

GoogLeNet(Inception V1)

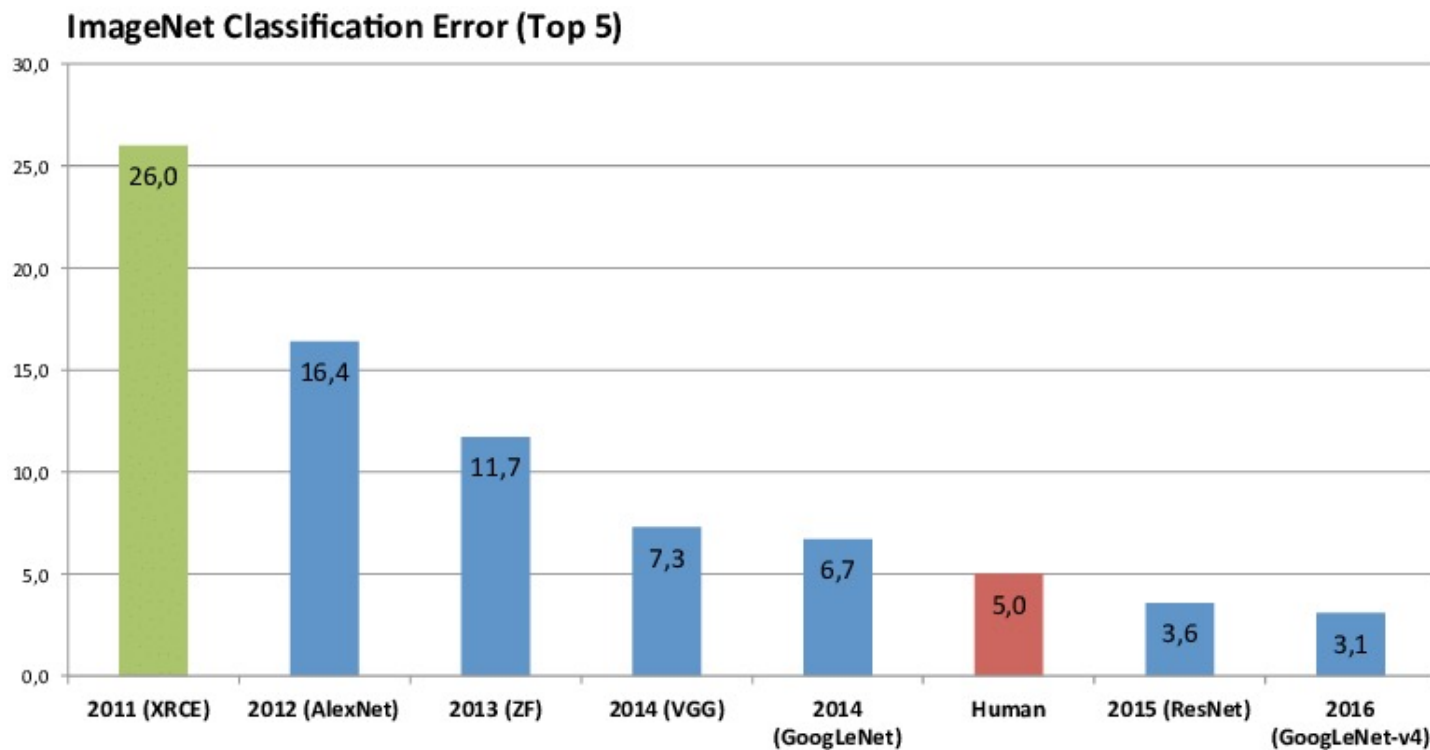
에이아이스쿨(AISchool) 대표
양진호 (솔라리스)

<http://aischool.ai>

<http://solarisailab.com>

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



GoogLeNet(Inception V1)

- Szegedy, Christian, et al, "Going deeper with convolutions.", CVPR 2015.
- <https://arxiv.org/pdf/1409.4842.pdf>

Going deeper with convolutions

Christian Szegedy
Google Inc.

Wei Liu
University of North Carolina, Chapel Hill

Yangqing Jia
Google Inc.

Pierre Sermanet
Google Inc.

Scott Reed
University of Michigan

Dragomir Anguelov
Google Inc.

Dumitru Erhan
Google Inc.

Vincent Vanhoucke
Google Inc.

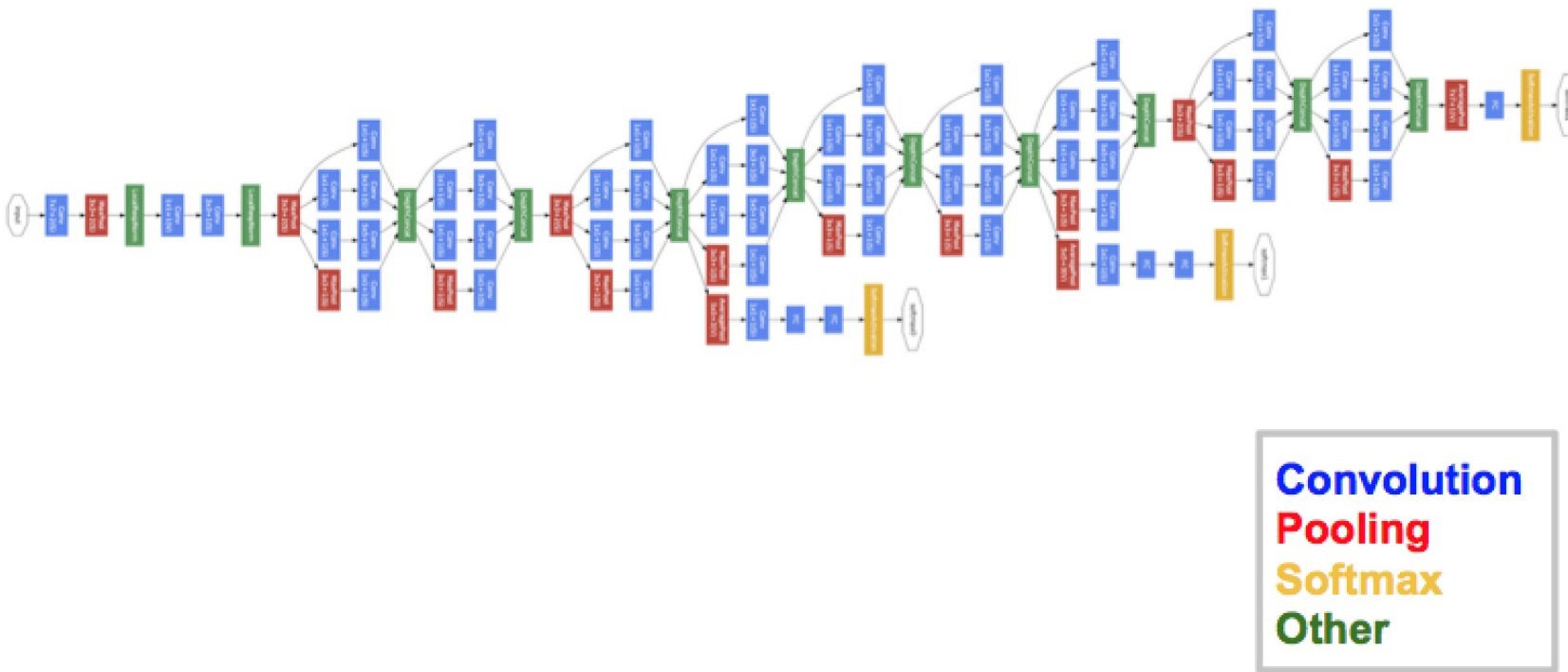
Andrew Rabinovich
Google Inc.

Abstract

We propose a deep convolutional neural network architecture codenamed Inception, which was responsible for setting the new state of the art for classification and detection in the ImageNet Large-Scale Visual Recognition Challenge 2014 (ILSVRC14). The main hallmark of this architecture is the improved utilization of the computing resources inside the network. This was achieved by a carefully crafted design that allows for increasing the depth and width of the network while keeping the computational budget constant. To optimize quality, the architectural decisions were based on the Hebbian principle and the intuition of multi-scale processing. One particular incarnation used in our submission for ILSVRC14 is called GoogLeNet, a 22 layers deep network, the quality of which is assessed in the context of classification and detection.

논문 리뷰 - Going Deeper with Convolutions

- **핵심 아이디어** : LeNet에 기반해서 Google이 제안한 CNN 모델인 GoogLeNet을 제안, AlexNet에 비해 12배나 적은 parameter로 더욱 좋은 성능을 보임. ILSVRC-2014 대회에서 1등을 차지함.
- Inception v1 모델을 제안



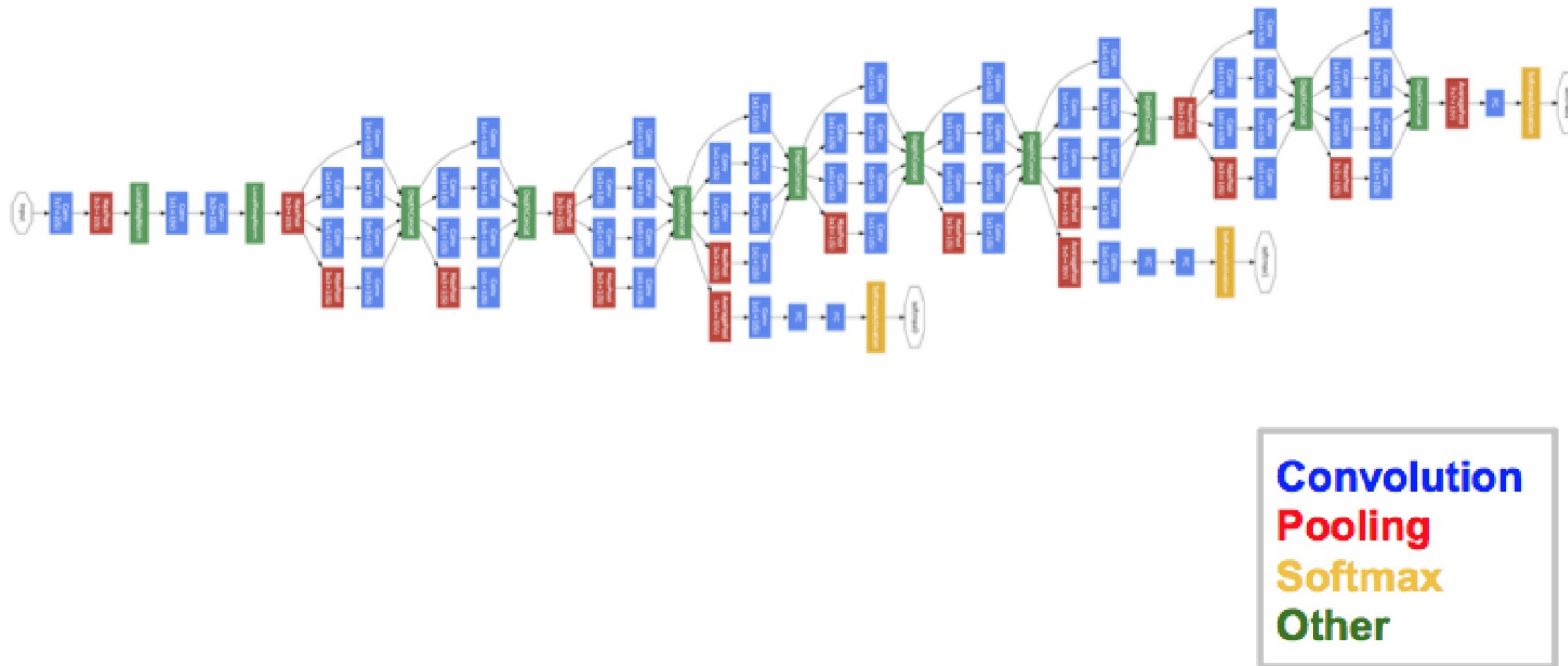
GoogLenet 핵심 아이디어

- 최대한 파라미터를 줄이면서 네트워크를 깊고 넓게 디자인하고자 함. (레이어가 깊더라도 연결이 Sparse하다면 파라미터수가 줄어듬.) 이는 Overfitting을 방지하는 효과를 얻을 수 있다.
- 하지만, 연산은 Dense하게 처리하는 것을 목표로 함
- 딥러닝 계산은 결국 매트릭스 곱으로 처리 되는데 이때 매트릭스의 값이 Sparse하다면(예를 들어, 100x100 매트릭스에 1개의 값만 있고 나머지는 모두 0이라면) 낭비되는 연산이 많아짐. 따라서 매트릭스 연산을 할때 데이터가 Sparse해지는 현상을 방지하는 것을 목표로 한다.
- 이를 위해, 여러 방면의 고민과 실험 끝에 영화 Inception에서 이름을 따온 Inception Module을 제안
- ILSVRC-2014에서 1등을 차지한 GoogLeNet은 Inception 모델의 여러 구현 중에 하나이다.



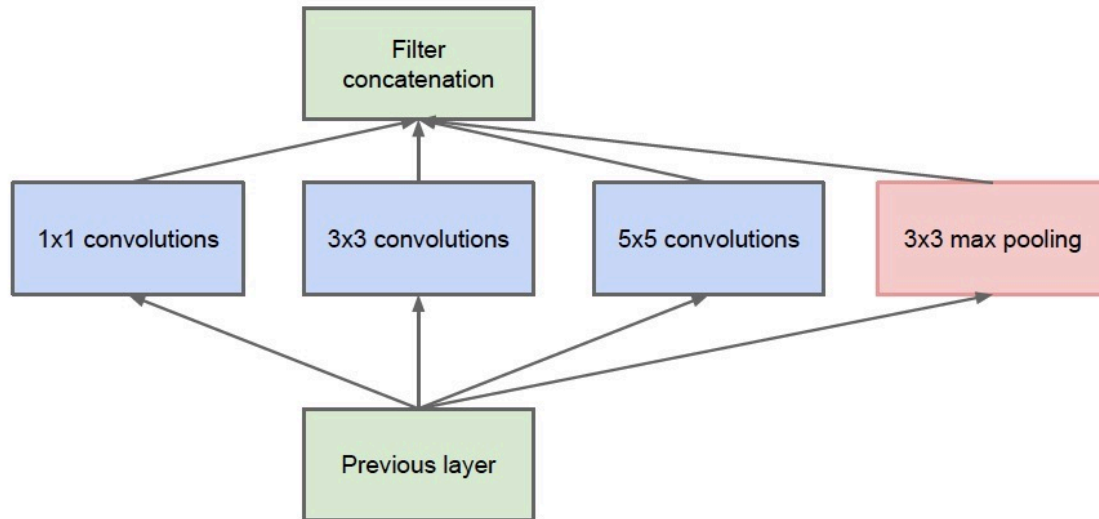
Model Architecture

- GoogLeNet의 전체 모델 훑어보기



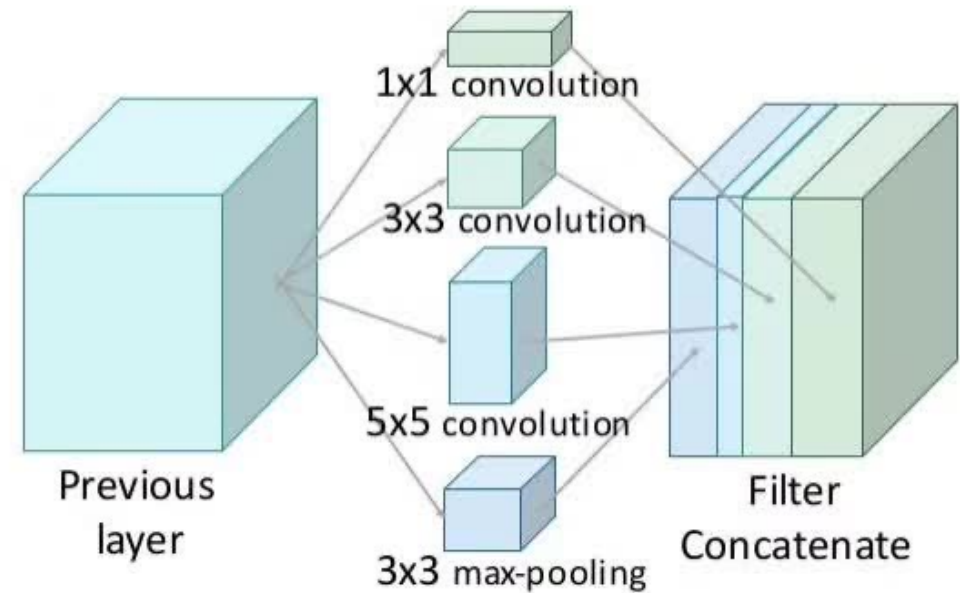
Inception Module - Naïve version

- 4개의 연산을 병렬적으로 수행한뒤 Concatenate로 합침



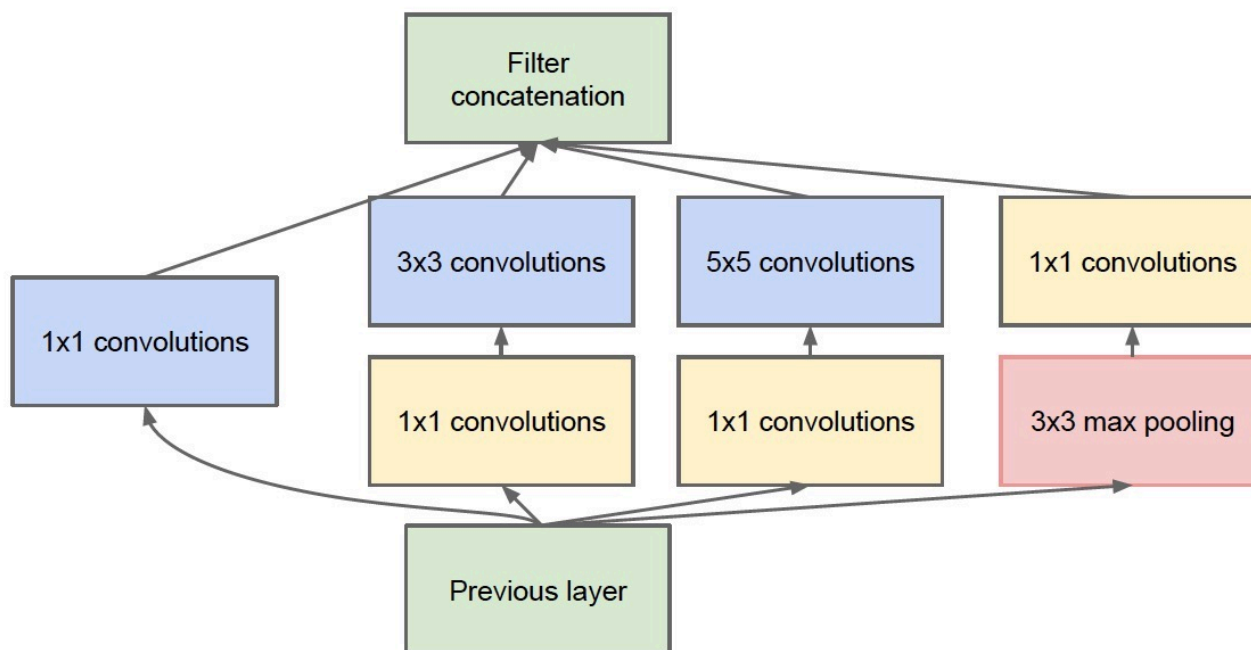
(a) Inception module, naïve version

Inception Module



Inception Module – Dimension Reduction version

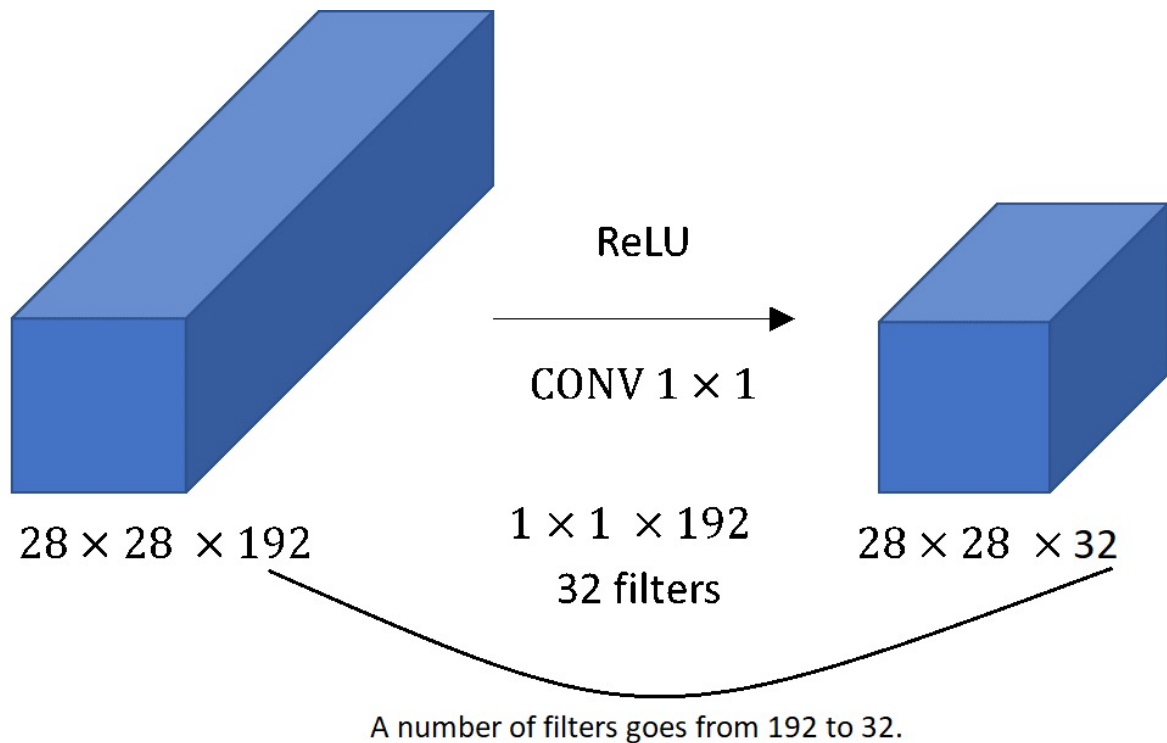
- 1x1 Convolution을 이용해서 Dimension Reduction을 수행해서 연산량을 축소함



(b) Inception module with dimension reductions

1x1 Convolution의 사용

- 1x1 Convolution의 장점
- Dimension Reduction
- More ReLU(More Non-Linearity)



GoogLeNet Model Architecture Summary

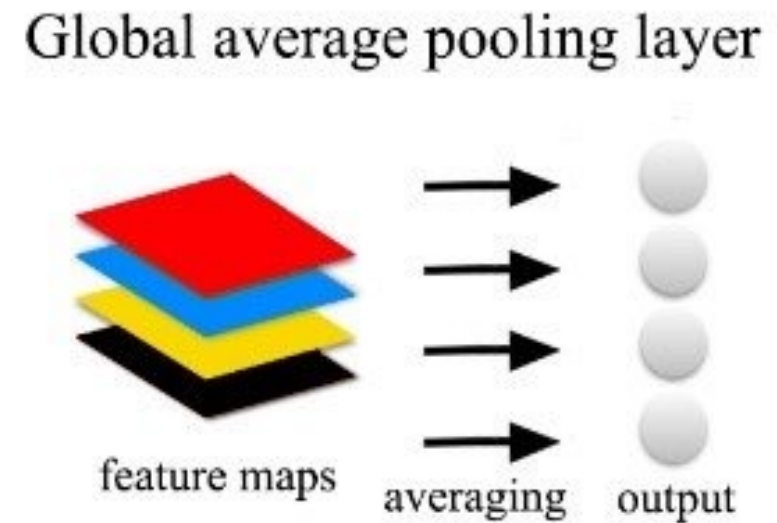
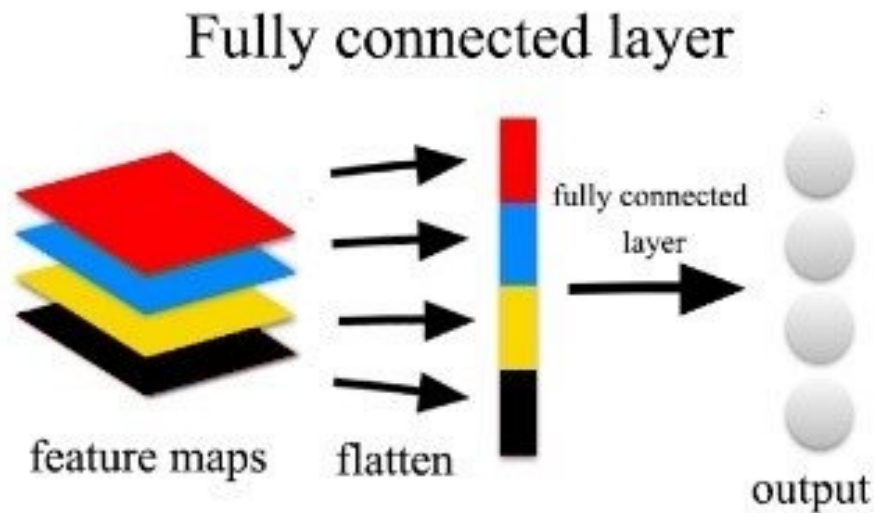
- 입력 이미지 : 224x224x3 크기의 이미지

type	patch size/ stride	output size	depth	#1×1	#3×3 reduce	#3×3	#5×5 reduce	#5×5	pool proj	params	ops
convolution	7×7/2	112×112×64	1							2.7K	34M
max pool	3×3/2	56×56×64	0								
convolution	3×3/1	56×56×192	2		64	192				112K	360M
max pool	3×3/2	28×28×192	0								
inception (3a)		28×28×256	2	64	96	128	16	32	32	159K	128M
inception (3b)		28×28×480	2	128	128	192	32	96	64	380K	304M
max pool	3×3/2	14×14×480	0								
inception (4a)		14×14×512	2	192	96	208	16	48	64	364K	73M
inception (4b)		14×14×512	2	160	112	224	24	64	64	437K	88M
inception (4c)		14×14×512	2	128	128	256	24	64	64	463K	100M
inception (4d)		14×14×528	2	112	144	288	32	64	64	580K	119M
inception (4e)		14×14×832	2	256	160	320	32	128	128	840K	170M
max pool	3×3/2	7×7×832	0								
inception (5a)		7×7×832	2	256	160	320	32	128	128	1072K	54M
inception (5b)		7×7×1024	2	384	192	384	48	128	128	1388K	71M
avg pool	7×7/1	1×1×1024	0								
dropout (40%)		1×1×1024	0								
linear		1×1×1000	1							1000K	1M
softmax		1×1×1000	0								

Average Pooling Instead Fully Connected Layers

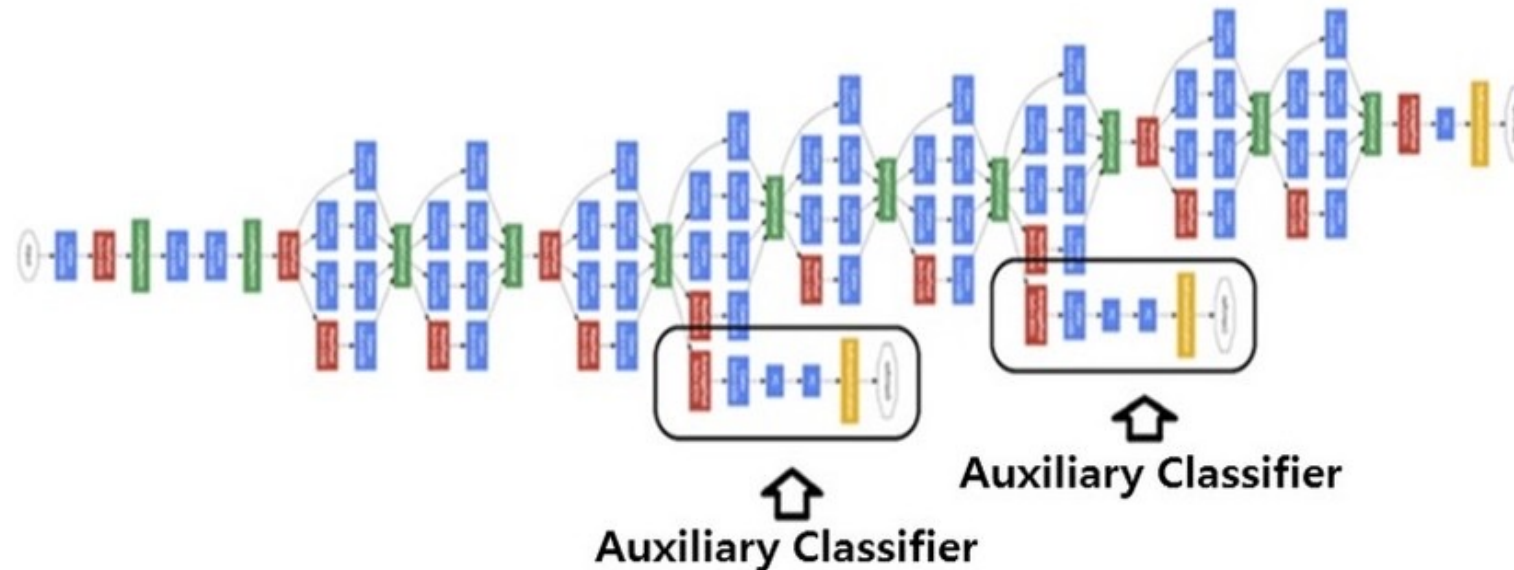
- Global Average Pooling의 장점

1. 학습해야할 파라미터 개수가 줄어들어 Overfitting의 가능성이 낮아짐
2. Spatial 정보의 평균을 취하므로 Input의 translation에 강건하다.



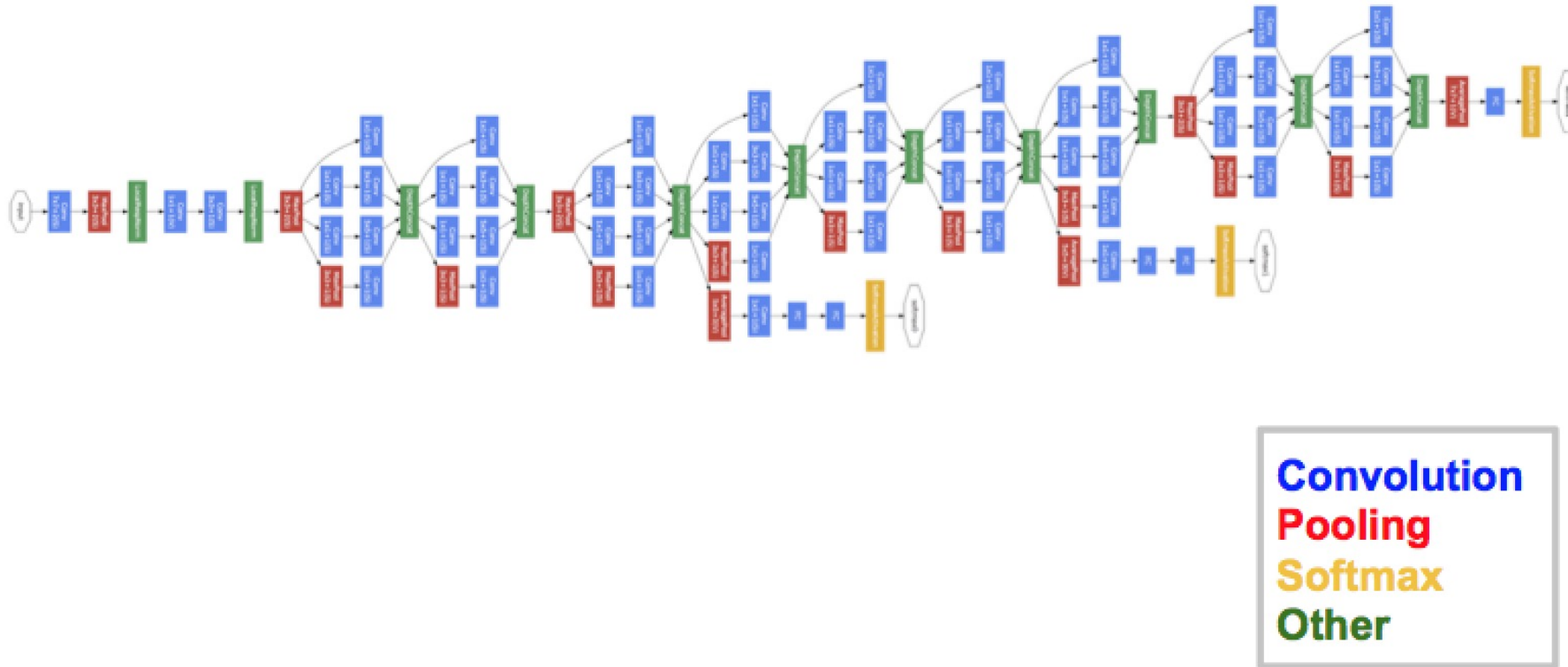
Auxiliary Classifier

- 레이어가 깊어지면서 발생할 수 있는 Vanishing Gradient Problem을 방지하고자 Auxiliary Classifier를 추가
- Test시에는 이를 사용하지 않는다.



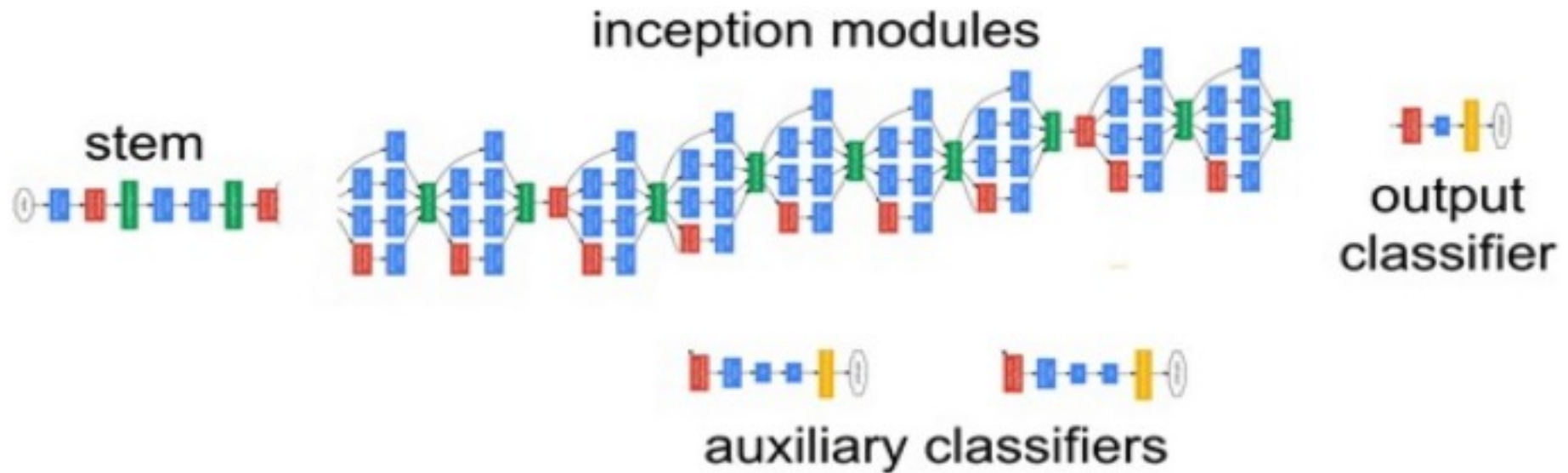
Model Architecture

- 전체 모델을 다시 한번 Remind



Model Architecture

- 전체 모델을 부분별로 살펴보면 아래와 같다.



Experiment Result

- 대회에서는 GoogLeNet 7개를 Ensemble Learning 해서 제출 함
- Specifically, we resize the image to 4 scales where the shorter dimension (height or width) is 256, 288, 320 and 352(x4) respectively, take the left, center and right square(x3) of these resized images (in the case of portrait images, we take the top, center and bottom squares).
- For each square, we then take the 4 corners and the center 224x224 crop as well as the square resized to 224x224(x6) , and their mirrored versions(x2).
- This results in $4 \times 3 \times 6 \times 2 = 144$ crops per image.

Team	Year	Place	Error (top-5)	Uses external data
SuperVision	2012	1st	16.4%	no
SuperVision	2012	1st	15.3%	Imagenet 22k
Clarifai	2013	1st	11.7%	no
Clarifai	2013	1st	11.2%	Imagenet 22k
MSRA	2014	3rd	7.35%	no
VGG	2014	2nd	7.32%	no
GoogLeNet	2014	1st	6.67%	no

Table 2: Classification performance

Number of models	Number of Crops	Cost	Top-5 error	compared to base
1	1	1	10.07%	base
1	10	10	9.15%	-0.92%
1	144	144	7.89%	-2.18%
7	1	7	8.09%	-1.98%
7	10	70	7.62%	-2.45%
7	144	1008	6.67%	-3.45%

GoogLeNet(Inception V1)의 의의

- 적은 파라미터수로도 높은 성능을 발휘하는 Inception Module이라는 창의적인 딥러닝 구조를 제안함
- 마지막 Feature Map에서 활용할 수 있는 Global Average Pooling 기법을 제안함

Thank you!
