

ILSVRC와 표준 CNN 모델들

에이아이스쿨(AISchool) 대표
양진호 (솔라리스)

<http://aischool.ai>

<http://solarisailab.com>

표준 CNN 모델 학습의 필요성

- CNN 모델의 **무수히 많은 디자인 요소들** (e.g. 레이어의 depth는 몇 depth로 할건지? 필터 사이즈는? Pooling 사이즈는?)
- **개개인이 모든 디자인 요소들의 가능성을 탐구하고 분석**하는 것은 현실적으로 불가능
- 따라서 **잘 동작하는 것이 검증된 표준 CNN 모델**을 사용하는 것이 필요함

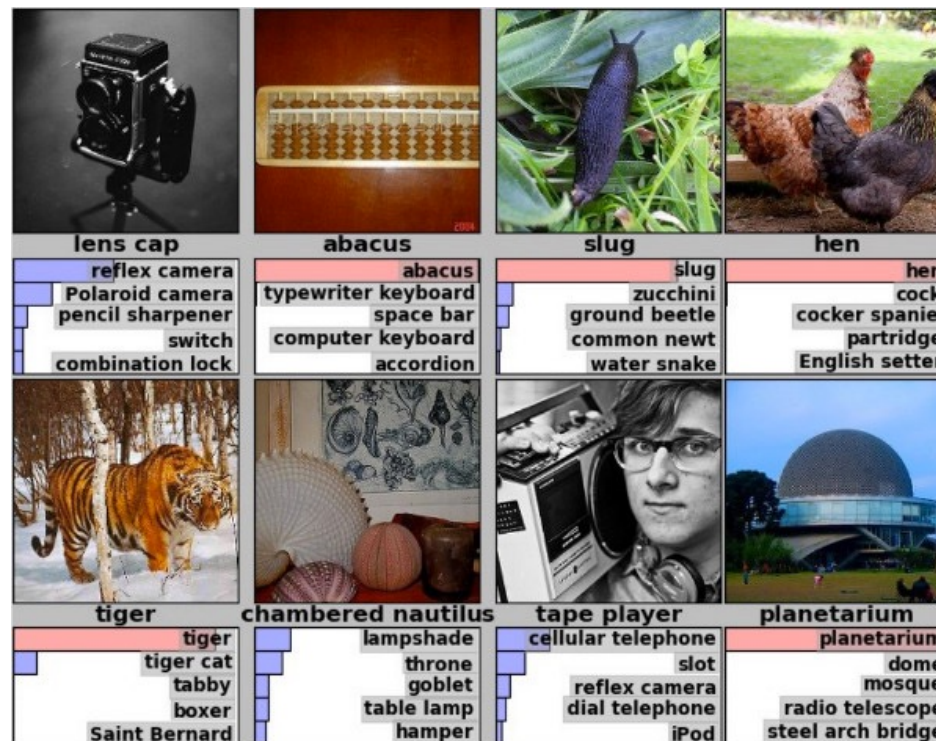
ImageNet Large-Scale Visual Recognition Challenge(ILSVRC)

- ImageNet Large-Scale Visual Recognition Challenge(ILSVRC)
- ILSVRC는 ImageNet의 데이터베이스의 일부분을 사용한다. **1000개의 카테고리**를 가지고 있고 **120만장의 training images**과 **5만장의 validation images**, **15만장의 testing images**로 구성되어 있다.



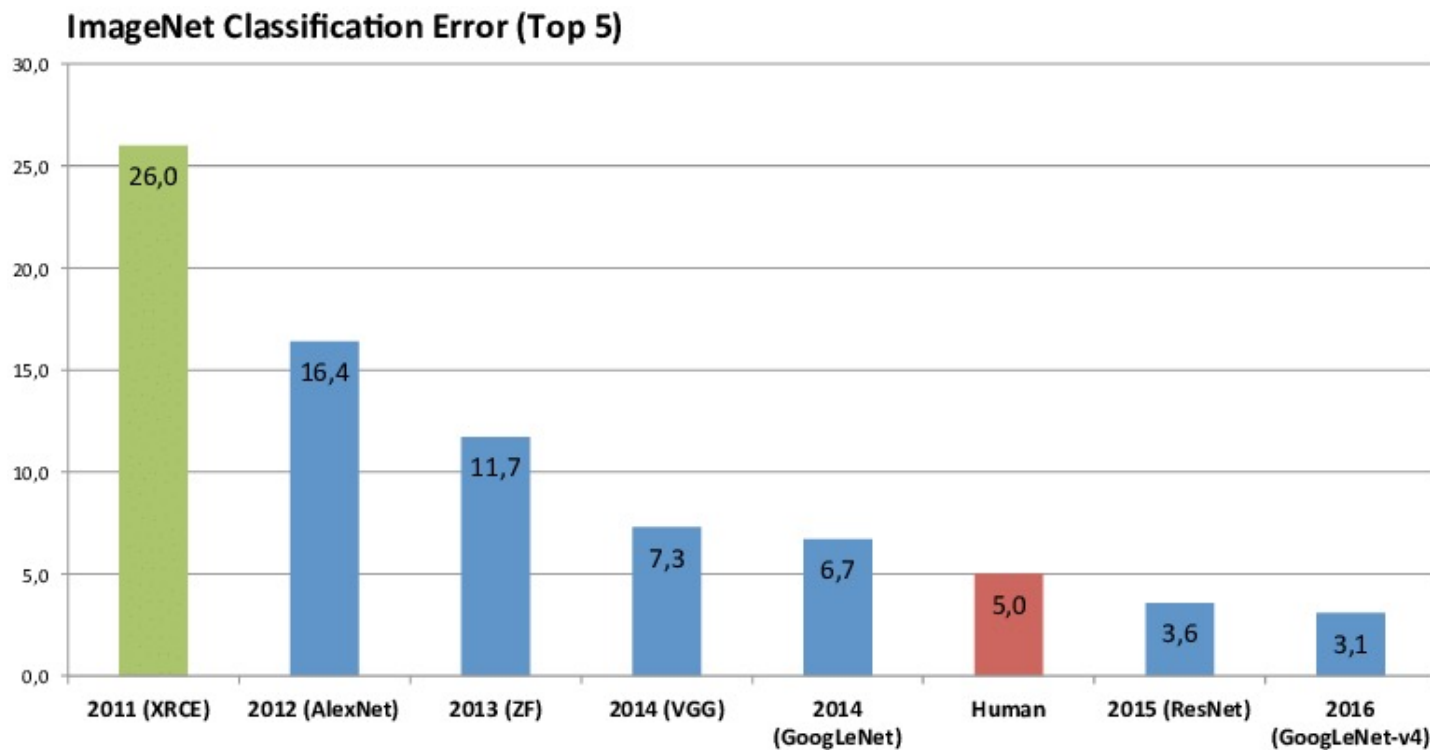
대회 성능 측정 방식 : Top-1 & Top-5

- ILSVRC 대회는 2개의 성능측정 방식이 존재합니다.
- Top-1** : 1000개의 Softmax 예측값중에 가장 확신의 정도가 큰 1개의 값이 정답과 일치하는지를 비교
- Top-5** : 1000개의 Softmax 예측값중에 가장 확신의 정도가 큰 5개의 값에 정답이 포함되어 있는지를 비교



연도별 ILSVRC 대회 우승 모델들

- 2012년 이전에는 CNN 외 기법들이 우승을 차지함
- 2012년 AlexNet이 큰 성능 gap을 만들면서 우승을 차지하면서 이후 대부분의 참가자들이 CNN 모델을 사용함
- 표준 CNN 모델들은 ILSVRC 대회에서 우승을 하거나 준우승을 차지한 모델들을 지칭함



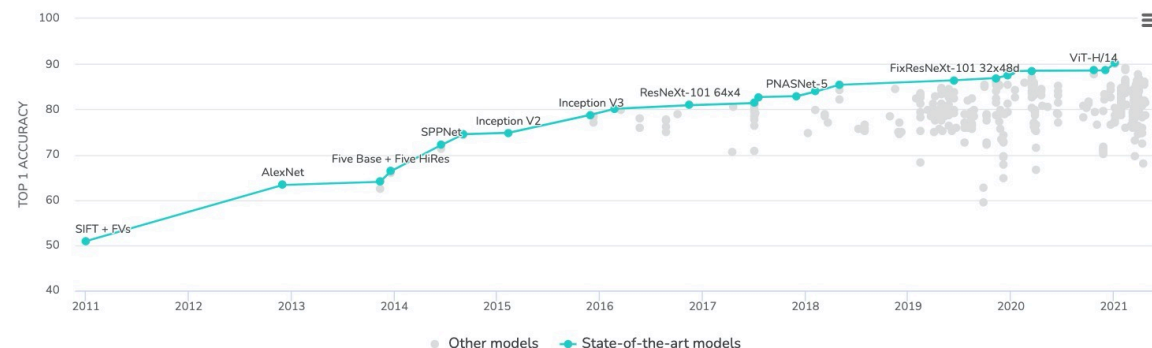
연도별 ILSVRC 대회 우승 모델들

- 최근 state-of-the-art(2021년 5월 기준)은 EfficientNet에 기반한 모델임
- <https://paperswithcode.com/sota/image-classification-on-imagenet>

Browse > Computer Vision > Image Classification > ImageNet dataset

Image Classification on ImageNet

Leaderboard Dataset



View Top 1 Accuracy

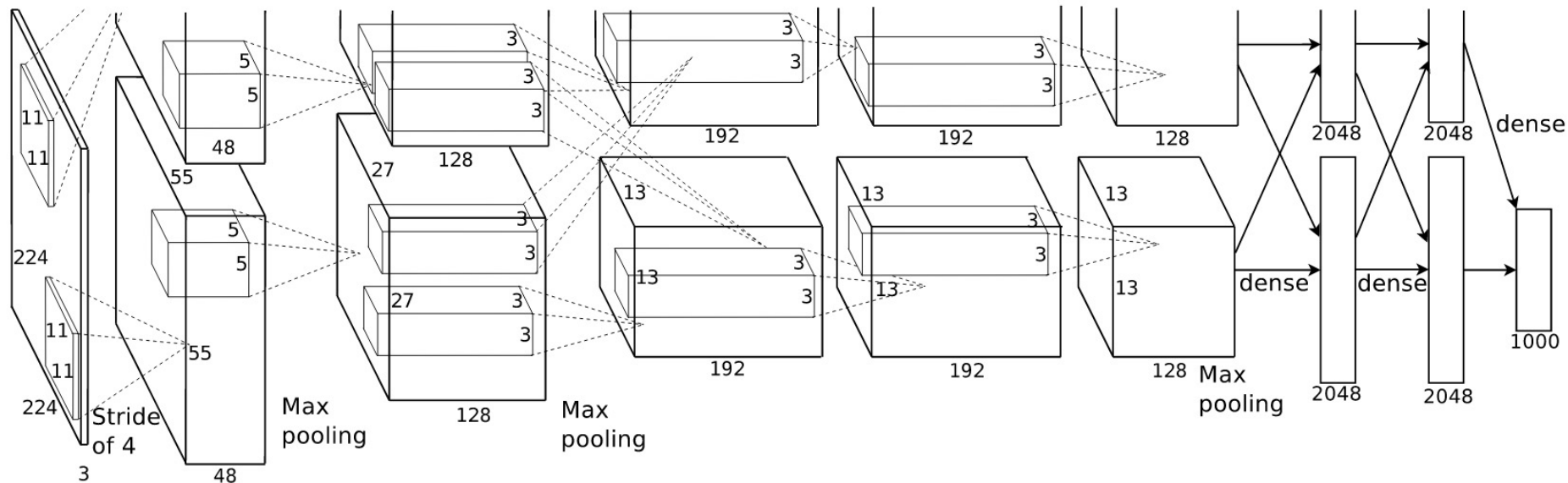
All models

Edit

RANK	MODEL	TOP 1 ACCURACY ↑	TOP 5 ACCURACY	NUMBER OF PARAMS	EXTRA TRAINING DATA	PAPER	CODE	RESULT	YEAR
1	Meta Pseudo Labels (EfficientNet-L2)	90.2%	98.8%	480M	✓	Meta Pseudo Labels	Code	Result	2021
2	Meta Pseudo Labels (EfficientNet-B6-Wide)	90%	98.7%	390M	✓	Meta Pseudo Labels	Code	Result	2021

AlexNet (2012년)

- 2012년 LeNet의 Scale을 키운 AlexNet을 제안
- ILSVRC-2012 1등



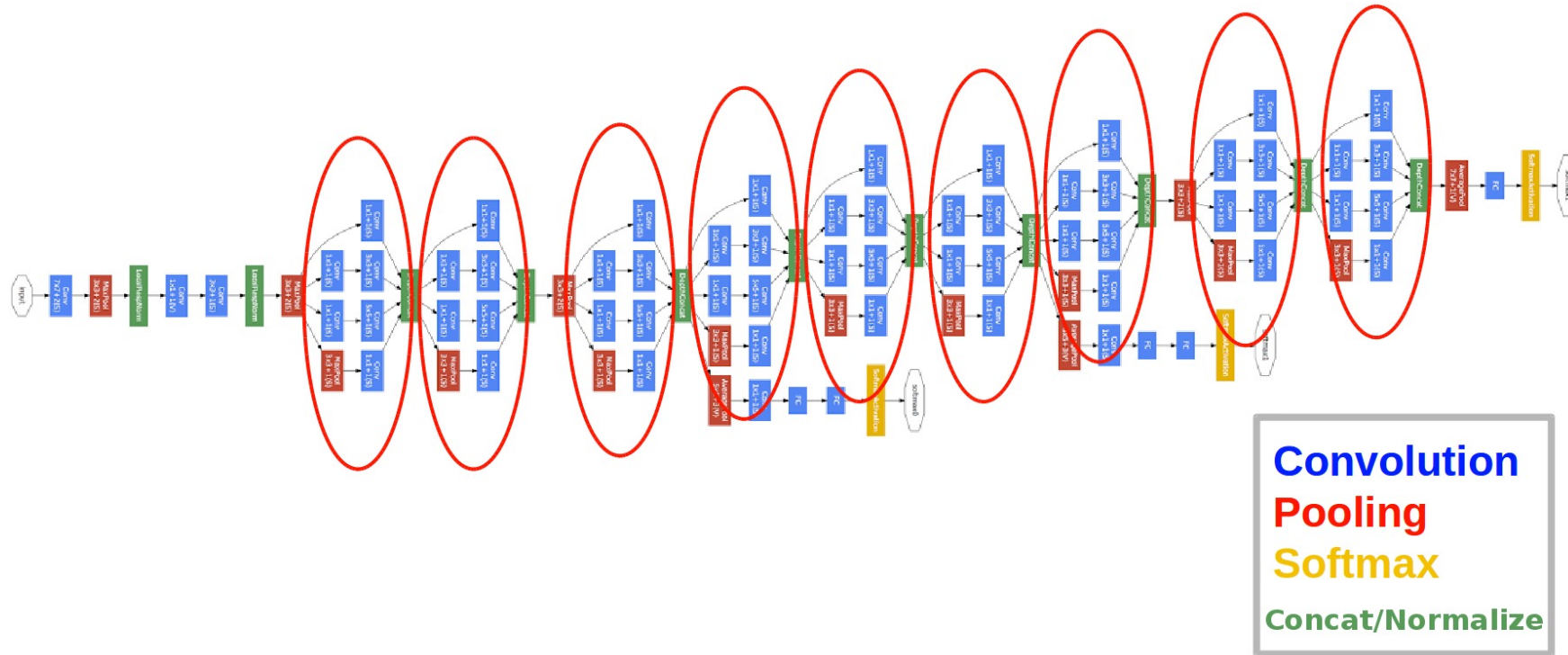
VGGnet (2014년)

- ILSVRC-2014 2등

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 LRN	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 conv1-256	conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256 conv3-256
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

GoogLeNet(Inception v1) (2014년)

- ILSVRC-2014 1등



TensorFlow에서 제공하는 표준 CNN 모델들 - tf.keras.applications

- 표준 CNN 모델들의 구현 Class와 ImageNet 데이터에 대한 Pre-Trained 모델을 TensorFlow에서 기본적으로 제공하고 있습니다.
- <https://keras.io/api/applications/>

Available models

Model	Size	Top-1 Accuracy	Top-5 Accuracy	Parameters	Depth
Xception	88 MB	0.790	0.945	22,910,480	126
VGG16	528 MB	0.713	0.901	138,357,544	23
VGG19	549 MB	0.713	0.900	143,667,240	26
ResNet50	98 MB	0.749	0.921	25,636,712	-
ResNet101	171 MB	0.764	0.928	44,707,176	-
ResNet152	232 MB	0.766	0.931	60,419,944	-
ResNet50V2	98 MB	0.760	0.930	25,613,800	-
ResNet101V2	171 MB	0.772	0.938	44,675,560	-
ResNet152V2	232 MB	0.780	0.942	60,380,648	-
InceptionV3	92 MB	0.779	0.937	23,851,784	159
InceptionResNetV2	215 MB	0.803	0.953	55,873,736	572
MobileNet	16 MB	0.704	0.895	4,253,864	88
MobileNetV2	14 MB	0.713	0.901	3,538,984	88
DenseNet121	33 MB	0.750	0.923	8,062,504	121
DenseNet169	57 MB	0.762	0.932	14,307,880	169
DenseNet201	80 MB	0.773	0.936	20,242,984	201
NASNetMobile	23 MB	0.744	0.919	5,326,716	-
NASNetLarge	343 MB	0.825	0.960	88,949,818	-
EfficientNetB0	29 MB	-	-	5,330,571	-
EfficientNetB1	31 MB	-	-	7,856,239	-
EfficientNetB2	36 MB	-	-	9,177,569	-
EfficientNetB3	48 MB	-	-	12,320,535	-
EfficientNetB4	75 MB	-	-	19,466,823	-
EfficientNetB5	118 MB	-	-	30,562,527	-
EfficientNetB6	166 MB	-	-	43,265,143	-
EfficientNetB7	256 MB	-	-	66,658,687	-

표준 CNN 모델들의 학습의 필요성

- 표준 CNN 모델들은 다른 컴퓨터 비전 문제영역 해결을 위한 모델들의 Backbone으로 기본적으로 사용되기 때문에 원리를 이해하는 것이 좋습니다.
- Backbone을 어떤 모델을 선택하느냐에 따라 모델의 성능이 달라질 수 있습니다.

Faster R-CNN ResNet50 V1 640x640	53	29.3	Boxes
Faster R-CNN ResNet50 V1 1024x1024	65	31.0	Boxes
Faster R-CNN ResNet50 V1 800x1333	65	31.6	Boxes
Faster R-CNN ResNet101 V1 640x640	55	31.8	Boxes
Faster R-CNN ResNet101 V1 1024x1024	72	37.1	Boxes
Faster R-CNN ResNet101 V1 800x1333	77	36.6	Boxes
Faster R-CNN ResNet152 V1 640x640	64	32.4	Boxes
Faster R-CNN ResNet152 V1 1024x1024	85	37.6	Boxes
Faster R-CNN ResNet152 V1 800x1333	101	37.4	Boxes
Faster R-CNN Inception ResNet V2 640x640	206	37.7	Boxes
Faster R-CNN Inception ResNet V2 1024x1024	236	38.7	Boxes

Thank you!
