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.

배성은

0. Abstract

A deep convolutional neural network architecture – Inception을 제안

주요 특징

- 네트워크 내부 컴퓨팅 리소스의 활용도 향상
- 22 layers deep network

1. Introduction

지난 3년간, 딥러닝 발전덕에 CNN 분야에서 많은 발전이 있었다.

- 이 발전은 주로 새로운 아이디어, 알고리즘 및 향상된 네트워크 아키텍쳐 덕분이다.
- 모바일 및 임베디드 컴퓨팅의 지속적인 관심들로 인해 알고리즘의 효율성. 특히 성능 및 메모리가 중요해졌다. 그래서 논문에서는 유연한 구조를 가지게끔 하였다.
- 추론시간에 1.5 billion이하의 연산만 수행하도록 설계해 현실에도 사용할 수 있게끔 설계하였다.
- GoogLeNet의 코드네임인 Inception은 We need to go deeper"에서 착안
  - Inception module 형태의 새로운 차원의 구조 도입
  - 네트워크의 깊이가 증가하였다는 의미

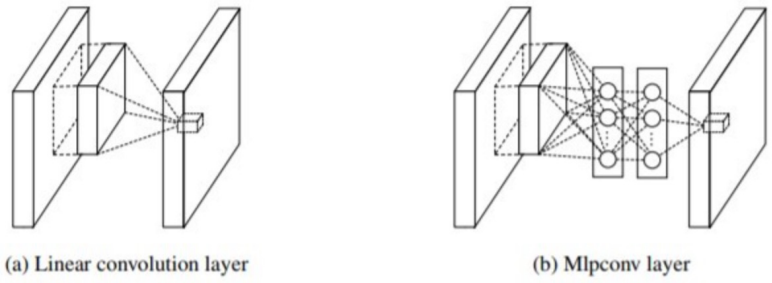
2. Related Work

1.
LeNet-5이후로 CNN 표준 구조를 가지고 있음 – Convolutional layer – 1개 or 그 이상 FC layer

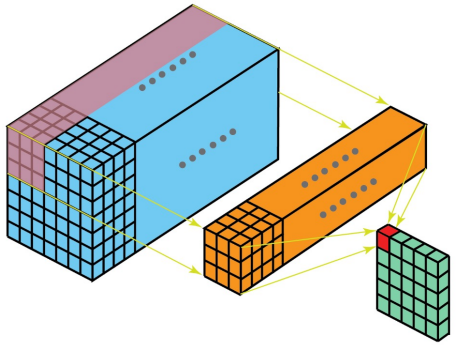
요즘 트렌드는 layer의 수와 사이즈를 늘리고, 오버피팅을 해결하기 위해 dropout을 적용
→ GoogLeNet도 같은 구조

2. Network in Network 논문
1x1 Convolutional layer가 추가, ReLu activation이 온다.

1x1 Convolutional layer의 목적
1) 병목현상을 제거하기 위한 차원축소
2) 네트워크 크기 제한


(a) Linear convolution layer

(b) Mlpconv layer



CCCP (Cascaded Cross Channel Pooling)
→ 1x1 convolutional layer과 연산방식 및 효과가 유사

3. Motivation and High Level Considerations

Deep neural network 성능을 향상시키는 직접적인 방법 → 네트워크 깊이 layer, 폭을 늘리는 것 node

이 당시에는 두가지 문제점

1) 크기가 커질수록 더욱 많은 파라미터를 가지게 된다. 과적합되기 쉽다.
고품질 훈련 데이터 셋 생성이 어려움.

2) 네트워크가 커지면 컴퓨팅 리소스가 급격히 증가함.
무차별적인 크기 증가보다 리소스의 효율적인 분배가 선호됨.

해결방법

- Dense한 FC구조 → Sparsely Connected 구조
- But, 컴퓨팅 환경은 균일하지 않은 sparse data구조를 다룰 때 비효율적임.
- 여러 시도들이 있었음.

- Inception 구조는 유사 sparse 구조를 시험하기 위해서 시작되었음.

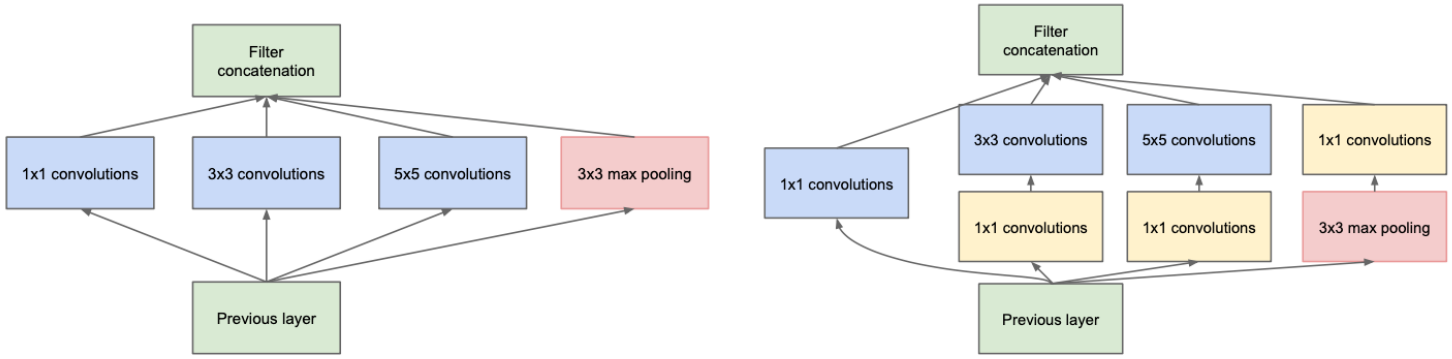
4. Architectural Details

Inception : CNN에서 최적의 local sparse structure로 근사화하고, 이를 dense components로 바꾸는 방법을 찾는 것.
즉, 최적의 local 구성을 찾고 공간적으로 반복 → Sparse matrix를 묶어 상대적으로 Dense한 Submatrix만듬

특징을 효과적으로 추출하기 위해 filter size를 1x1, 3x3, 5x5로 병렬 수행함. (a)
그러나 Layer가 깊어짐에 따라 연산량이 늘어나 비용이 많이 듬.

그래서 (b) 1x1을 추가해 차원을 축소한다. 비선형적 특징도 추가

- 과도한 연산량 문제없이 각 단계에서 유닛수를 증가가능
- Visual 정보가 다양한 scale로 처리되고, 다음 layer는 동시에 서로 다른 layer에서 특징을 추출할 수 있다.



(a) Inception module, naïve version (b) Inception module with dimension reductions

Figure 2: Inception module

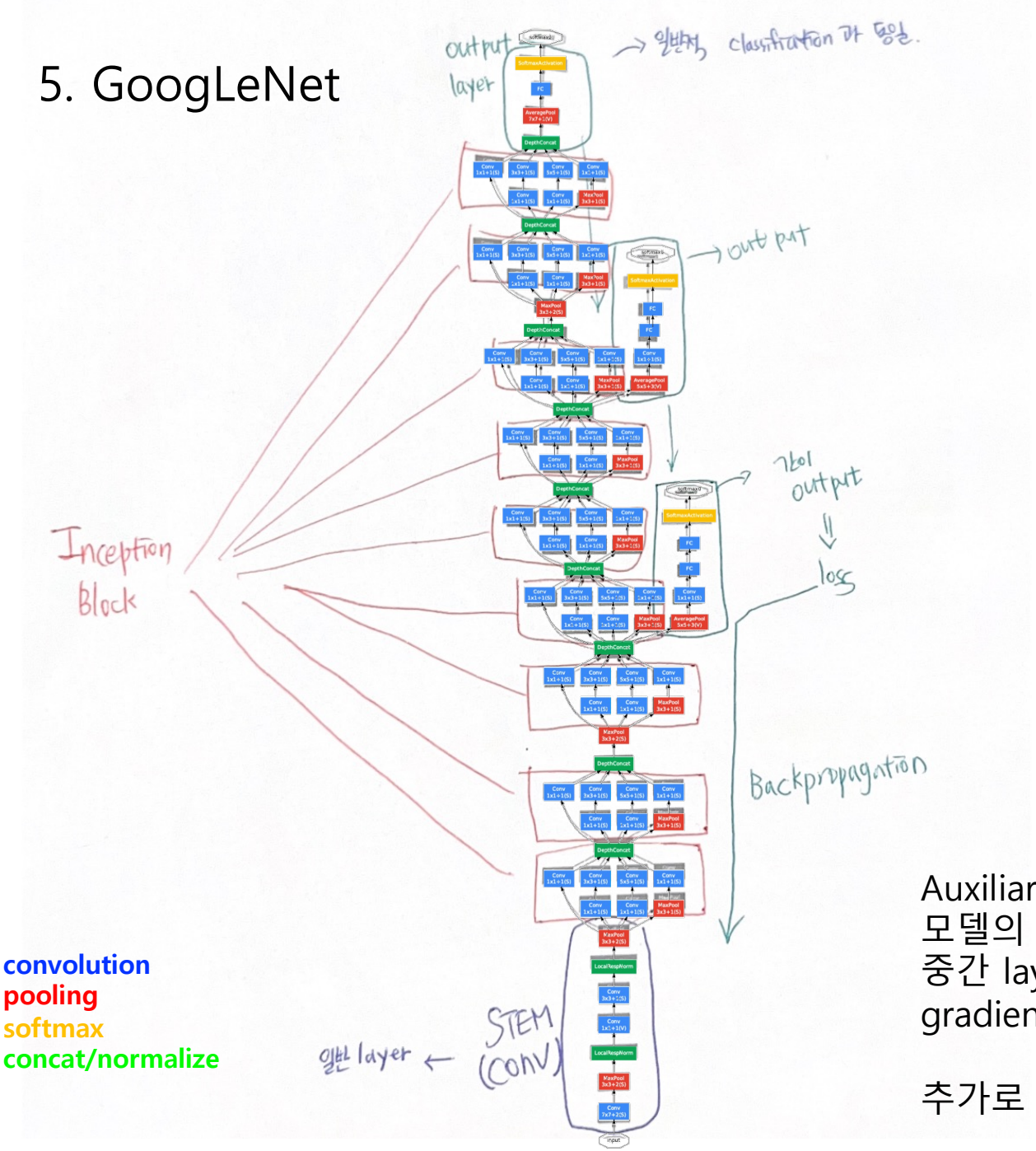
5. GoogLeNet

Inception 여러 개 쌓아 구성한 네트워크
당시 22개로 깊은 네트워크를 구성했음에도 AlexNet대비 12배 적은 파라미터 사용

	type	patch size/ stride	output size	depth	#1×1	#3×3 reduce	#3×3	#5×5 reduce	#5×5	pool proj	params	ops
stem(conv)	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	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	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	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
output	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								

Table 1: GoogLeNet incarnation of the Inception architecture

5. GoogLeNet



convolution
pooling
softmax
concat/normalize

Auxiliary classifier 적용부분
모델의 깊이가 매우 깊으면, gradient vanishing문제가 발생할 수 있어 중간 layer에 추가해서 결과를 출력하고 추가적인 역전파를 일으켜 gradient가 전달될 수 있게끔 해 정규화 효과가 나타남.

추가로 지나치게 영향을 주는 것을 막기 위해 0.3 가중치를 주었다.

6. Training Methodology

0.9 momentum stochastic gradient descent
Learning rate 8 epochs마다 4%씩 감소

이미지의 가로 세로 비율을 3:4와 4:3으로 유지하면서 본래 사이즈의 8~100%가 포함되도록 다양한 크기의 patch를 사용하였다.
Photometric distortions을 통해 학습데이터를 늘렸다.

* 광도(Photometric) 왜곡은 밝기, 색조, 대비, 채도 및 노이즈를 조정하여 동일한 이미지의 더 많은 다양성을 표시함으로써 새로운 이미지를 생성합니다.

Adjust hue of an image

귀의 예에서 Hue(또는 색상 모양 매개변수)를 조정하여 이미지를 수정하고 새로운 샘플을 생성하여 훈련 네트에서 더 많은 가변성을 생성했습니다.

https://wikidocs.net/163078

7. ILSVRC 2014 Classification Challenge Setup and Results

두가지 숫자 보고
- 상위 1개 정확도
- 상위 5개 오류율

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%

Table 3: GoogLeNet classification performance break down

8. ILSVRC 2014 Detection Challenge Setup and Results

Team	Year	Place	mAP	external data	ensemble	approach
UvA-Euvision	2013	1st	22.6%	none	?	Fisher vectors
Deep Insight	2014	3rd	40.5%	ImageNet 1k	3	CNN
CUHK DeepID-Net	2014	2nd	40.7%	ImageNet 1k	?	CNN
GoogLeNet	2014	1st	43.9%	ImageNet 1k	6	CNN

Table 4: Detection performance

Team	mAP	Contextual model	Bounding box regression
Trimps-Soushen	31.6%	no	?
Berkeley Vision	34.5%	no	yes
UvA-Euvision	35.4%	?	?
CUHK DeepID-Net2	37.7%	no	?
GoogLeNet	38.02%	no	no
Deep Insight	40.2%	yes	yes

Table 5: Single model performance for detection

9. Conclusions

Inception구조는 Sparse 구조를 Dense 구조로 근사화하여 성능을 개선하였다.

연산량이 약간 증가하더라도 상당한 품질향상이 있다.
그리고 Context를 활용하지 않고 bounding box regression를 수행하지 않았음에도 경쟁력이 있다.