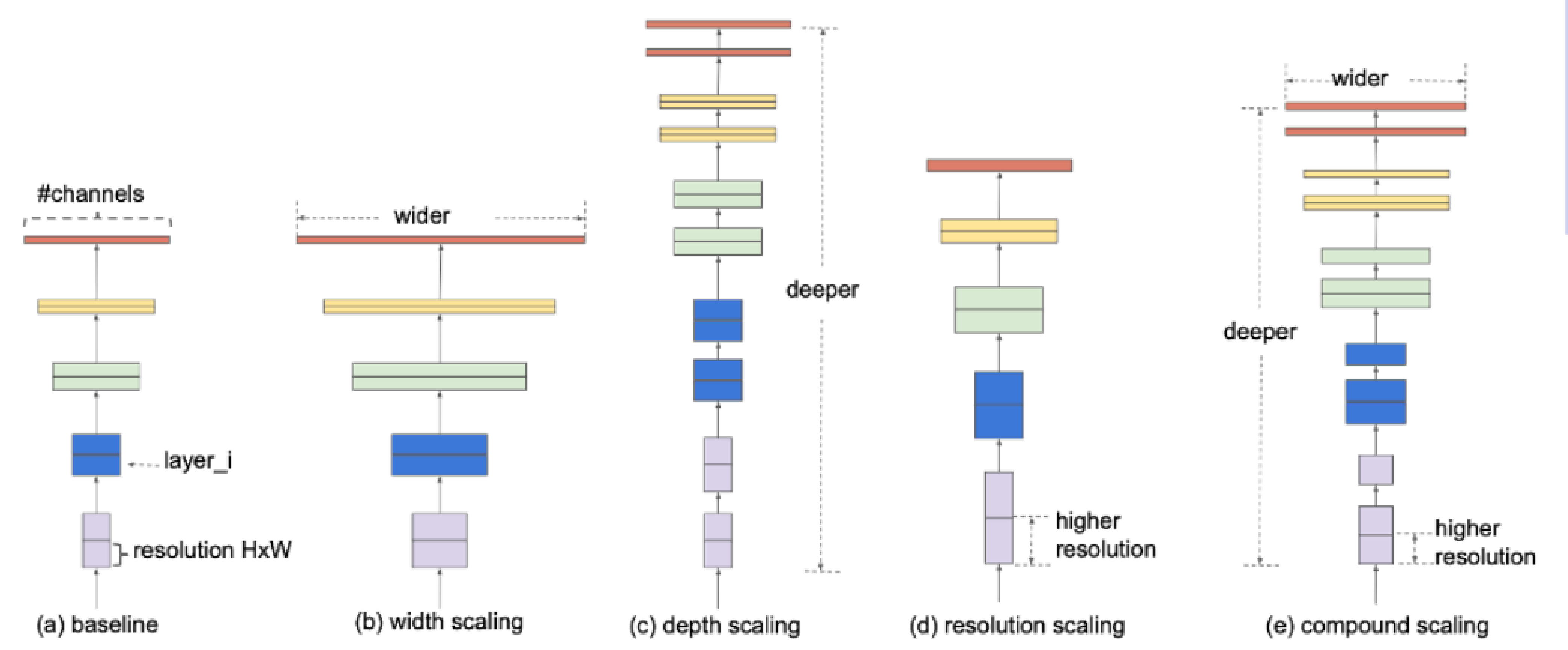
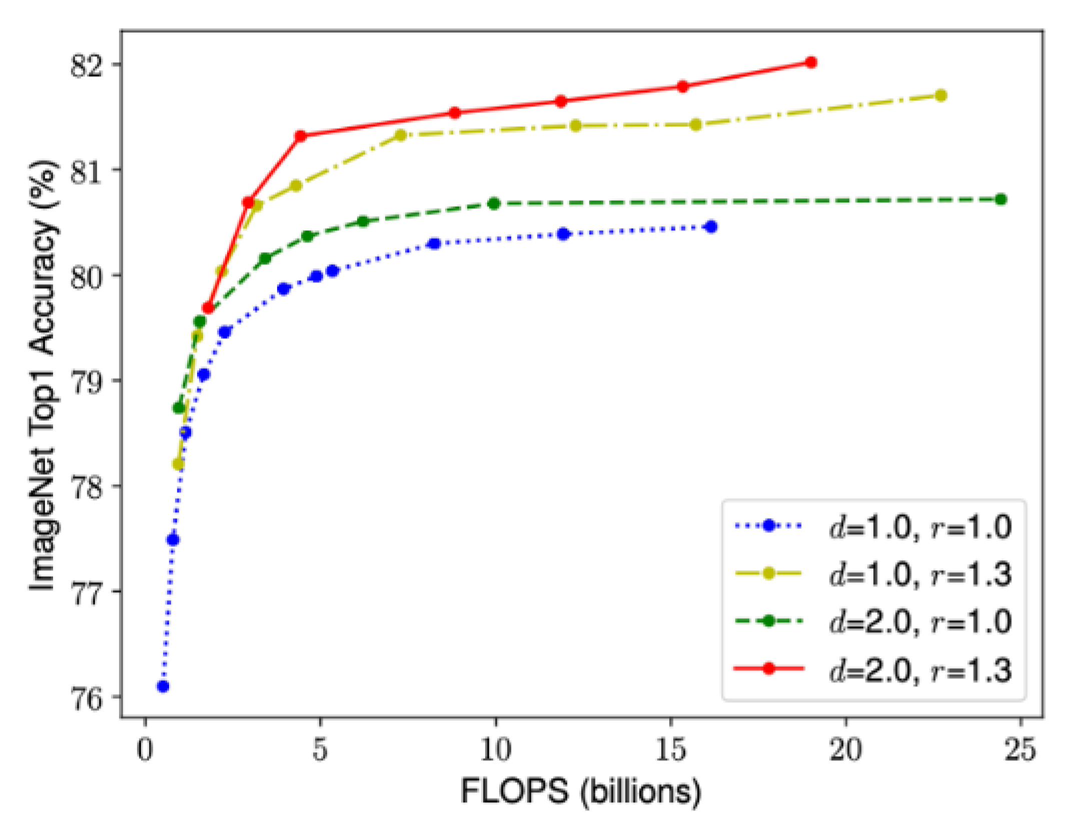
BOAZ 분석 19기 학기 논문 리뷰 세션

EfficientNet

분석 19기 우아라, 정은진





Compound Scaling

depth(d)와 resolution(r)을 고정한 채로 width만 조절하며 정확도의 변화를 측정하는 실험 진행

동일한 FLOPS 에서 width/depth/resolution 조합에 따라 다양한 성능 차이를 보임 → 최적의 조합을 찾아내야 함.

더 높은 해상도의 크기를 조절할 때는 신경망의 깊이가 증가해야 하고, 더욱 세밀한 패턴 포착을 위해 너비도 증가시켜야 함.

Compound Scaling Method

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

α ,β ,γ 는 grid search 를 이용하여 탐색

ϕ 는 사용자가 제어할 수 있는 factor 로 가용한 resource 에 따라 적당한 값을 취함

depth 는 제곱하면 2배의 FLOPS, width, resolution 은 제곱하면 4배의 FLOPS

최종적인 FLOPS는 $(\alpha \cdot \beta^2 \cdot \gamma^2)^{\phi}$ 에 의해 결정

$\alpha \cdot \beta^2 \cdot \gamma^2 \approx 2$ 라는 제약을 사용

→ 최종 FLOPS는 대략 2^{ϕ} 배로 증가

EfficientNet Architecture

$$ACC(m) \times [FLOPS(m)/T]^{w}$$ (m : 모델, T : target FLOPS, w : 정확도와 FLOPS 사이 trade-off)

EfficientNet은 baseline network를 MNasNet과 유사한 구조로 만들었음. (위 수식)

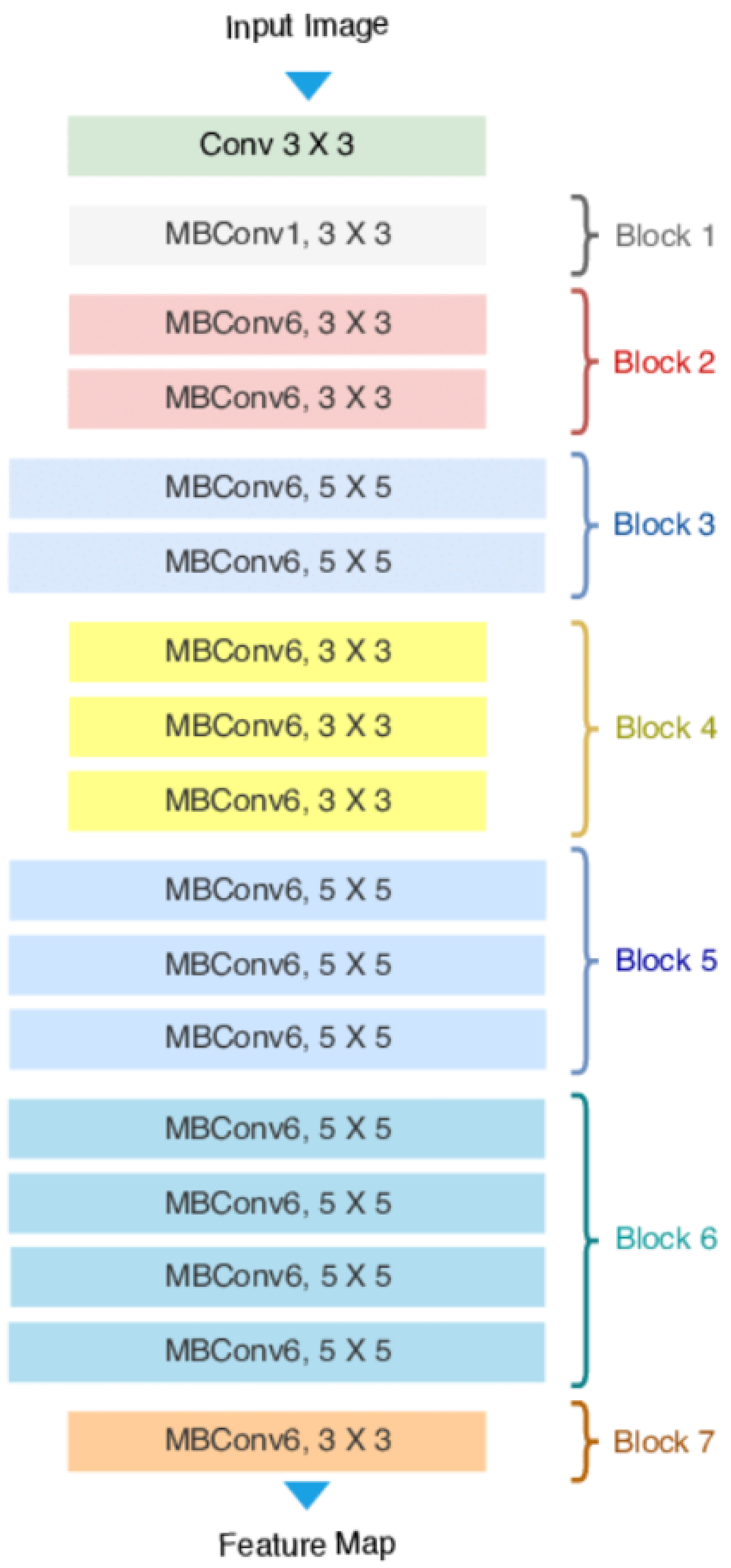
EfficientNet은 MnasNet과 달리 Latency 대신 FLOPS를 최적화 함. (Target hardware device를 정하지 않았기 때문)

이렇게 만들어진 모델이 EfficientNet-B0.

EfficientNet Architecture

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 Architecture

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-B0가 baseline 을 Scaling 하는 방법

STEP1.

ϕ를 1로 고정 후 grid search를 수행하여 a, b, γ 값을 찾음 (α =1.2,β =1.1,γ =1.15)

STEP2.

α , β , γ 를 고정시키고, ϕ를 변화시키면서 EfficientNetB1~B7을 찾음.

Experiments

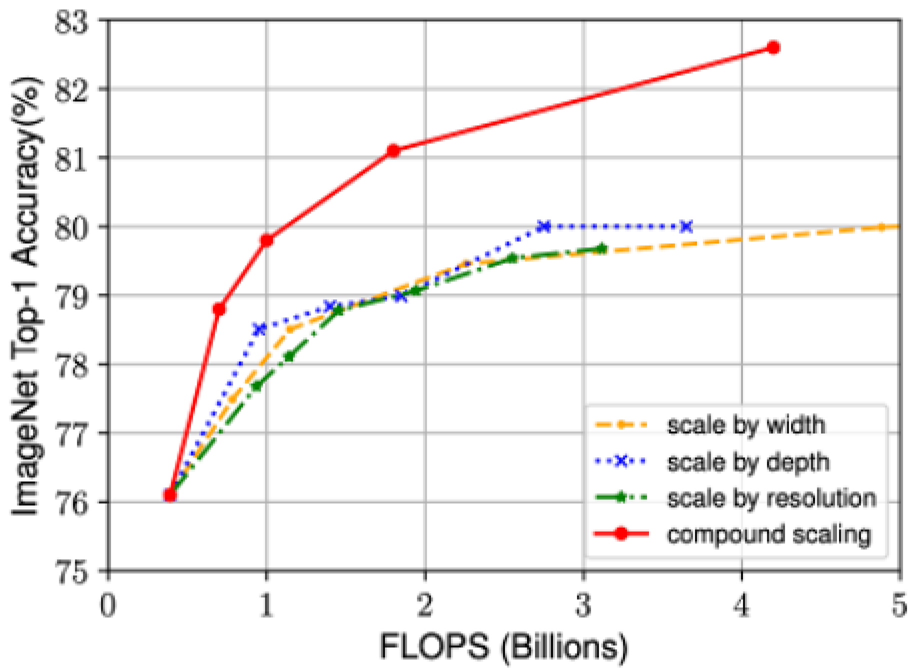


Figure 8. **Scaling Up EfficientNet-B0 with Different Methods.**

EfficientNet-B0에서 single-scaling과 compound scaling기법을 비교한 실험

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%**

MobileNetV1과 MobileNetV2, ResNet-50에 대해 각각 depth / width / resolution / compound scaling을 적용하고 FLOPS를 유사하게 맞춘 실험

→ compound scaling 성능이 월등히 좋음

Experiments

Model	Top-1 Acc.	Top-5 Acc.	#Params	Ratio-to-EfficientNet	#FLOPs	Ratio-to-EfficientNet
EfficientNet-B0	**77.1%**	**93.3%**	**5.3M**	**1x**	**0.39B**	**1x**
ResNet-50 (He et al., 2016)	76.0%	93.0%	26M	4.9x	4.1B	11x
DenseNet-169 (Huang et al., 2017)	76.2%	93.2%	14M	2.6x	3.5B	8.9x
EfficientNet-B1	**79.1%**	**94.4%**	**7.8M**	**1x**	**0.70B**	**1x**
ResNet-152 (He et al., 2016)	77.8%	93.8%	60M	7.6x	11B	16x
DenseNet-264 (Huang et al., 2017)	77.9%	93.9%	34M	4.3x	6.0B	8.6x
Inception-v3 (Szegedy et al., 2016)	78.8%	94.4%	24M	3.0x	5.7B	8.1x
Xception (Chollet, 2017)	79.0%	94.5%	23M	3.0x	8.4B	12x
EfficientNet-B2	**80.1%**	**94.9%**	**9.2M**	**1x**	**1.0B**	**1x**
Inception-v4 (Szegedy et al., 2017)	80.0%	95.0%	48M	5.2x	13B	13x
Inception-resnet-v2 (Szegedy et al., 2017)	80.1%	95.1%	56M	6.1x	13B	13x
EfficientNet-B3	**81.6%**	**95.7%**	**12M**	**1x**	**1.8B**	**1x**
ResNeXt-101 (Xie et al., 2017)	80.9%	95.6%	84M	7.0x	32B	18x
PolyNet (Zhang et al., 2017)	81.3%	95.8%	92M	7.7x	35B	19x
EfficientNet-B4	**82.9%**	**96.4%**	**19M**	**1x**	**4.2B**	**1x**
SENet (Hu et al., 2018)	82.7%	96.2%	146M	7.7x	42B	10x
NASNet-A (Zoph et al., 2018)	82.7%	96.2%	89M	4.7x	24B	5.7x
AmoebaNet-A (Real et al., 2019)	82.8%	96.1%	87M	4.6x	23B	5.5x
PNASNet (Liu et al., 2018)	82.9%	96.2%	86M	4.5x	23B	6.0x
EfficientNet-B5	**83.6%**	**96.7%**	**30M**	**1x**	**9.9B**	**1x**
AmoebaNet-C (Cubuk et al., 2019)	83.5%	96.5%	155M	5.2x	41B	4.1x
EfficientNet-B6	**84.0%**	**96.8%**	**43M**	**1x**	**19B**	**1x**
EfficientNet-B7	**84.3%**	**97.0%**	**66M**	**1x**	**37B**	**1x**
GPipe (Huang et al., 2018)	84.3%	97.0%	557M	8.4x	-	-

We omit ensemble and multi-crop models (Hu et al., 2018), or models pretrained on 3.5B Instagram images (Mahajan et al., 2018).

모든 EfficientNet은 비슷한 성능에서 훨씬 적은 파라미터 보유

특히 EfficienNet-B7은 84.3% 정확도를 달성하면서 Gpipe대비 8.4배 작음

Experiments

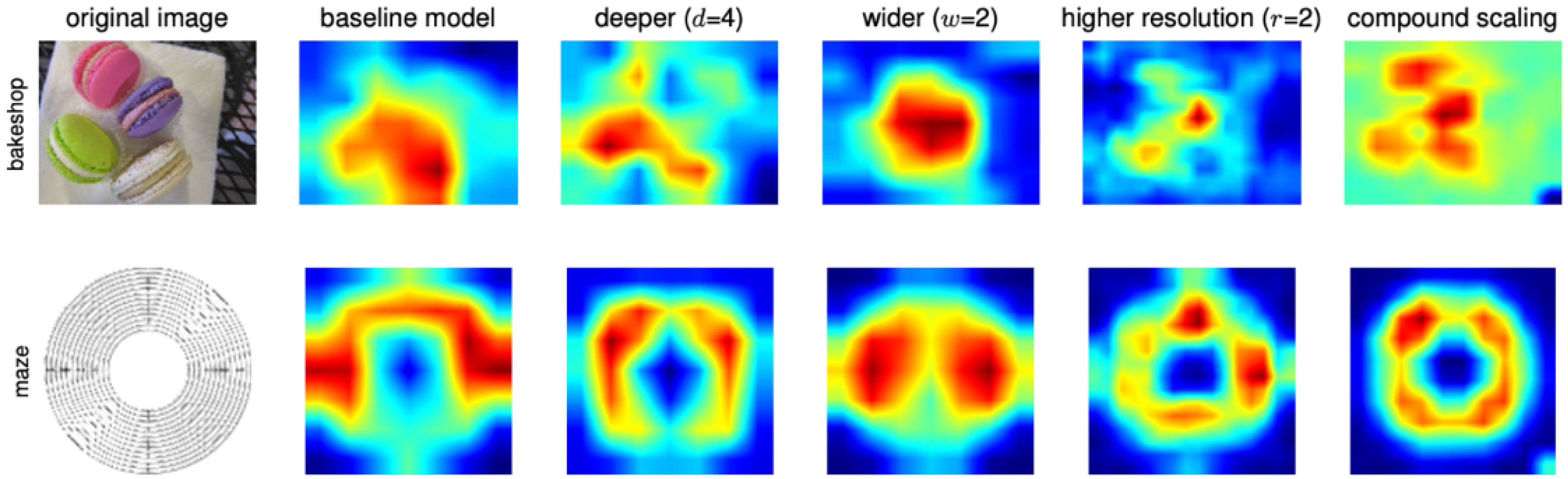


Figure 7. **Class Activation Map (CAM) (Zhou et al., 2016) for Models with different scaling methods-** Our compound scaling method allows the scaled model (last column) to focus on more relevant regions with more object details. Model details are in Table 7.

compound 방법이 다른 방법에 비해서 성능이 좋은 이유를 분석 하기 위해 class activation map을 뽑아 본 것
다른 맵들에 비해 중요한 영역을 잘 찾는 것을 볼 수 있음

Result

~~~~~~~~~~~~은진이 내용~~~~~~~~~~~~~~~~
~~~~~~~~~~~~

THANK YOU!