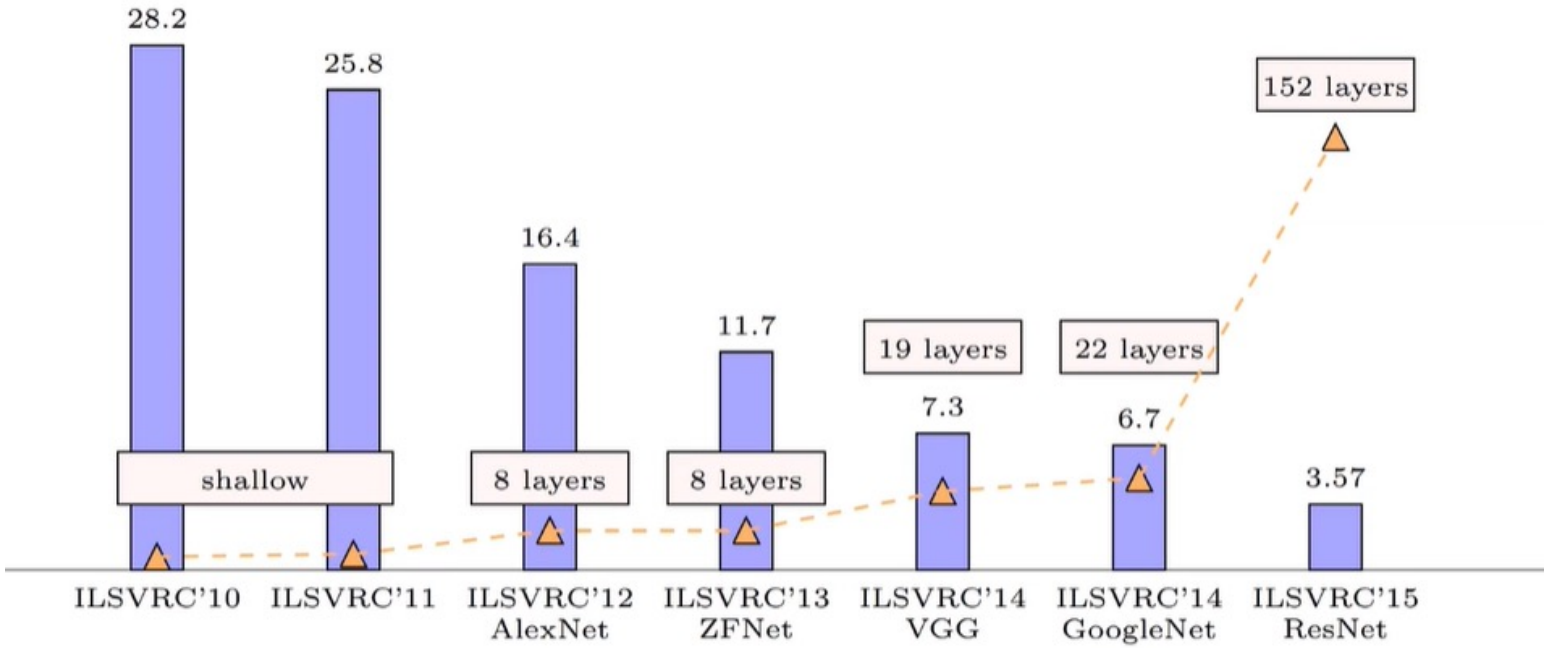
EfficientNet

에이아이스쿨(AISchool) 대표
양진호 (솔라리스)
http://aischool.ai
http://solarisailab.com

연도별 ILSVRC 대회 우승 모델들

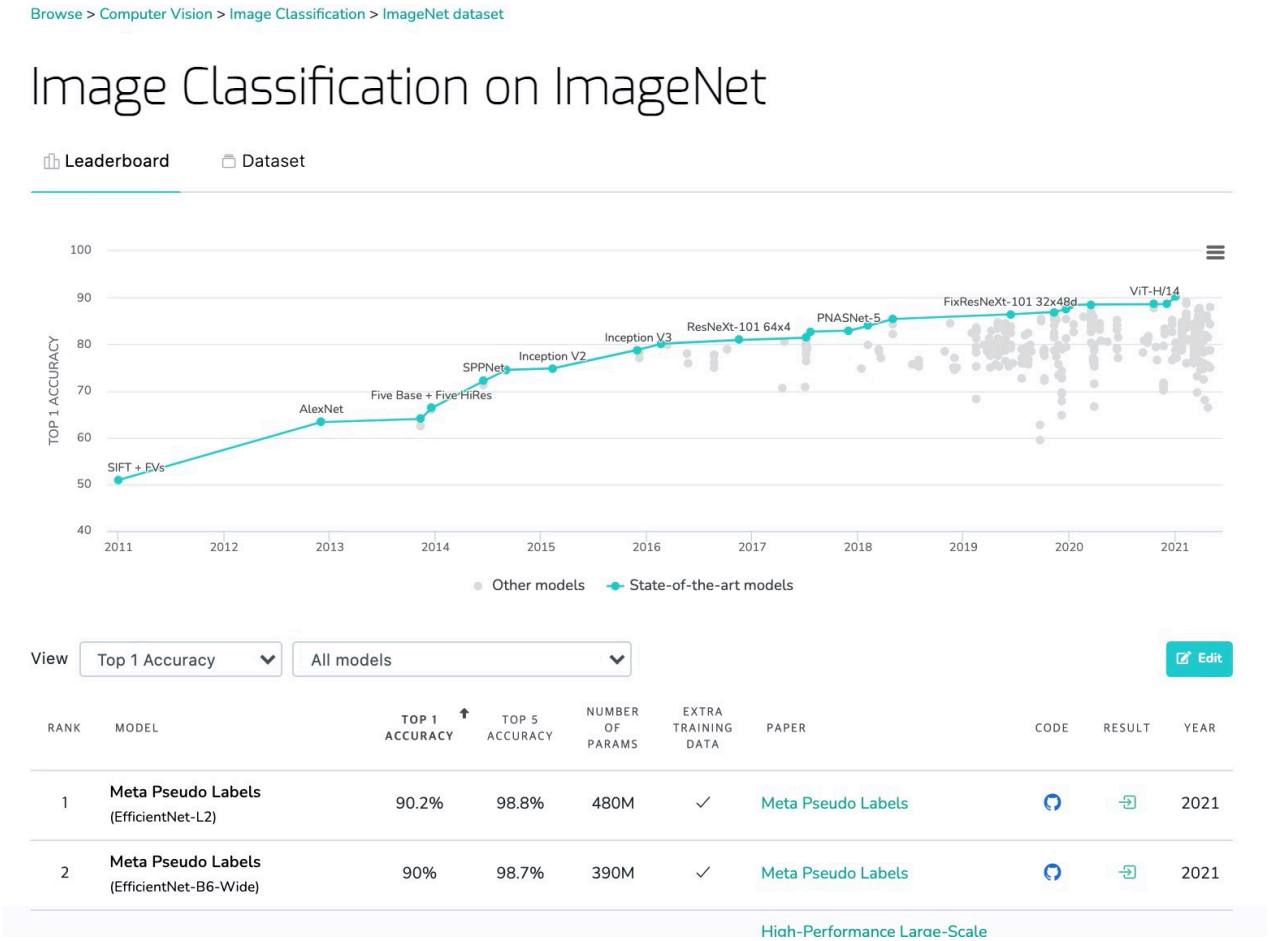
- ResNet은 Layer Depth를 기존 모델 대비 폭발적으로 늘리면서 큰 성능향상을 가져옴



Reference : https://medium.com/analytics-vidhya/7-cnn-architectures-evolved-from-2010-2015-fd5869bd744e

State-of-the-art ImageNet Image Classification

- https://paperswithcode.com/sota/image-classification-on-imagenet



EfficientNet

- Tan, Mingxing, and Quoc V. Le. "Efficientnet: Rethinking model scaling for convolutional neural networks." arXiv preprint arXiv:1905.11946 (2019).
- https://arxiv.org/pdf/1905.11946.pdf

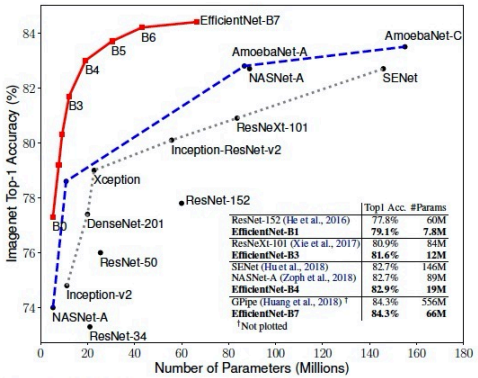
EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks

Mingxing Tan [1] Quoc V. Le [1]

Abstract

Convolutional Neural Networks (ConvNets) are commonly developed at a fixed resource budget, and then scaled up for better accuracy if more resources are available. In this paper, we systematically study model scaling and identify that carefully balancing network depth, width, and resolution can lead to better performance. Based on this observation, we propose a new scaling method that uniformly scales all dimensions of depth/width/resolution using a simple yet highly effective *compound coefficient*. We demonstrate the effectiveness of this method on scaling up MobileNets and ResNet.

To go even further, we use neural architecture search to design a new baseline network and scale it up to obtain a family of models, called *EfficientNets*, which achieve much better accuracy and efficiency than previous ConvNets. In particular, our EfficientNet-B7 achieves state-of-the-art 84.3% top-1 accuracy on ImageNet, while being **8.4x smaller** and



	Top1 Acc.	#Params
ResNet-152 (He et al., 2016)	77.8%	60M
EfficientNet-B1	**79.1%**	**7.8M**
ResNeXt-101 (Xie et al., 2017)	80.9%	84M
EfficientNet-B3	**81.6%**	**12M**
SENet (Hu et al., 2018)	82.7%	146M
NASNet-A (Zoph et al., 2018)	82.7%	89M
EfficientNet-B4	**82.9%**	**19M**
GPipe (Huang et al., 2018) †	84.3%	556M
EfficientNet-B7	**84.3%**	**66M**

†Not plotted

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.

CNN의 성능을 높일 수 있는 요소 : Depth (d), Width (w), Resolution (r)

- 2012년에 AlexNet이 등장한 이후로 후속 연구로 제안된 CNN 구조들은 네트워크를 더욱 깊게 쌓으면서 성능을 향상시켜왔다. 특히, ResNet이 등장한 이후 100-depth 이상의 깊은 CNN도 많이 사용되고 있다.
- 이렇게 CNN의 성능을 높일 수 있는 요소를 저자들은 좀더 구조화해서 아래 4가지로 요소로 분석하고, 각 요소들의 scaling에 따른 성능 변화를 실험해보았다.

① baseline CNN 모델을 선택한다.

② CNN의 Width(필터의 개수)를 늘린다.

③ CNN의 Depth(레이어의 개수)를 늘린다.

④ CNN의 인풋 이미지의 Resolution(크기)를 늘린다.

CNN의 성능을 높일 수 있는 요소 : Depth (d), Width (w), Resolution (r)

- CNN의 성능을 높일 수 있는 요소 – Width, Depth, Resolution

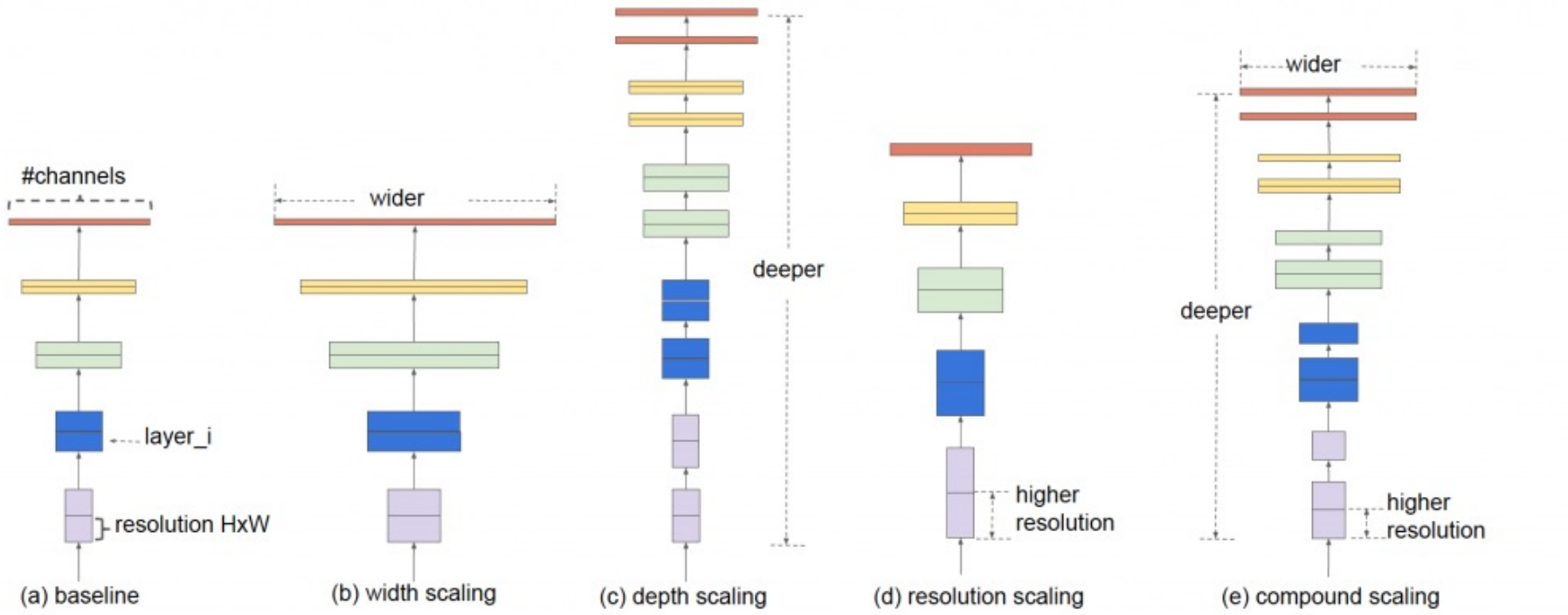


Figure 2. **Model Scaling.** (a) is a baseline network example; (b)-(d) are conventional scaling that only increases one dimension of network width, depth, or resolution. (e) is our proposed compound scaling method that uniformly scales all three dimensions with a fixed ratio.

CNN의 성능을 높일 수 있는 요소 : Depth (d), Width (w), Resolution (r)

- 각각의 요소들을 바꿔가며 실험한 결과, 아래와 같은 결과를 얻을 수 있었다.

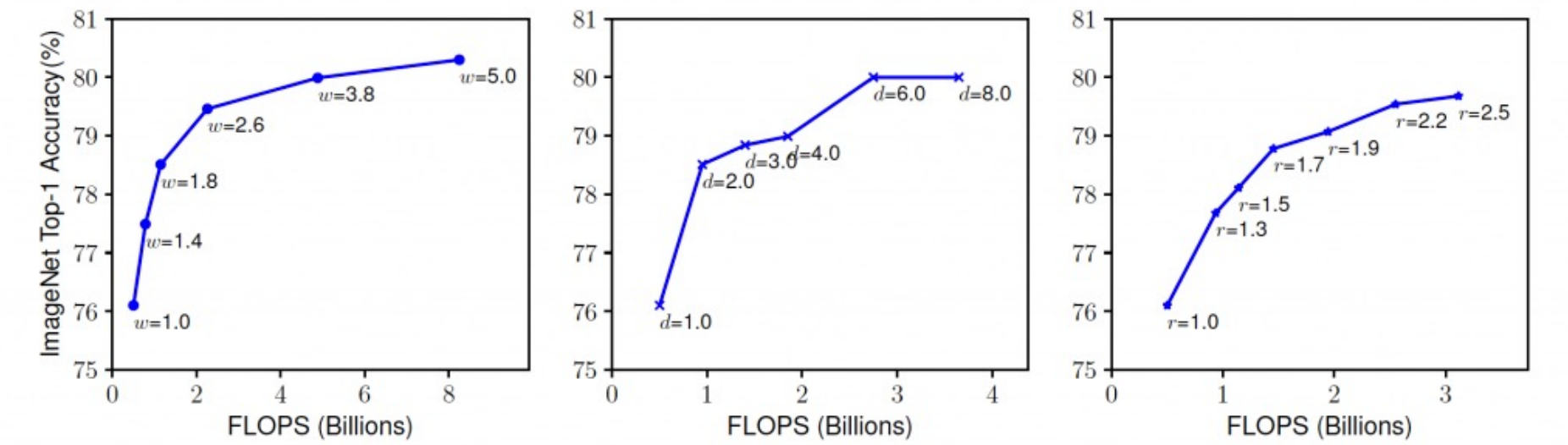


Figure 3. **Scaling Up a Baseline Model with Different Network Width (w), Depth (d), and Resolution (r) Coefficients.** Bigger networks with larger width, depth, or resolution tend to achieve higher accuracy, but the accuracy gain quickly saturate after reaching 80%, demonstrating the limitation of single dimension scaling. Baseline network is described in Table 1.

Compound scaling method

- 저자들은 Width, Depth, Resolution이 서로 독립적이지 않다는 사실에 기반해서 3개의 요소를 조화롭게 scaling할 수 있는 Compound scaling method를 새롭게 제안하였다. Compound scaling method는 아래와 같은 제약조건하에서 CNN의 Width, Depth, Resolution을 함께 늘려준다.

$$\begin{aligned} \text{depth: } d &= \alpha^\phi \\ \text{width: } w &= \beta^\phi \\ \text{resolution: } r &= \gamma^\phi \\ \text{s.t. } \alpha \cdot \beta^2 \cdot \gamma^2 &\approx 2 \\ \alpha \geq 1, \beta &\geq 1, \gamma \geq 1 \end{aligned}$$

- 이때 α, β, γ 값을 적절하게 설정한뒤 ϕ를 원하는 컴퓨팅 리소스를 고려해서 늘려주게 되면, 컴퓨팅 리소스에 비례해서 성능이 증가하는 효율적인 CNN 모델을 만들 수 있다. (EfficientNet-B0(가장 작은 모델) ~ EfficientNet-B7(가장 큰 모델))

Compound scaling method

- 직관적으로 생각해보았을 때, 높은 해상도(Resolution)를 가진 이미지를 CNN의 인풋으로 넣게 되면 그로부터 더욱 잘 정제된 특징을 여러개의 필터들(Width)이 학습할 수 있고, 더 큰 receptive field로부터 더욱더 추상적인 특징을 CNN 레이어(Depth)에서 학습할 수 있다. 따라서 각 요소들을 같이 증가시켜주게되며 더욱 큰 시너지 효과를 기대할 수 있을 것이다.
- 예를 들어, 아래와 같이 Width의 증가에 따른 성능 변화를 실험해본 결과, Depth와 Resolution을 변경하지 않은 상태에서는 금방 Width의 증가에 따른 성능 증가가 saturation되지만 Depth와 Resolution을 같이 늘려줄 경우, 천천히 saturation되는 모습을 볼 수 있다.

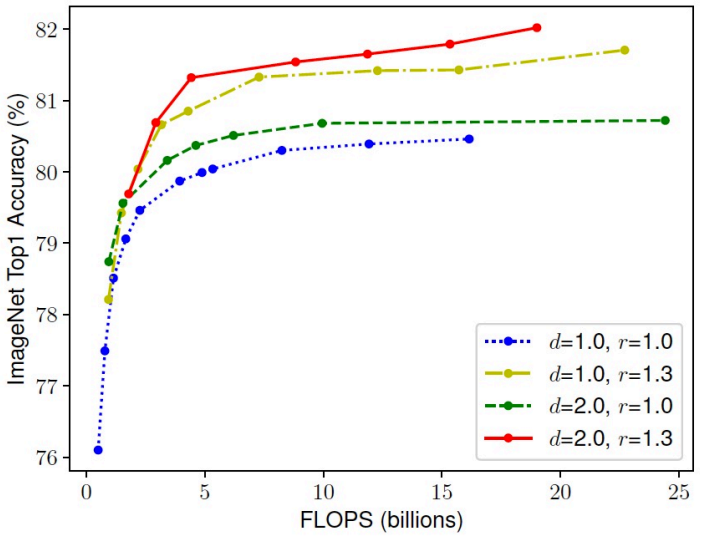


Figure 4. **Scaling Network Width for Different Baseline Networks**. Each dot in a line denotes a model with different width coefficient (w). All baseline networks are from Table 1. The first baseline network (d=1.0, r=1.0) has 18 convolutional layers with resolution 224x224, while the last baseline (d=2.0, r=1.3) has 36 layers with resolution 299x299.

Model Architecture & Experiments

- 확장된 EfficientNet을 만들기 위한 baseline 모델(EfficientNet-B0)은 Neural Architecture Search(NAS)를 이용해서 최적의 구조를 찾아내었다.

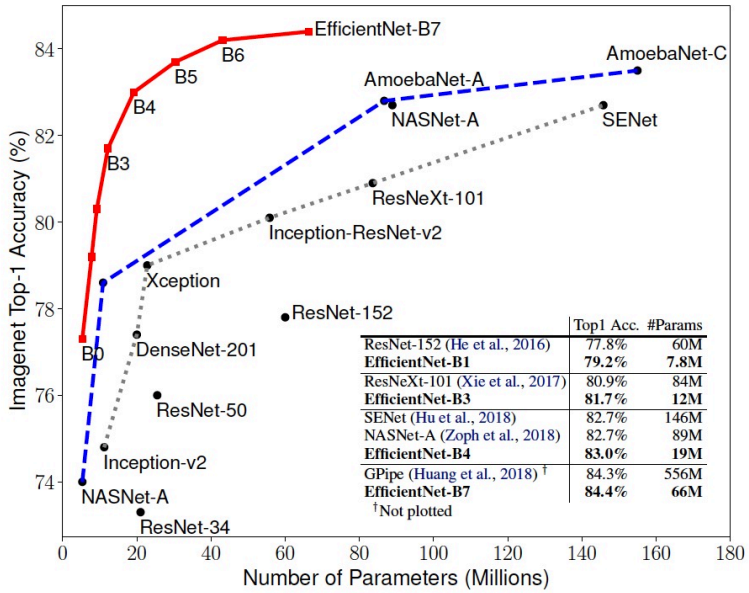
Table 1. **EfficientNet-B0 baseline network** – Each row describes a stage i with $\hat{L}_i$ layers, with input resolution $\langle \hat{H}_i, \hat{W}_i \rangle$ and output channels $\hat{C}_i$. Notations are adopted from equation 2.

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-B0을 ϕ에 비례해서 크게 확장할 경우 Efficient-B1~B7 모델이 된다.

Model Architecture & Experiments

- 이렇게 만든 EfficientNet의 성능을 살펴보면 아래와 같다.
- 보이는 바와 같이 ImageNet 데이터셋에 대해서 기존의 State-Of-The-Art CNN에 비해 더 높은 정확도를 보여주는 모습을 볼 수 있다.



	Top1 Acc.	#Params
ResNet-152 (He et al., 2016)	77.8%	60M
EfficientNet-B1	**79.2%**	**7.8M**
ResNeXt-101 (Xie et al., 2017)	80.9%	84M
EfficientNet-B3	**81.7%**	**12M**
SENet (Hu et al., 2018)	82.7%	146M
NASNet-A (Zoph et al., 2018)	82.7%	89M
EfficientNet-B4	**83.0%**	**19M**
GPipe (Huang et al., 2018) †	84.3%	556M
EfficientNet-B7	**84.4%**	**66M**

†Not plotted

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 84.4% 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.

Model Architecture & Experiments

- 왜 EfficientNet이 잘 동작하는지 CAM(Class Activation Map)으로 시각화해보면 아래와 같다.

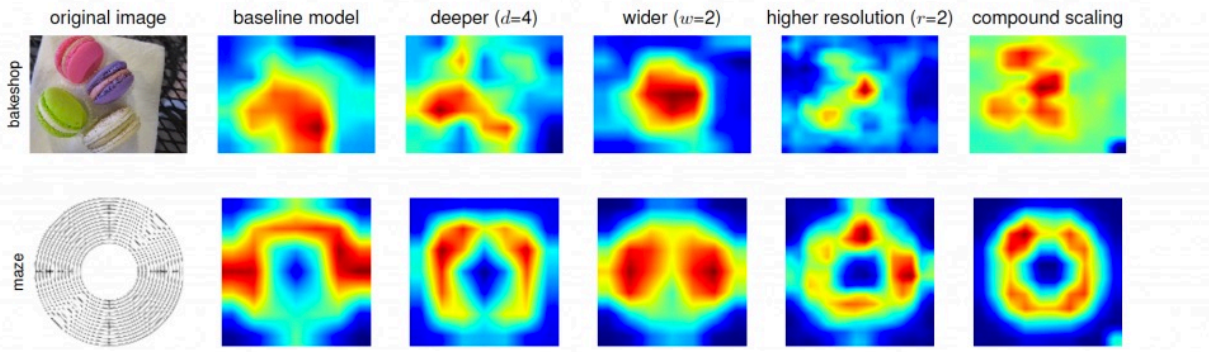


Figure 7. **Class Activation Map (CAM) (Zhou et al., 2016) for Models with different scaling methods-** Our compound scaling method allows the scaled model (last column) to focus on more relevant regions with more object details. Model details are in Table 7.

- 실험 결과를 통해 Width, Depth, Resolution을 조화롭게 증가시켜주는 것이 이미지의 중요특징을 훨씬 정확하게 파악하고 있다는 사실을 알 수 있다.

Table 7. **Scaled Models Used in Figure 7.**

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%**

Model Architecture & Experiments

- 또한 기존 모델들보다 적은 파라미터수를 가지고 있기때문에 매우 빠른 속도로 Inference를 진행할 수 있다. 따라서 실제 문제에 적용할 때 기존 모델에 비해 훨씬 효율적이다.

Table 4. **Inference Latency Comparison** – Latency is measured with batch size 1 on a single core of Intel Xeon CPU E5-2690.

	Acc. @ Latency		Acc. @ Latency
ResNet-152	77.8% @ 0.554s	GPipe	84.3% @ 19.0s
EfficientNet-B1	78.8% @ 0.098s	EfficientNet-B7	84.4% @ 3.1s
Speedup	**5.7x**	**Speedup**	**6.1x**

EfficientNet의 의의

- 적은 수의 파라미터를 가지면서 좋은 성능을 보여주는 State-of-the-art(SOTA) CNN 모델을 제안
- Depth, Width, Resolution을 조화롭게 늘리는 Compound scaling method를 제안

Thank you!