 구현되며, 이로 인해 깊은 네트워크에서도 안정적인 학습이 가능합니다.

▼ Inception Networks

- **개념:** Inception 네트워크는 여러 크기의 필터와 풀링 레이어를 병렬로 연결하여 다양한 특징을 추출할 수 있게 설계되었습니다.
- **특징:** 이러한 복잡한 구조로 인해 Inception 네트워크는 다양한 스케일의 특징을 잘 잡아낼 수 있습니다.

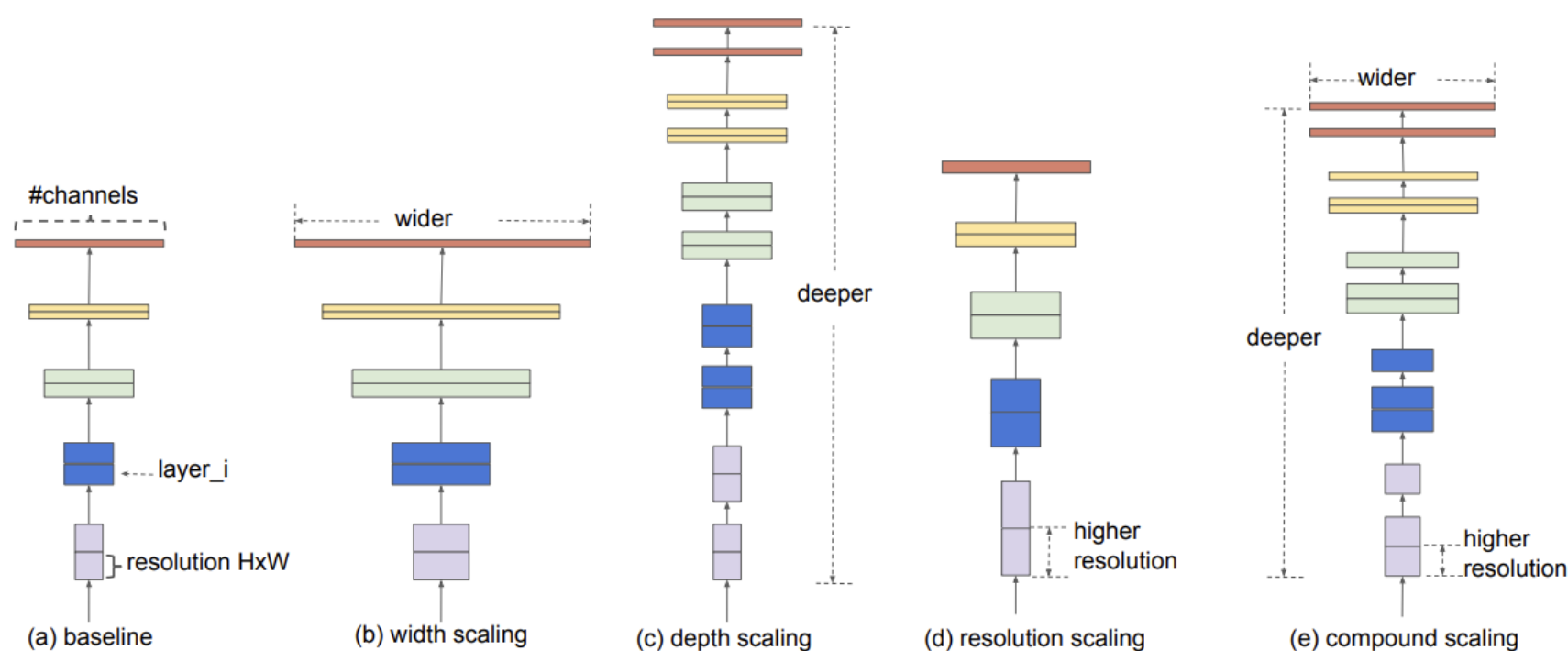


Figure 2. Model Scaling. (a) is a baseline network example; (b)-(d) are conventional scaling that only increases one dimension of network width, depth, or resolution. (e) is our proposed compound scaling method that uniformly scales all three dimensions with a fixed ratio.

▼ Figure 2. Model Scaling

(a) 기존 네트워크 예시

- 이 부분은 기본 네트워크 아키텍처를 나타냅니다. 이 기존 네트워크는 다른 확장 방법들과 비교될 기초가 됩니다.

(b)-(d) 전통적인 스케일링 방법

- 이 부분들은 네트워크의 너비, 깊이, 또는 해상도 중 하나만을 증가시키는 전통적인 스케일링 방법을 보여줍니다.
- 예를 들어, (b)는 네트워크의 너비만 증가시킨 경우, (c)는 깊이만 증가시킨 경우, 그리고 (d)는 해상도만 증가시킨 경우를 각각 나타냅니다.

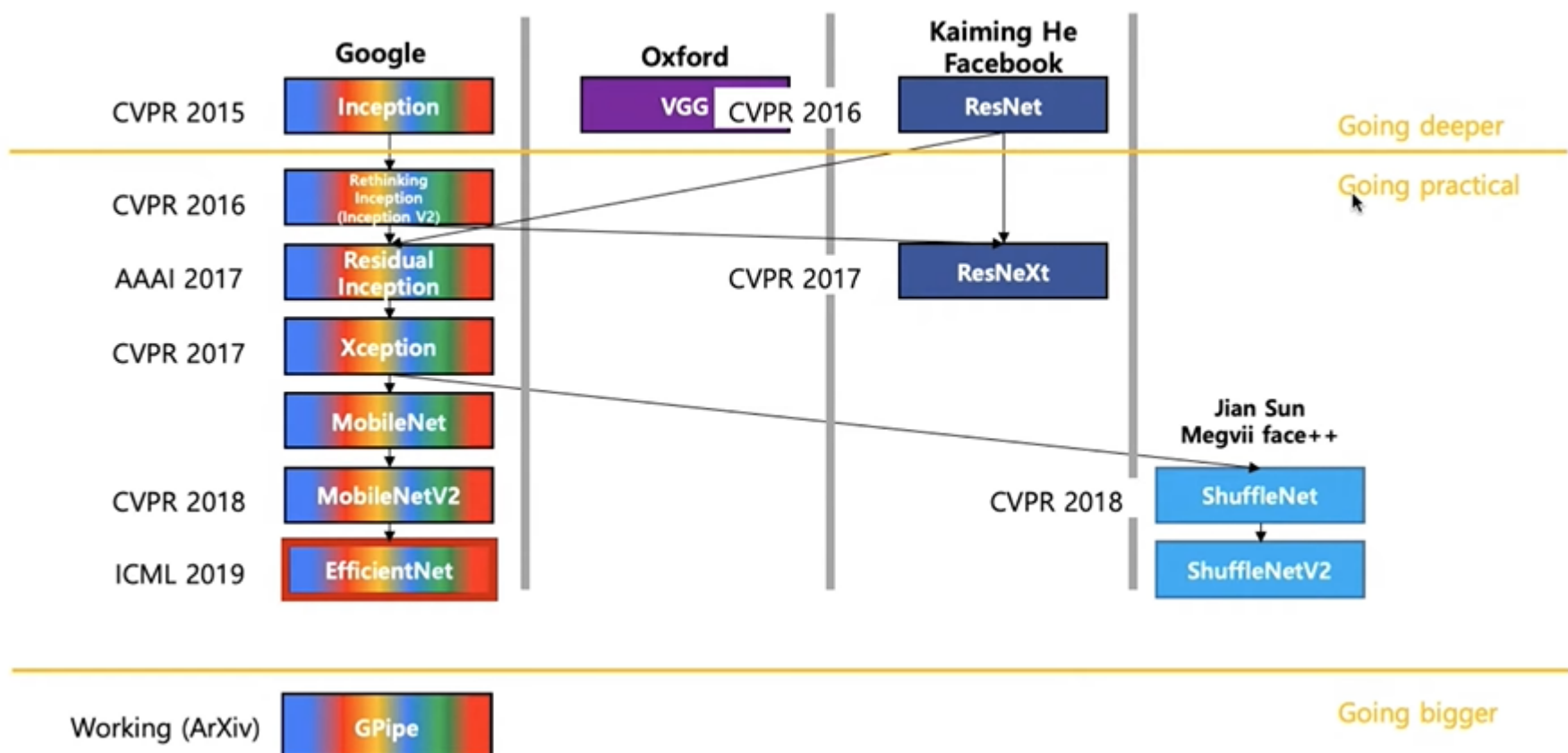
(e) 제안된 복합 스케일링 방법

- 이 부분은 논문에서 제안하는 "복합 스케일링 방법"을 나타냅니다.
- 이 방법은 너비, 깊이, 해상도 모든 세 가지 차원을 일정한 비율로 균일하게 확장합니다.

이 그림은 전통적인 스케일링 방법이 하나의 차원만을 확장하는 반면, 복합 스케일링 방법은 모든 차원을 균형 있게 확장한다는 차이점을 시각적으로 보여줍니다. 이러한 접근법은 네트워크의 성능과 효율성을 동시에 높일 수 있음을 의미합니다.

- ConvNet 모델을 어떻게 효율적으로 확장(scaling)할 수 있을까?
- 깊이(depth), 너비(width), 해상도(resolution) 등 여러 차원을 어떻게 균형 있게 조절하여 성능을 향상시킬 수 있을까?
- 주어진 Recourse(메모리와 FLOPs) 제약 내에서, 최대 성능(정확도)을 얻을 수 있을까?

2. 관련 연구



ConvNet Accuracy

1. **역사적 발전:** AlexNet이 2012년 ImageNet 대회에서 우승한 이후, ConvNets는 점점 더 정확해지고 있습니다. 예를 들어, 2014년의 GoogleNet은 약 6.8M의 파라미터로 74.8%의 top-1 정확도를, 2017년의 SENet은 145M 파라미터로 82.7%의 top-1 정확도를 달성했습니다.
2. **최근 동향:** GPipe는 557M의 파라미터를 사용하여 84.3%의 top-1 정확도를 달성했습니다. 이 모델은 너무 크기 때문에 특별한 파이프라인 병렬처리 라이브러리를 사용해야 합니다.
3. **전이 학습과 다른 작업에서의 성능:** 더 나은 ImageNet 모델은 전이 학습과 객체 탐지 등 다른 컴퓨터 비전 작업에서도 더 좋은 성능을 보입니다.

but, 이미 하드웨어 메모리 한계에 도달했기 때문에, 더 나은 정확도를 달성하려면 효율성이 필요합니다.

ConvNet Efficiency

1. **과다 매개변수화 문제:** 깊은 ConvNets은 종종 Over-Parameterized 됨
2. **모델 압축:** 모델 크기를 줄이기 위한 일반적인 방법입니다. 하지만 이는 정확도를 희생할 수 있습니다.
3. **모바일 네트워크:** SqueezeNets, MobileNets, ShuffleNets 등의 효율적인 모바일 크기의 ConvNets이 개발되고 있습니다.

but, 작은 모델에는 NasNet 방법이 적합할 수 있으나, 큰 모델에 적용할 수 있을지는 불분명(Unclear)

Model Scaling

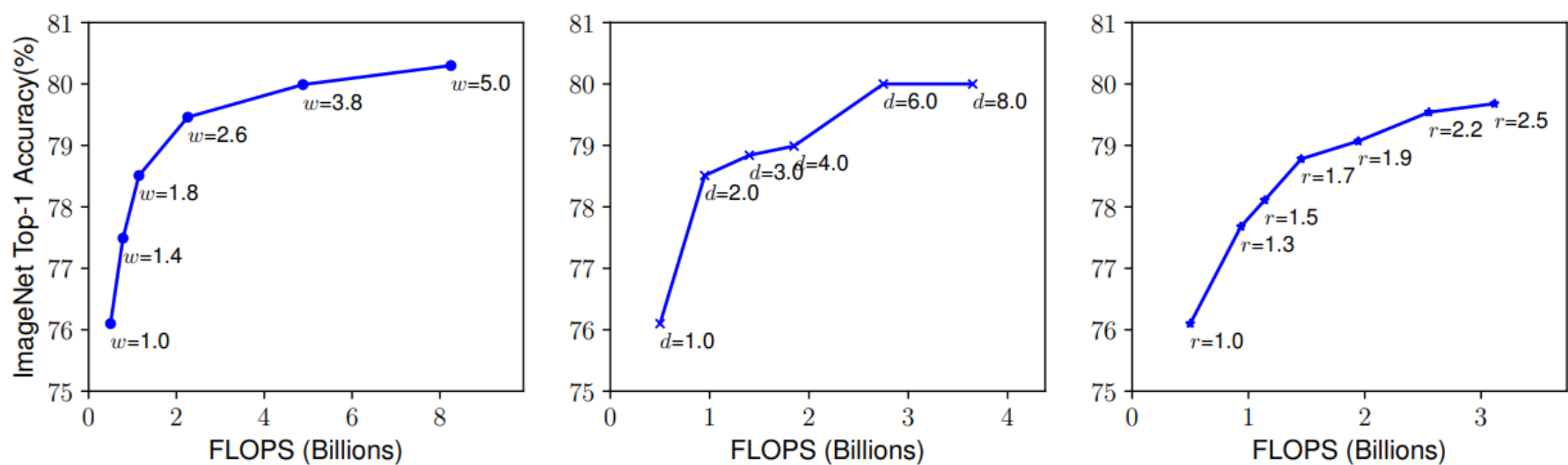
1. **다양한 방법**: ResNet은 네트워크 깊이를 조정하여 스케일을 조절할 수 있고, WideResNet과 MobileNets은 네트워크 너비를 조정할 수 있습니다.
2. **입력 이미지 크기의 중요성**: 더 큰 입력 이미지 크기는 더 높은 정확도를 가져오지만, 계산 부하도 증가시킵니다.

so, 네트워크의 너비, 깊이, 해상도 모든 세 가지 차원에 대한 ConvNet 스케일링을 체계적으로 연구합니다.

3. 논문에서 제시한 아이디어

- resolution이 큰 모델을 사용하게 된다면 특징을 더 잡아내기 위하여 더 많은 레이어를 사용하게 되고 이미지의 패턴을 찾아내기 위해 더 많은 채널을 사용할거라고 생각
- 복합 계수(compound coefficient)를 사용하여 깊이, 너비, 해상도를 균일하게 확장.
- Scaling을 공식화(Formulation)하여 문제를 해결하고자 함
- 신경 아키텍처 탐색(Neural Architecture Search)을 통해 새로운 기준 모델을 설계하고 이를 확장하여 EfficientNets라는 새로운 모델 계열을 개발.

4. 구현



depth 증가

- 증가 할수록 vanishing gradient problem로 인해 훈련시키기 어렵고 정확도 또한 일정 레이어 이상이 되면 더 이상 크게 향상 되지 않는것을 확인 할 수 있다.

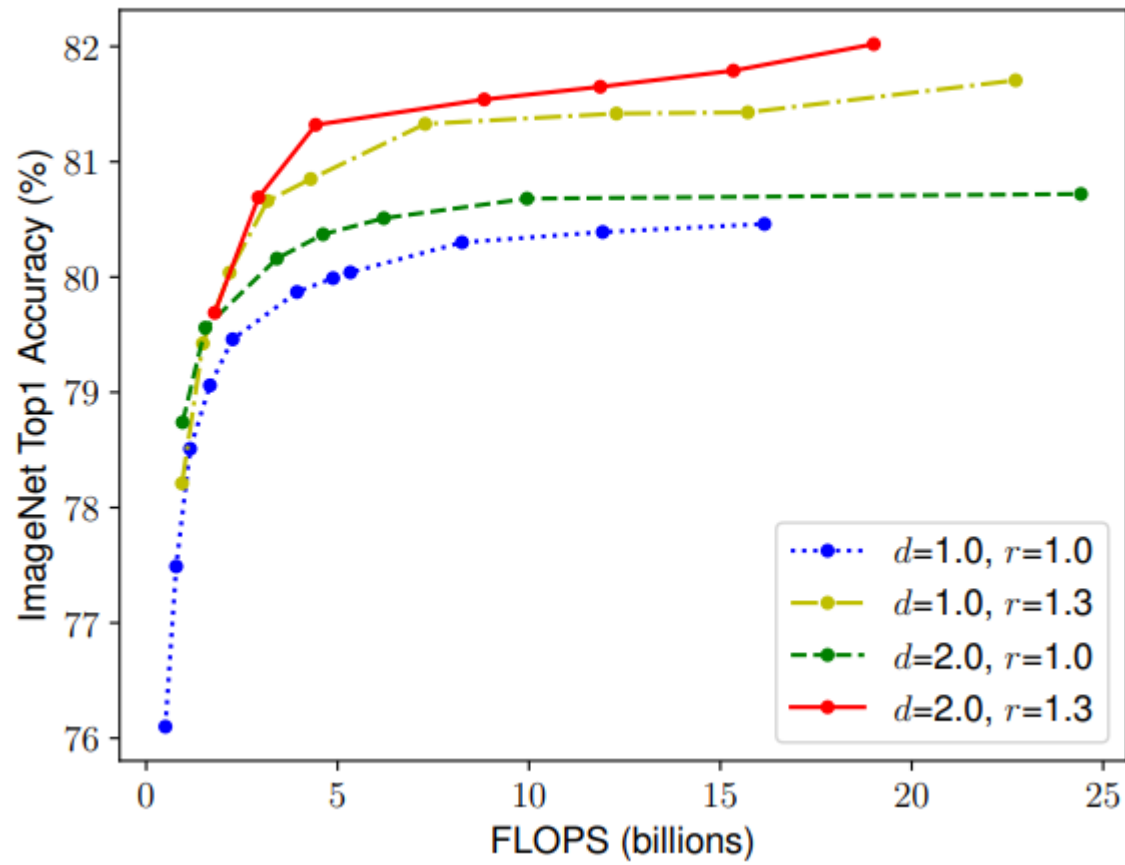
Width 증가

- Width가 증가 할 수록 초반에는 정확도가 빠르게 나중에는 더이상 증가 하지 않음

resolution 증가

- resolution이 증가할 수록 정확도 증가 but 일정 해상도 이상이 되면 오히려 정확도가 줄어드는 현상 발생

세가지 모두 accuracy를 향상 시키지만 일정 수준을 넘어가면 더이상 증가 x



depth 와 resolution 고정 Width를 조정한 실험
 $d=1.0, r=1.0$ 을 고정 → 정확도 빠르게 포화
 $d=2.0, r=1.3$ 동일한 연산 대비 더 나은 정확도

위의 실험을 통해서 depth, width, resolution을 균형있게 조율하는것이 중요하다.

▼ 이전 연구

너비와 깊이를 임의로 균형을 맞추려고 하였으나 다 수동 조정이 필요
 위 논문은 수동이 아닌 균일하게 확장하는 새로운 복합 스케일링 방법을 제안

Compound scaling

$$\alpha \cdot \beta^2 \cdot \gamma^2 \approx 2$$

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

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

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

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

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

STEP 1: $\phi=1$ 로 고정한뒤 작게 grid search를 수행하여 α, β, γ 를 찾는다. 큰 모델에서는 실험에 들어가는 자원이 많기 때문에 작은 baseline network에서 α, β, γ 를 먼저 찾는다.

$\alpha=1.2, \beta=1.1, \gamma=1.15$

STEP 2: step1에서 구한 α, β, γ 를 사용하여 ϕ 를 변화시키면서 전체적인 크기를 키운다.

$$ACC(m) \times [FLOPS(\hat{m})/T]^w$$

- optimization goal
- accuracy and FLOPS 최적화를 찾는것이 목표
- T는 목표 FLOPS w = -0.07 정확도와 FLOPS간의 균형을 제어하는 하이퍼파라미터

Table 1. EfficientNet-B0 baseline network – Each row describes a stage i with \hat{L}_i layers, with input resolution $\langle \hat{H}_i, \hat{W}_i \rangle$ and output channels \hat{C}_i . Notations are adopted from equation 2.

Stage i	Operator $\hat{\mathcal{F}}_i$	Resolution $\hat{H}_i \times \hat{W}_i$	#Channels \hat{C}_i	#Layers \hat{L}_i
1	Conv3x3	224×224	32	1
2	MBConv1, k3x3	112×112	16	1
3	MBConv6, k3x3	112×112	24	2
4	MBConv6, k5x5	56×56	40	2
5	MBConv6, k3x3	28×28	80	3
6	MBConv6, k5x5	14×14	112	3
7	MBConv6, k5x5	14×14	192	4
8	MBConv6, k3x3	7×7	320	1
9	Conv1x1 & Pooling & FC	7×7	1280	1

- efficientnet-b0 기본 네트워크 구조
- MnasNet에서 구조를 따옴
-

5. 실험

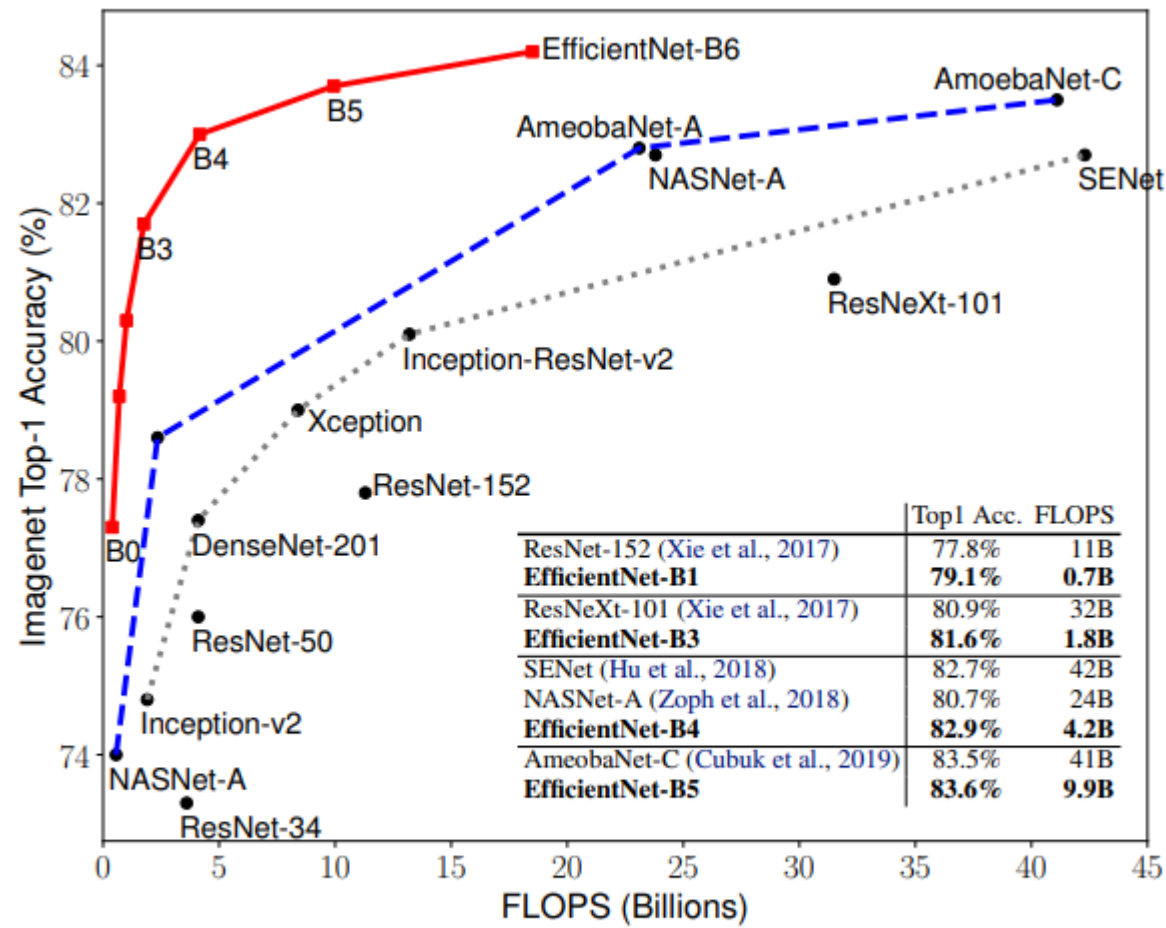


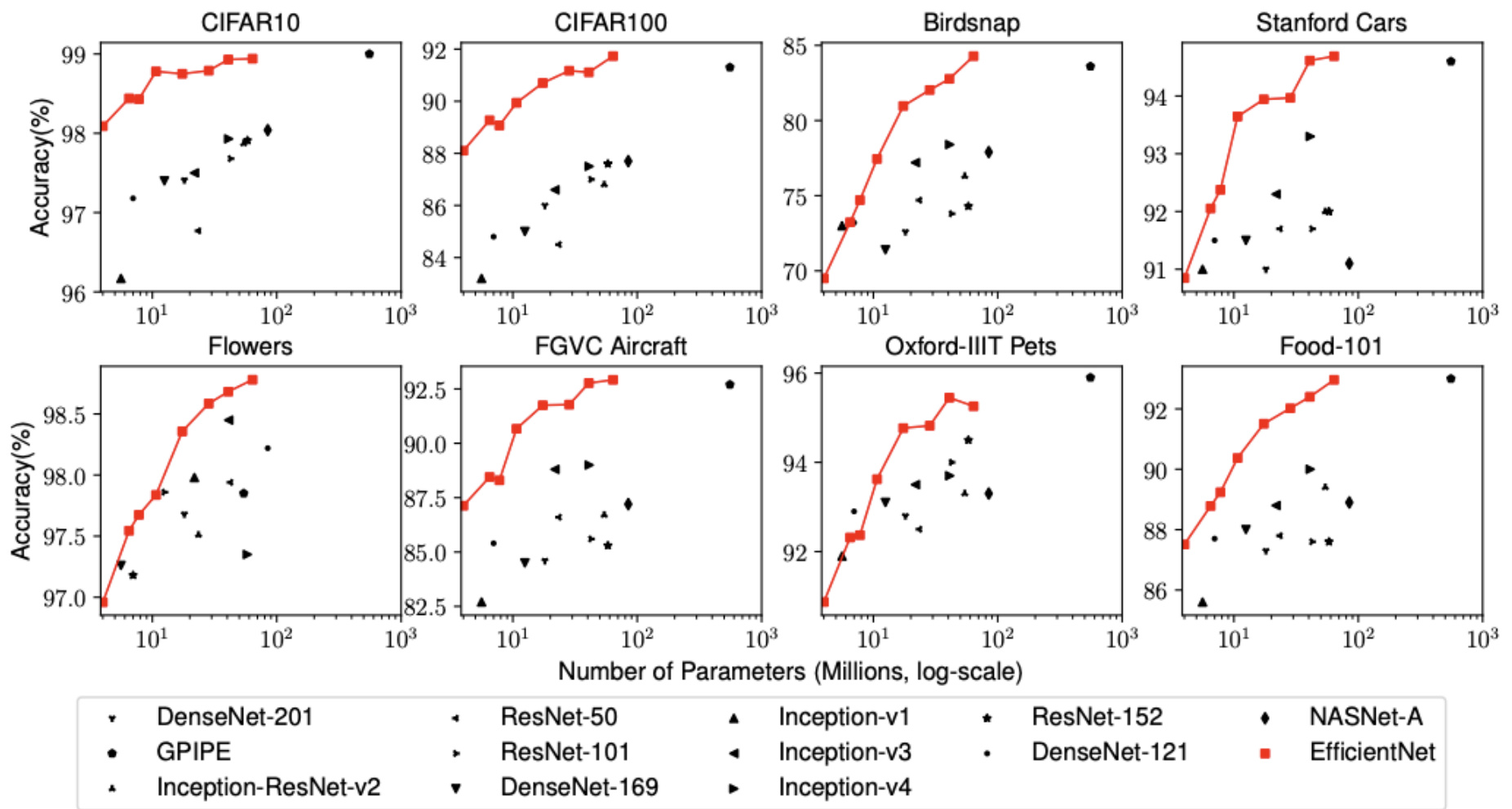
Figure 5. FLOPS vs. ImageNet Accuracy – Similar to Figure 1 except it compares FLOPS rather than model size.

- efficientnet을 사용해 model scaling을때 다른 기존 모델에 비해 적은 연산으로 더 높은 정확도를 갖는것을 확인 할 수 있다.

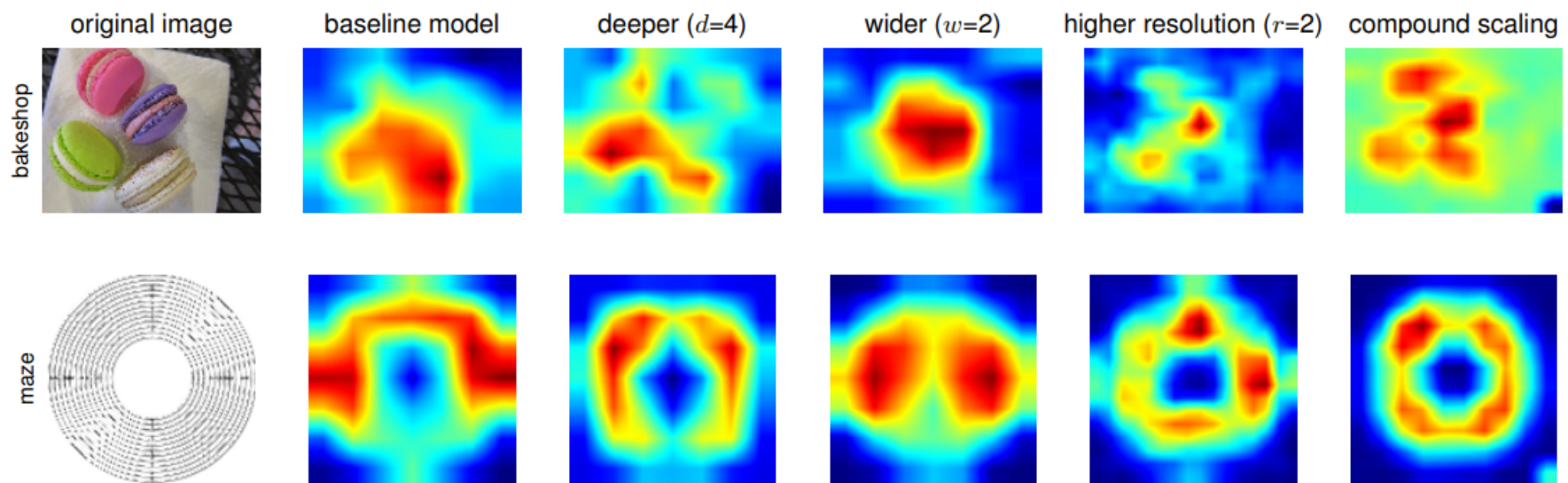
Table 3. Scaling Up MobileNets and ResNet.

Model	FLOPS	Top-1 Acc.
Baseline MobileNetV1 (Howard et al., 2017)	0.6B	70.6%
Scale MobileNetV1 by width ($w=2$)	2.2B	74.2%
Scale MobileNetV1 by resolution ($r=2$)	2.2B	72.7%
compound scale ($d=1.4, w=1.2, r=1.3$)	2.3B	75.6%
Baseline MobileNetV2 (Sandler et al., 2018)	0.3B	72.0%
Scale MobileNetV2 by depth ($d=4$)	1.2B	76.8%
Scale MobileNetV2 by width ($w=2$)	1.1B	76.4%
Scale MobileNetV2 by resolution ($r=2$)	1.2B	74.8%
MobileNetV2 compound scale	1.3B	77.4%
Baseline ResNet-50 (He et al., 2016)	4.1B	76.0%
Scale ResNet-50 by depth ($d=4$)	16.2B	78.1%
Scale ResNet-50 by width ($w=2$)	14.7B	77.7%
Scale ResNet-50 by resolution ($r=2$)	16.4B	77.5%
ResNet-50 compound scale	16.7B	78.8%

- 기존 모델 mobilenet과 resnet에 efficientnet을 접목했을때의 성능 비교
- depth, width, resolution을 각각 조절해봤을때 보다 compound scale했을때 더 좋은 결과가 나타남



- ImageNet에서 EfficientNet-B7은 84.3%의 top-1 정확도를 달성.
- EfficientNets 모델은 기존 ConvNets보다 크기가 작고 추론 속도가 빠름.
- 전이 학습에서도 CIFAR-100, Flowers 등에서 최고 수준의 정확도를 달성.



- compound scaling 한 모델이 다른 스케일링 방법을 사용한 모델에 비해 객체의 디테일과 관련되어 있는 영역에 더 집중하는 경향을 보임

6. 결론

- 이 논문에서는 ConvNet 스케일링을 체계적으로 연구하고, 네트워크의 너비, 깊이, 해상도를 신중하게 균형있게 조절하는 것이 중요하지만 누락된 부분이라고 지적합니다.
- 이 문제를 해결하기 위해, 간단하면서도 매우 효과적인 복합 스케일링 방법을 제안합니다. 이 방법을 사용하면 기본 ConvNet을 원하는 리소스 제약 조건에 맞게 체계적으로 확장할 수 있으며, 모델의 효율성을 유지할 수 있습니다. 이 복합 스케일링 방법을 활용하여, 모바일 크기의 **EfficientNet** 모델을 매우 효과적으로 확장할 수 있음을 보여줍니다.
- 결과적으로, 이 모델은 ImageNet과 다른 다섯 개의 일반적으로 사용되는 전이 학습 데이터셋에서, 훨씬 적은 수의 매개변수와 FLOPS를 사용하면서도 최첨단의 정확도를 뛰어넘을 수 있습니다.

7. 기타

▼ neural architecture search

NAS는 Neural Architecture Search의 약자로, 기계 학습 알고리즘이 자동으로 신경망 구조를 설계하는 과정입니다. NAS는 최적의 네트워크 구조를 찾기 위해 다양한 구조와 하이퍼파라미터를 탐색하고, 그 중 가장 성능이 좋은 모델을 선택합니다. 이는 기존에는 사람이 수행하던 신경망 설계 작업을 자동화하여, 더 빠르고 효율적으로 높은 성능의 모델을 찾을 수 있게 합니다.

▼ FLOPs

FLOPs는 "FLoating point OPerations"의 약자로, 부동소수점 연산의 횟수를 의미합니다. 이는 컴퓨터가 실수를 다룰 때 얼마나 많은 연산을 수행하는지를 나타내는 지표로 사용됩니다.

ex) 식 $A \times X + B$ 의 FLOPs를 계산하려면 다음과 같이 해석할 수 있습니다:

1. $A \times X$: 곱셈 1회
2. $+B$: 덧셈 1회

따라서 이 식의 FLOPs는 2가 됩니다.

▼ MobileNet

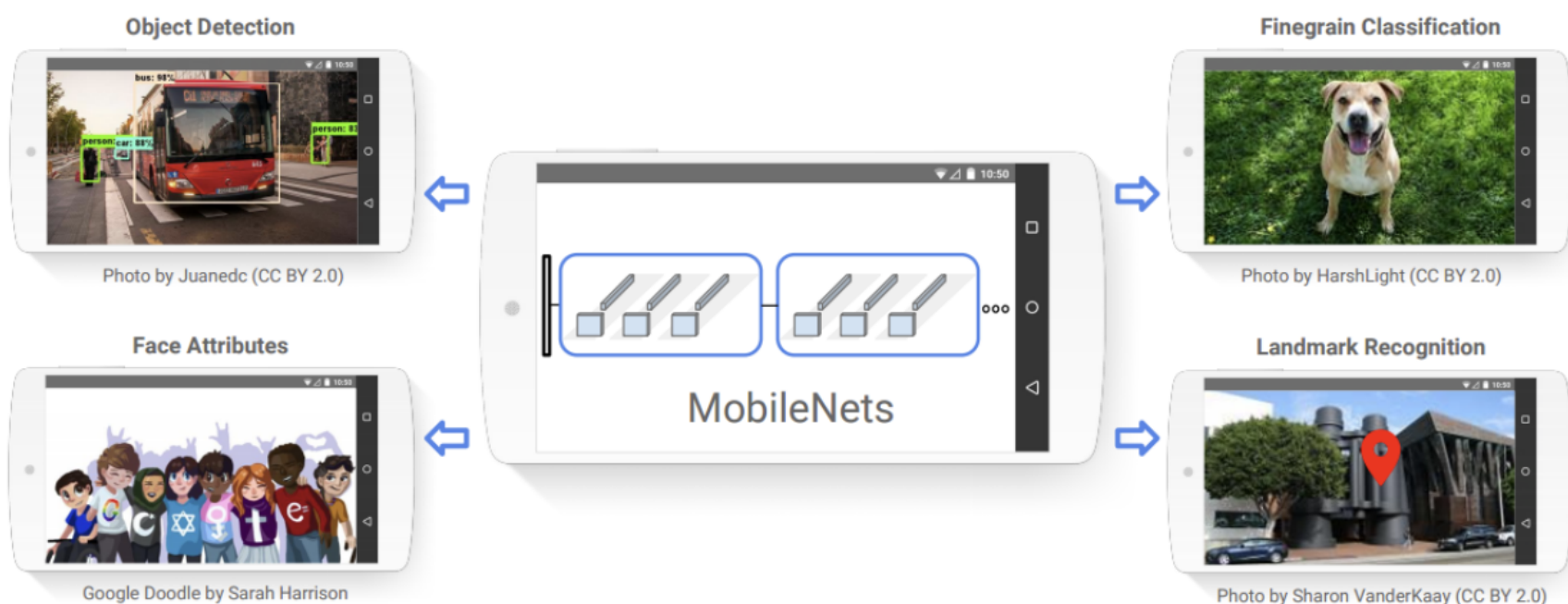


Figure 1. MobileNet models can be applied to various recognition tasks for efficient on device intelligence.

MobileNet은 경량화된 신경망 구조로, 주로 모바일 기기나 엣지 컴퓨팅에서 사용됩니다. MobileNet은 작은 모델 크기와 빠른 추론 속도를 가지고 있으며, 높은 성능까지 유지하는 것이 목표입니다. Depthwise separable convolution이라는 특별한 유형의 컨볼루션을 사용하여 연산량을 줄입니다.

▼ ResNet

ResNet은 Residual Network의 약자로, "잔차 연결(residual connections)"을 통해 깊은 네트워크에서도 효율적으로 학습할 수 있습니다. ResNet은 레이어의 입력을 레이어의 출력에 더하는 간단한 연산을 추가함으로써, 역전파가 더 잘 이루어지게 하고, 따라서 더 깊은 네트워크를 효과적으로 학습할 수 있습니다.

▼ Gpipe

GPipe는 큰 신경망을 효율적으로 학습하기 위한 파이프라인 병렬화 기술입니다. 모델이 너무 클 경우, 단일 하드웨어 리소스로는 학습이 어려울 수 있습니다. GPipe는 이러한 대규모 모델을 여러 하드웨어 리소스에 분할하여 병렬로 학습합니다. 이는 특히 매개변수가 많은 큰 모델에서 유용합니다.

▼ grid search

▼ 기존에 depth, width, resolution 중 주로 하나의 차원만 확장하는 방법을 사용한 이유

2가지 ~ 3개의 차원을 임의로 확장할 경우 최적의 정확도와 효율성은 얻지 못했기 때문에

▼

8. 참고자료

<https://kmhana.tistory.com/26>

[MAT 3편] NAS(Network architecture search)란 무엇일까?

8만 AI 팀이 협업하는 데이터 사이언스 플랫폼. AI 경진대회와 대상 맞춤 온/오프라인 교육, 문제 기반 학습 서비스를 제공합니다.

 <https://dacon.io/en/codeshare/4879>



https://greeksharifa.github.io/computer_vision/2022/03/01/EfficientNet/

1. resnet에 대해서 모르겠다..
2. grid search
3. 전에는 왜 하나씩 증가시켰는지?
4. Max Accuracy에 대한 내용
5. 검증에서 사용된 모델에 간략한 이해정도?