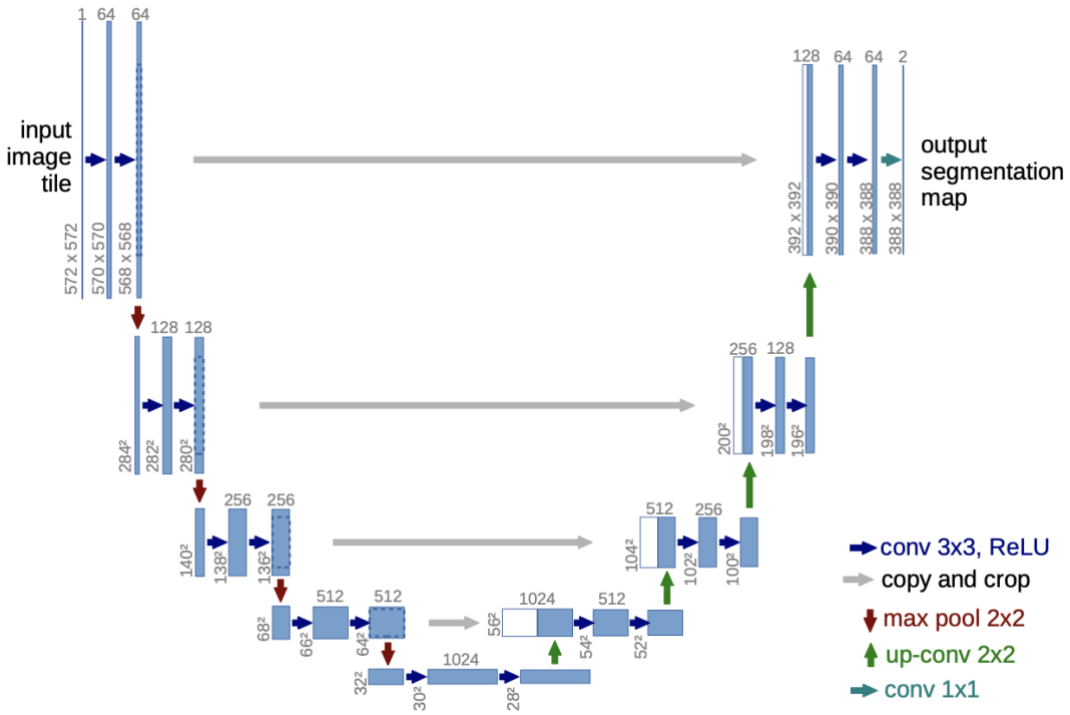
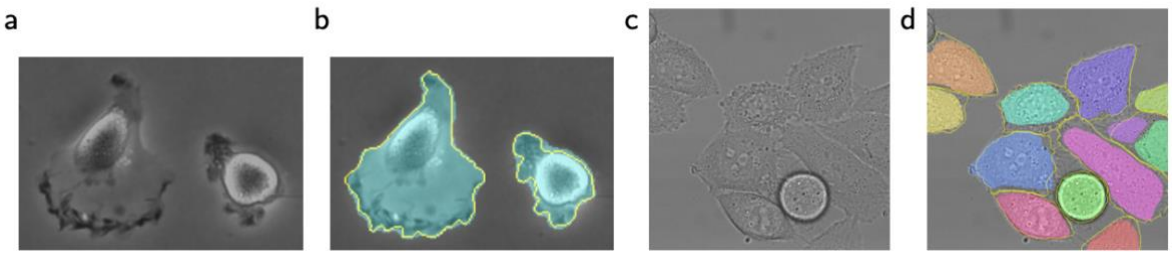
Week19 paper summary

1. U-Net: Convolutional Networks for Biomedical Image Segmentation



U-net 은 Biomedical 분야에서 이미지 분할(image segmentation)을 목적으로 제안된 End-to-End 방식의 fully-convolutional network 기반 모델이다. 이미지의 정보를 얻기 위한 네트워크와 지역화(Localization)을 위한 네트워크가 U 자 대칭 형태로 이루어져 있다. Localization 을 위한 네트워크에서는 높은 해상도의 segmentation 결과를 위해 up-sampling 을 여러 번 진행한다. 또한 얕은 layer 의 feature map 과 깊은 layer 의 feature map 을 결합하는 skip-connection 구조를 FCN 네트워크에서 차용하였다. 이를 통해 convergence time 을 줄일 수 있고 깊은 network 에서 나타나는 gradient vanishing 문제를 해결할 수 있다. 결과적으로는 적은 데이터로도 Biomedical image segmentation 에서 좋은 성능을 보여주었다.



이것이 광학 현미경에서 얻은 이미지로 수행한 cell segmentation task이다. A,c와 같이 얻은 사진을 b,d처럼 잘 구별하는지 알아보는 실험이다.

Name	PhC-U373	DIC-HeLa
IMCB-SG (2014)	0.2669	0.2935
KTH-SE (2014)	0.7953	0.4607
HOUS-US (2014)	0.5323	-
second-best 2015	0.83	0.46
u-net (2015)	**0.9203**	**0.7756**

이것이 실험의 결과이다. A 사진이 PhC-U373 데이터셋에 있던 이미지이고 c가 DIC-HeLa 데이터셋에 있던 이미지이다. 확실히 성능이 높아진 것을 확인할 수 있다.

2. Focal Loss for Dense Object Detection

1) Focal Loss의 필요성

Object Detection에는 크게 두 가지 종류의 알고리즘이 있다. R-CNN 계열의 two-stage detector와 YOLO,SSD 계열의 one-stage detector이다.

Two-stage detector는 먼저 localization을 한 다음에 classification이 순차적으로 이루어지고 one-stage detector는 localization과 classification을 동시에 처리한다. 정확도와 성능으로는 two-stage detector가 좋지만 연산 속도가 오래 걸리는 단점이 있다.

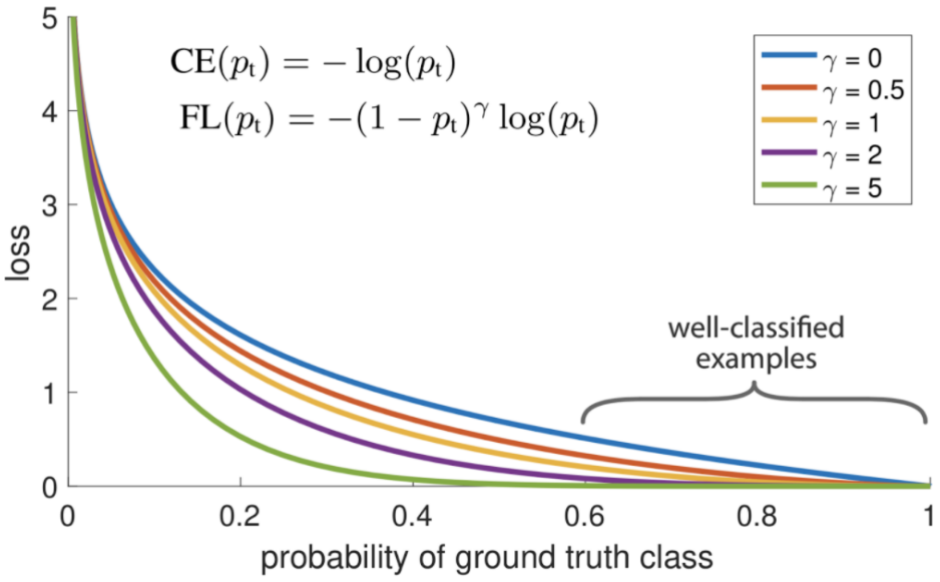
Focal Loss는 one-stage detector의 정확도와 성능을 개선하기 위해 고안되었다.

2) Cross Entropy Loss의 문제점

Cross-entropy loss는 잘 분류한 경우보다 잘못 예측한 경우에 페널티를 부여하는 것에 초점을 맞춘다.

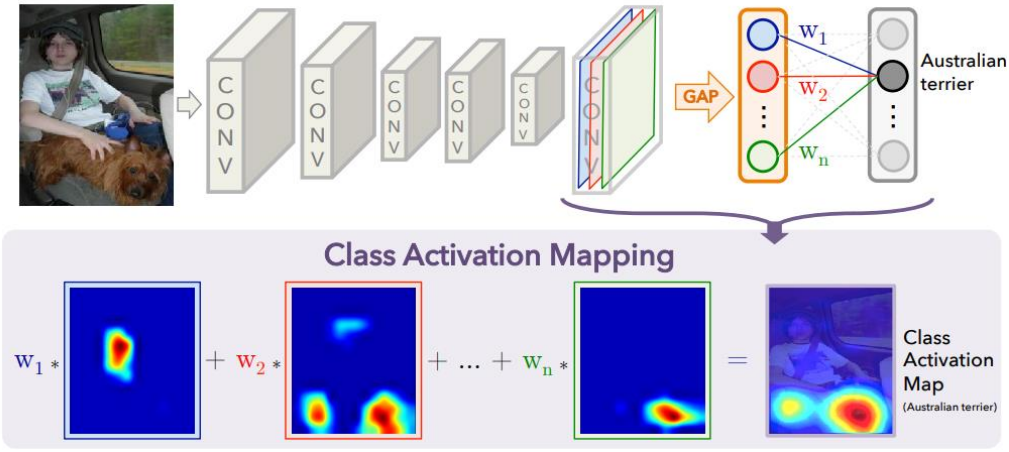
만약 p의 값이 1이면 잘 예측하였음에도 따로 보상이 없다. 다만 페널티가 없어진다.

P를 0에 가깝게 예측하게 되면 페널티는 무한대로 굉장히 커지게 된다.



위의 그래프는 Focal Loss를 나타낸다. 식을 비교하면 (1-p(t))^(gamma) term이 추가된 것을 확인할 수 있다. 기본적인 cross entropy는 감마가 0일 때이다.
추가된 식의 역할은 easy example에 사용되는 Loss의 가중치를 줄이기 위함이다.

3. learning deep features for discriminative localization



1) Global Average Pooing(GAP)

GAP의 개념은 pooling을 알고 있다면 쉽게 이해할 수 있다. 보통은 max pooling이 가장 유명한 개념인데 GAP는 평균값을 추출하여 요약하는 pooling layer이다. 이는 overfitting을 방지하기 위해 Regularization 장치로 연구되었다. GAP의 경우 전체 object의 Localization을 식별하며 average pooling이 모든 discriminative regions의 식별을 장려하는 반면 GAP는 이미지의 가장 discriminative한 regions를 식별한다는 점에서 차이가 있다.

2) Class Activatioin Map(CAM)

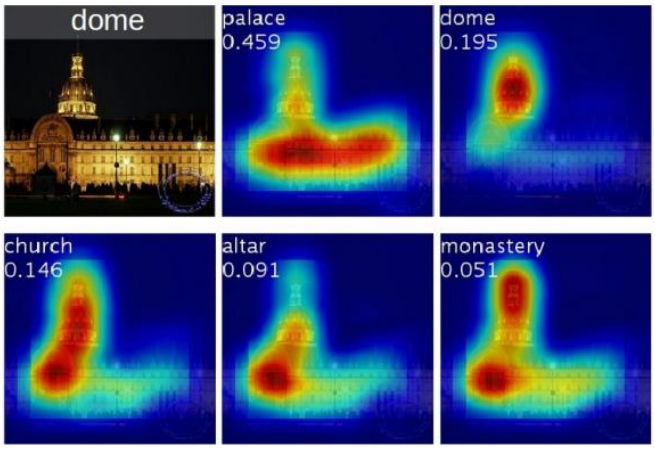


Figure 4. Examples of the CAMs generated from the top 5 predicted categories for the given image with ground-truth as dome. The predicted class and its score are shown above each class activation map. We observe that the highlighted regions vary across predicted classes e.g., *dome* activates the upper round part while *palace* activates the lower flat part of the compound.

CAM은 마지막 convolutional feature map의 내적과 FC layer의 class별 weight의 합으로 설명할 수 있다. 또한 마지막 convolutional feature map의 해상도에 따라 CAM은 input image의 discriminative regions를 시각화하기 위해 upsamping하며 softmax/sigmoid output function을 통해 전달되어 Image-level classification label을 생성한다.

3) Experiments

Table 1. Classification error on the ILSVRC validation set.

Networks	top-1 val. error	top-5 val. error
VGGnet-GAP	33.4	12.2
GoogLeNet-GAP	35.0	13.2
AlexNet*-GAP	44.9	20.9
AlexNet-GAP	51.1	26.3
GoogLeNet	31.9	11.3
VGGnet	31.2	11.4
AlexNet	42.6	19.5
NIN	41.9	19.6
GoogLeNet-GMP	35.6	13.9

이것은 conv layer를 제거한 후의 실험 결과이다. Top-5 error와 top-1 error가 약간 증가한 모습을 볼 수 있다.

Table 2. Localization error on the ILSVRC validation set. *Back-prop* refers to using [22] for localization instead of CAM.

Method	top-1 val.error	top-5 val. error
GoogLeNet-GAP	**56.40**	**43.00**
VGGnet-GAP	57.20	45.14
GoogLeNet	60.09	49.34
AlexNet*-GAP	63.75	49.53
AlexNet-GAP	67.19	52.16
NIN	65.47	54.19
Backprop on GoogLeNet	61.31	50.55
Backprop on VGGnet	61.12	51.46
Backprop on AlexNet	65.17	52.64
GoogLeNet-GMP	57.78	45.26

Table 3. Localization error on the ILSVRC test set for various weakly- and fully- supervised methods.

Method	supervision	top-5 test error
GoogLeNet-GAP (heuristics)	weakly	**37.1**
GoogLeNet-GAP	weakly	42.9
Backprop [22]	weakly	46.4
GoogLeNet [24]	full	26.7
OverFeat [21]	full	29.9
AlexNet [24]	full	34.2

이 표들은 CAM 방법이 이전의 방법들보다 더 잘 localize하는 것을 보여준다.

4. EfficientNet: Rethinking Model Scaling for Convolutional Neural Net work

1) Abstract

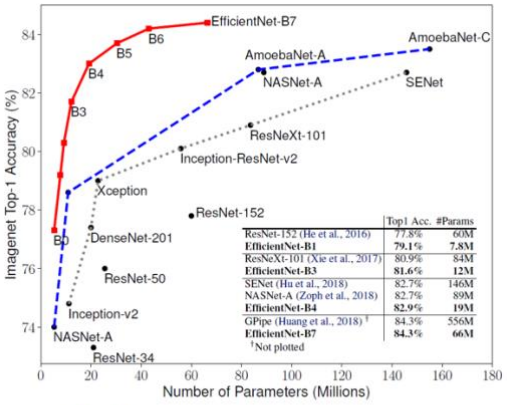


Figure 1. **Model Size vs. ImageNet Accuracy.** All numbers are for single-crop, single-model. Our EfficientNets significantly outperform other ConvNets. In particular, EfficientNet-B7 achieves new state-of-the-art 84.3% top-1 accuracy but being 8.4x smaller and 6.1x faster than GPipe. EfficientNet-B1 is 7.6x smaller and 5.7x faster than ResNet-152. Details are in Table 2 and 4.

한정된 자원으로 최대의 효율을 내기 위해 Model scaling을 시스템적으로 분석하여 더 나은 성능을 얻고자 한다. 따라서 scaling 방법으로 compound coefficient를 제안한다.

이를 바탕으로 찾은 모델인 EfficientNet은 기존의 ConvNet보다 8.4배 작으면서 6.1배 빠르고 더 높은 정확도를 갖는다.

2) 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

이 baseline network에 기반해서 시작한다. 작은 baseline network에 대해서 먼저 좋은 알파,베타, 감마 값을 찾고 그 다음에 전체적인 크기를 키운다.

3) Experiments

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**

전체적으로 8.4배 적은 연산량으로 더 높은 정확도를 가지는 것을 알 수 있다.

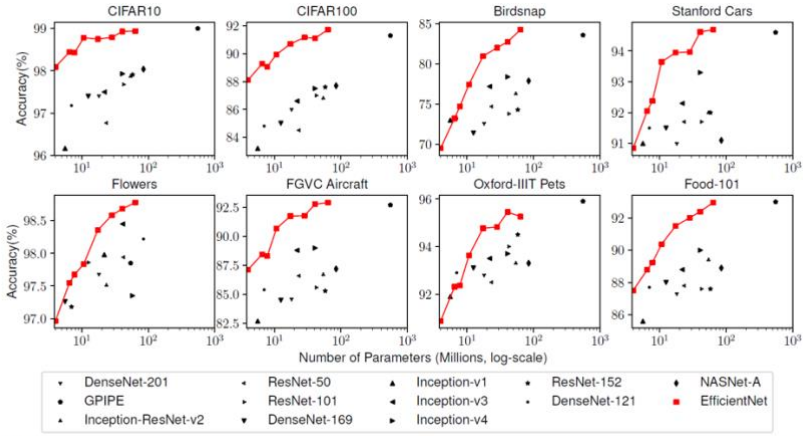


Figure 6. **Model Parameters vs. Transfer Learning Accuracy –** All models are pretrained on ImageNet and finetuned on new datasets.

어떤 결과는 18배 적거나 5.7배 등의 더 빠른 추론 시간을 보여주기도 한다.