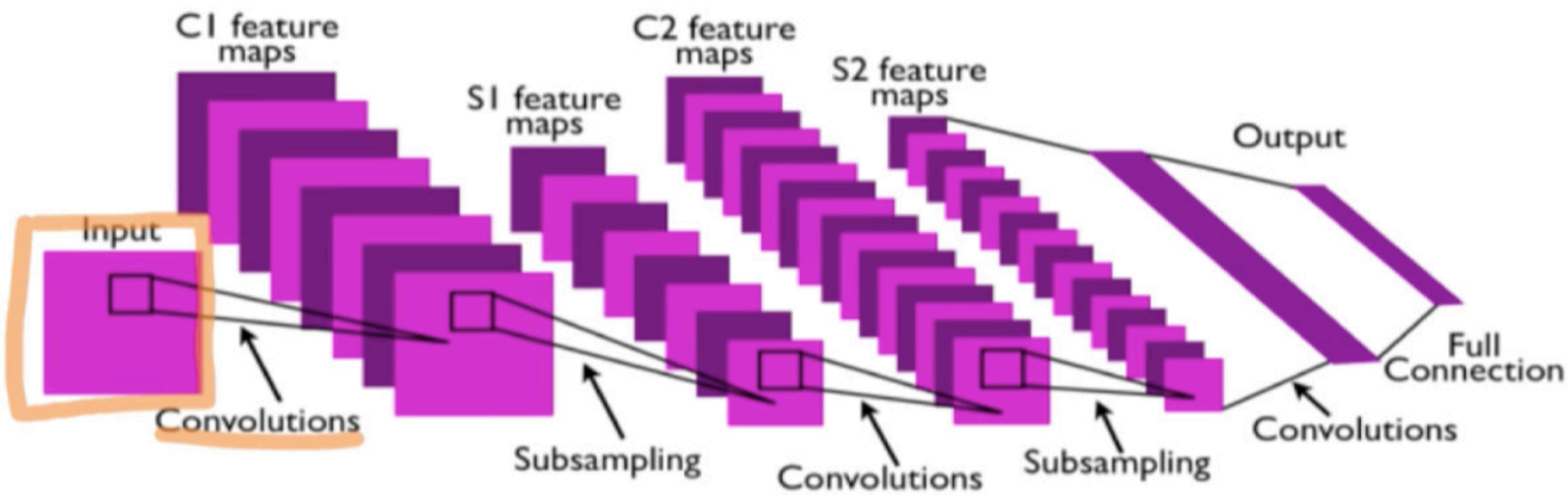
본격 CNN 논문 파헤치기

이전에...

CNN

CNN

Convolutional Neural Network



이번 발표에서 주의깊게 살펴봐야 할 부분은 다음과 같다.

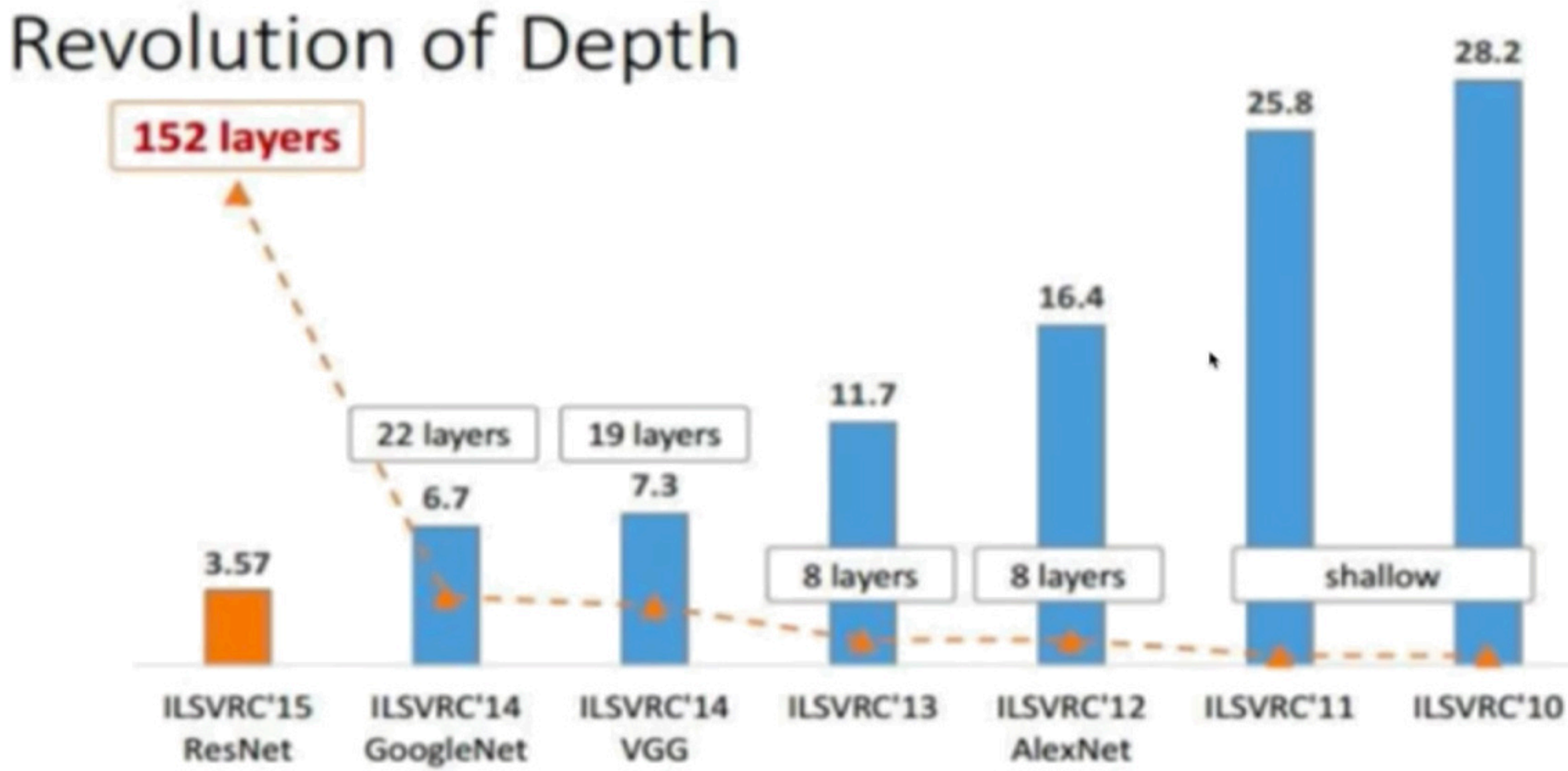
- Terminology & Notation
- Structure model
- implementation
- Pros and Cons

오늘 보게 될 논문들은 크게 4가지.

- AlexNet
- VGG
- GoogLeNet
- ~~ResNet~~

ILSVRC
(IMAGENET LARGE SCALE VISUAL RECOGNITION COMPETITION)

Top-5 Classification Error



ALEXNET

WHY ALEXNET

- 기존의 컴퓨터 비전 기법들을 사용한 알고리즘들의 성능을 월등히 능가하는 성능을 냈다.
- 딥러닝 혁명을 이끈 초기 딥러닝 모델 중 하나로 컨볼루션 신경망 구조를 가졌다.
- 1등이 15.3% 2등이 26.2%(2등 컴퓨터 비전 기법)

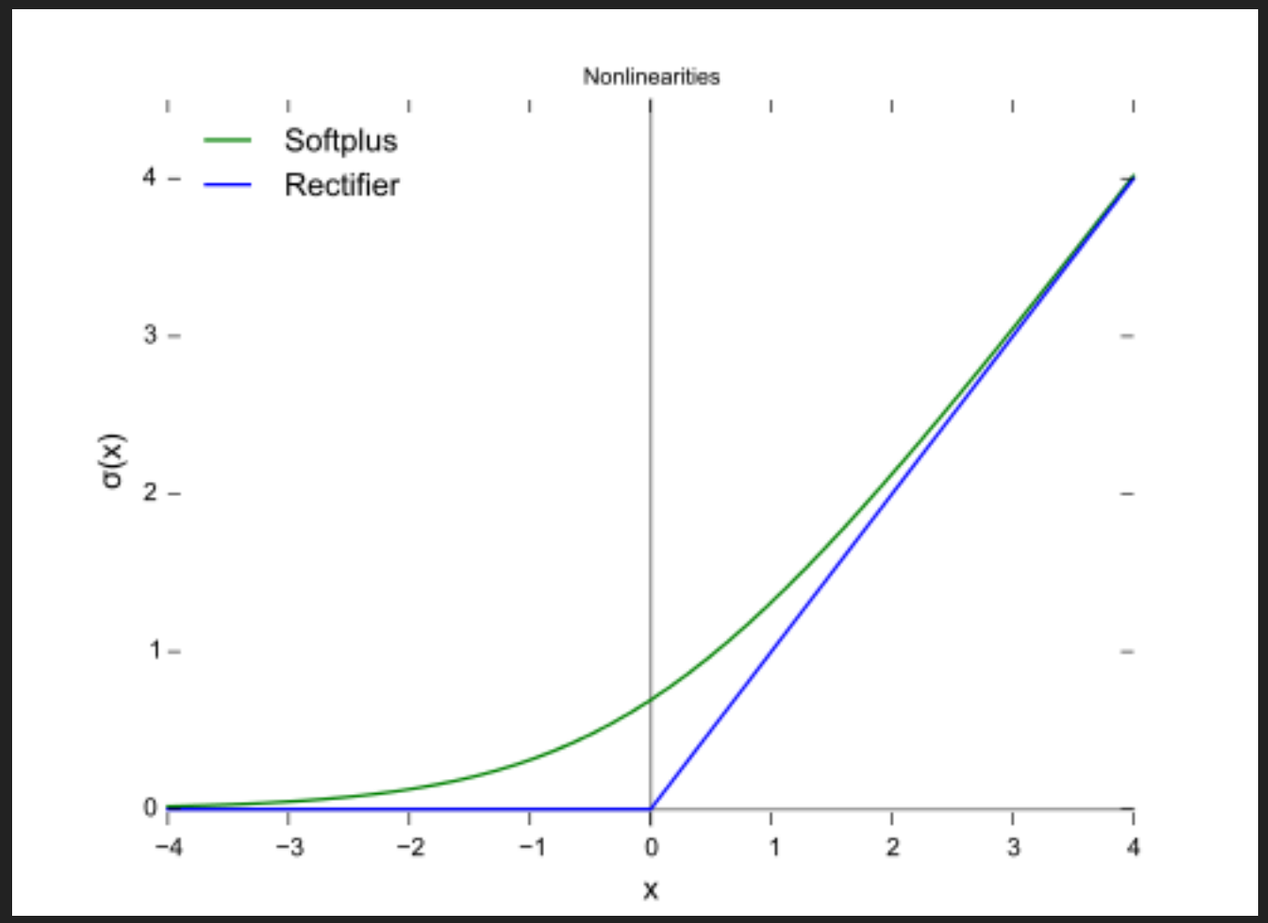
ALEXNET

- ReLu(Rectified Linear Unit)
- LRN(Local Response Normalization)
- Regularization - overfitting 문제를 피하기 위한 해결 방법
  - Data augmentation
  - Dropout

RELU(RECTIFIED LINEAR UNIT)

$$f = \begin{cases} (x < 0) & f(x) = 0 \\ (x \geq 0) & f(x) = x \end{cases}$$

- 정류한 선형 유닛?
- Solving Vanishing gradient problem



- SPARSE ACTIVATION: 0이하의 입력에 대해 0을 출력함으로 부분적으로 활성화 시킬 수 있다.
- EFFICIENT GRADIENT PROPAGATION: GRADIENT의 VANISHING이 없으며 GRADIENT가 EXPLODING되지 않는다.
- EFFICIENT COMPUTATION: 선형함수이므로 미분 계산이 매우 간단하다.

LRN(LOCAL RESPONSE NORMALIZATION)

딥러닝 입력 데이터 정규화 (Normalizing inputs)

학습을 수행하기 전에 값의 범위를 `normalization` 하는 것은 중요하다. 그 이유는 아래와 같다.

> *입력 변수가 MLP에서와 같이 선형 적으로 결합된다면 적어도 이론 상으로는 입력을 표준화하는 것이 거의 필요하지 않습니다.*
> *그 이유는 해당* `weight` *와* `bais` *를 변경하여 입력 벡터를 재조정하면 이전과 완전히 똑같은 결과를 남길 수 있기 때문입니다.*
> *그러나 입력을* `Standardization` *하면 학습을 더 빨리하고 지역 최적의 상태에 빠지게 될 가능성을 줄이는 다양한 실용적인 이유가 있습니다.*
> *또한, 표준화 된 입력을 통해* `Gradient Descent` *및* `Bayesian estimation` *을 보다 편리하게 수행 할 수 있습니다.*

Normalization

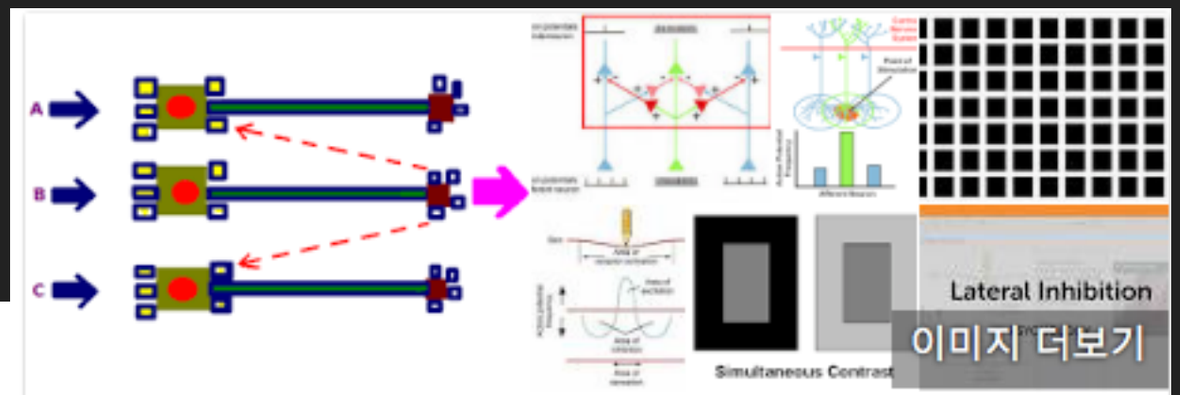
수식 : (요소값 - 최소값) / (최대값 - 최소값)

설명 : 전체 구간을 `0~1` 사이의 값으로 맞춰 준다.

LRN(LOCAL RESPONSE NORMALIZATION)

- Generalization 에러를 줄이기 위해 정규화
- It implements a form of **lateral inhibition** inspired by real neurons

$$b^i_{x,y} = a^i_{x,y} / \left(k + \alpha \right.$$



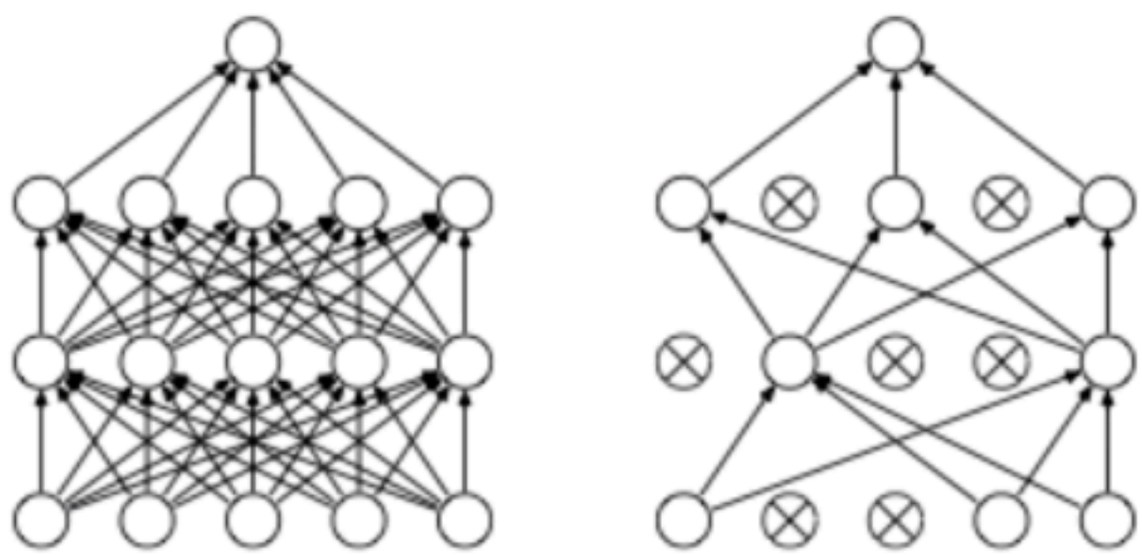
측면 억제

측면 억제는 신경생리학 용어로, 한 영역에 있는 신경 세포가 상호 간 연결되어 있을 때 한 그 자신의 축색이나 자신과 이웃 신경세포를 매개하는 중간신경세포를 통해 이웃에 있는 신경 세포를 억제하려는 경향이다. 위키백과

REGULARIZATION

- Dropout
- Data augmentation

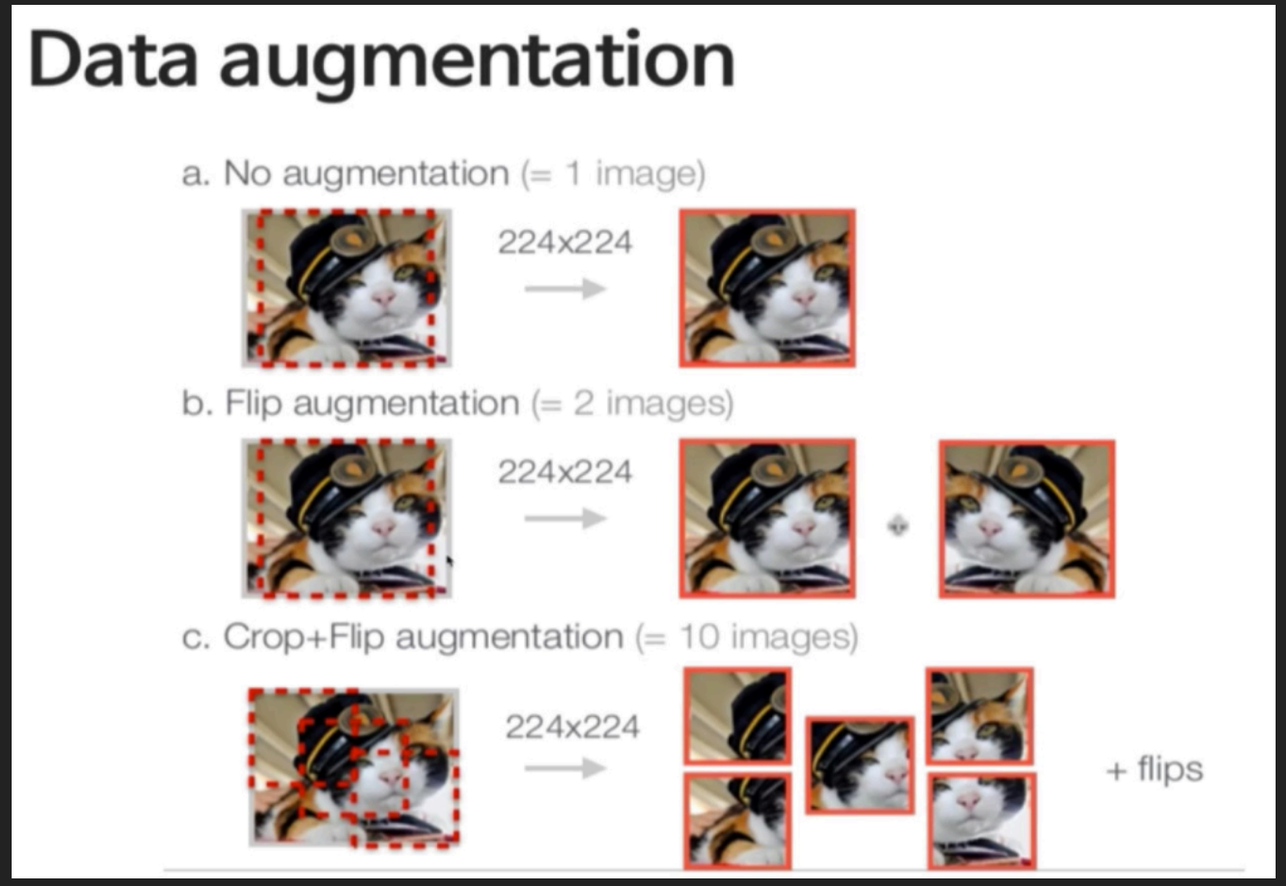
Dropout: A Simple Way to Prevent Neural Networks from Overfitting [Srivastava et al. 2014]

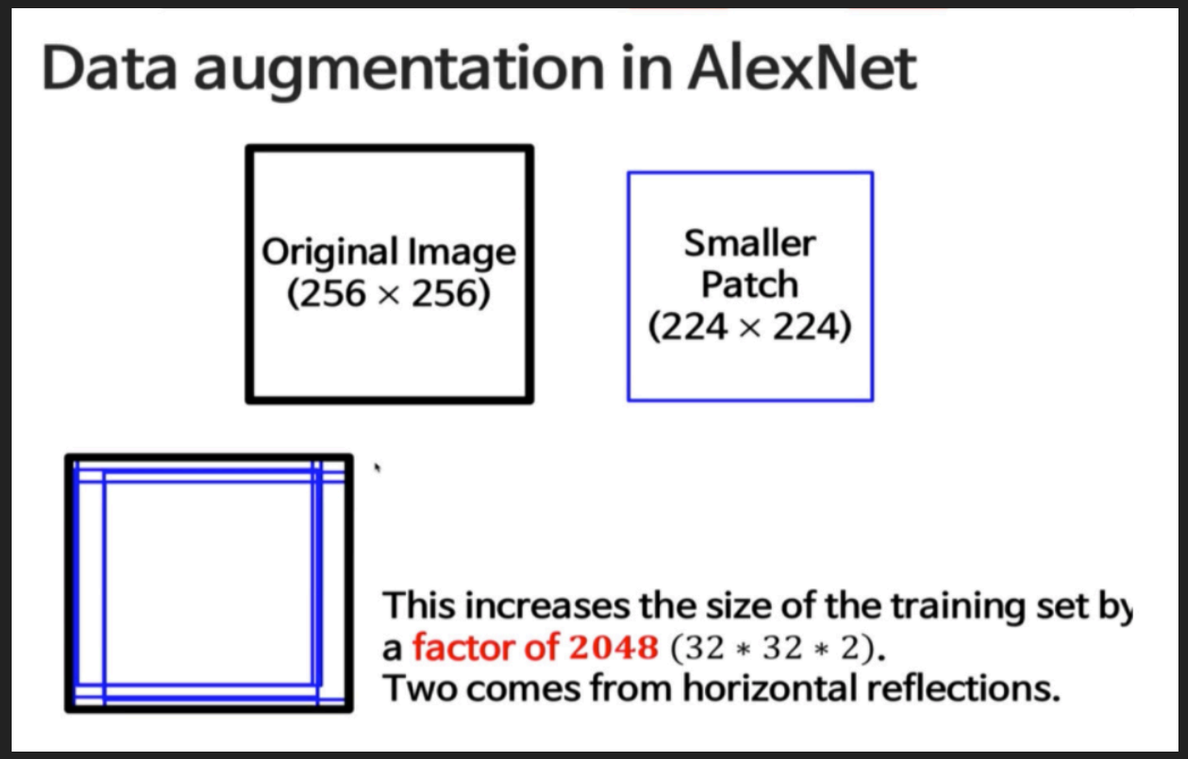


(a) Standard Neural Net (b) After applying dropout.

REGULARIZATION

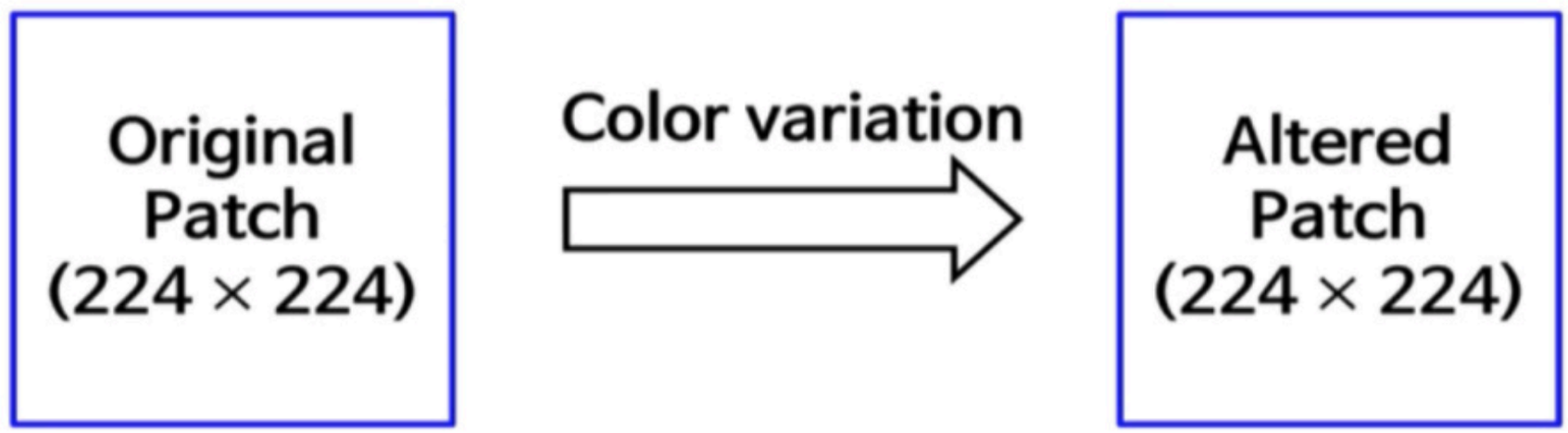
- Dropout
- Data augmentation





REGULARIZATION

Data augmentation in AlexNet



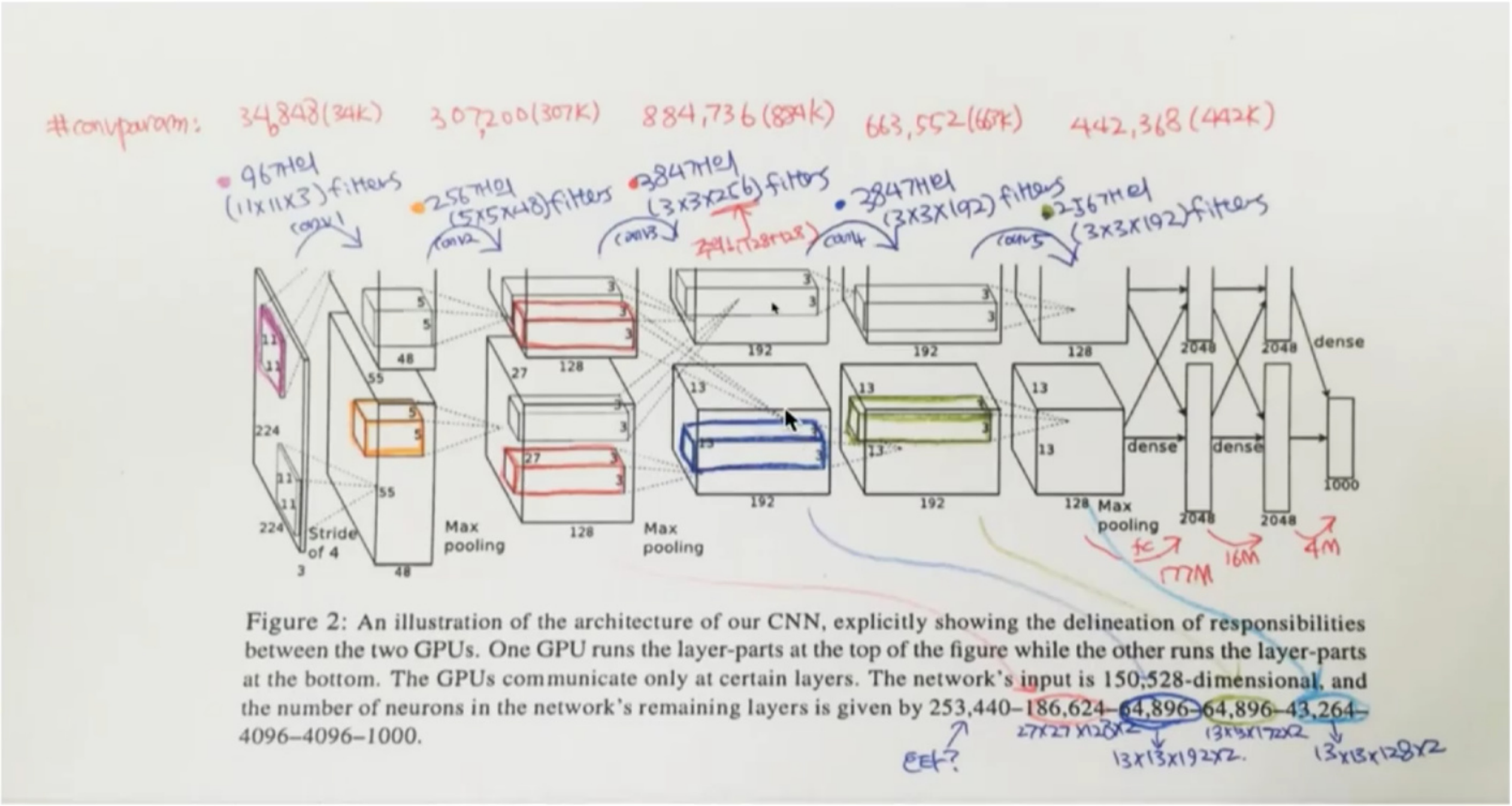
To each RGB image pixel, following quantity is added:

$$[p_1, p_2, p_3][\alpha_1\lambda_1, \alpha_2\lambda_2, \alpha_3\lambda_3]^T$$

where p_i and λ_i are ith eigenvector and eigenvalue of 3×3 covariance matrix of RGB pixel values.

Probabilistically, not a single patch will be same at the training phase! (a factor of infinity!)

AlexNet


#conv param: 34,848(34K) 307,200(307K) 884,736(884K) 663,552(663K) 442,368(442K)
96개의 (11x11x3) filters
256개의 (5x5x48) filters
384개의 (3x3x256) filters
384개의 (3x3x192) filters
256개의 (3x3x192) filters
conv1
conv2
conv3
conv4
conv5
224
224
3
Stride of 4
55
55
48
48
Max pooling
27
27
128
128
Max pooling
13
13
192
192
192
192
128
128
Max pooling
dense
dense
2048
2048
2048
2048
1000
fc
77M
16M
4M
Figure 2: An illustration of the architecture of our CNN, explicitly showing the delineation of responsibilities between the two GPUs. One GPU runs the layer-parts at the top of the figure while the other runs the layer-parts at the bottom. The GPUs communicate only at certain layers. The network's input is 150,528-dimensional, and the number of neurons in the network's remaining layers is given by 253,440–186,624–64,896–64,896–43,264–4096–4096–1000.
253,440
186,624
64,896
64,896
43,264
27x27x128x2
13x13x192x2
13x13x192x2
13x13x128x2

VGG

LOOKING THROUGH VGG PAPER

VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION

Karen Simonyan* **& Andrew Zisserman+**
Visual Geometry Group, Department of Engineering Science, University of Oxford
{karen,az}@robots.ox.ac.uk

ABSTRACT

In this work we investigate the effect of the convolutional network depth on its accuracy in the large-scale image recognition setting. Our main contribution is a thorough evaluation of networks of increasing depth using an architecture with very small (3×3) convolution filters, which shows that a significant improvement on the prior-art configurations can be achieved by pushing the depth to 16–19 weight layers. These findings were the basis of our ImageNet Challenge 2014 submission, where our team secured the first and the second places in the localisation and classification tracks respectively. We also show that our representations generalise well to other datasets, where they achieve state-of-the-art results. We have made our two best-performing ConvNet models publicly available to facilitate further research on the use of deep visual representations in computer vision.

VGG?

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					

VGG

- 논문에 모든 conv 계층에 **3x3** conv 을 활용한 이유를 설명하는데, 그 이유는 파라미터의 개수가 적기 현저하게 적기 때문에.
  예를 들어 3x3 conv 레이어 3개와 7x7 conv 레이어 1개가 있다고 가정.
  3 * (3^2C^2) = 27C^2 // 7^2C^2 = 49C^2
  이전에 CNN에서 말했듯이 parameter의 개수가 적다면 regularization 측면에서 이점을 발휘한다.
- LRN은 성능향상에 도움이 되지 않는다고 함.
- 대부분은 Validation data는 AlexNet과 유사
- Data augmentation에서도 유사하나 한가지 추가 된 것은 image-rescaling을 활용했다.(화면비율을 동일한 비가 되도록 이미지를 줄이거나 늘리는 행위)

GOOGLENET

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.

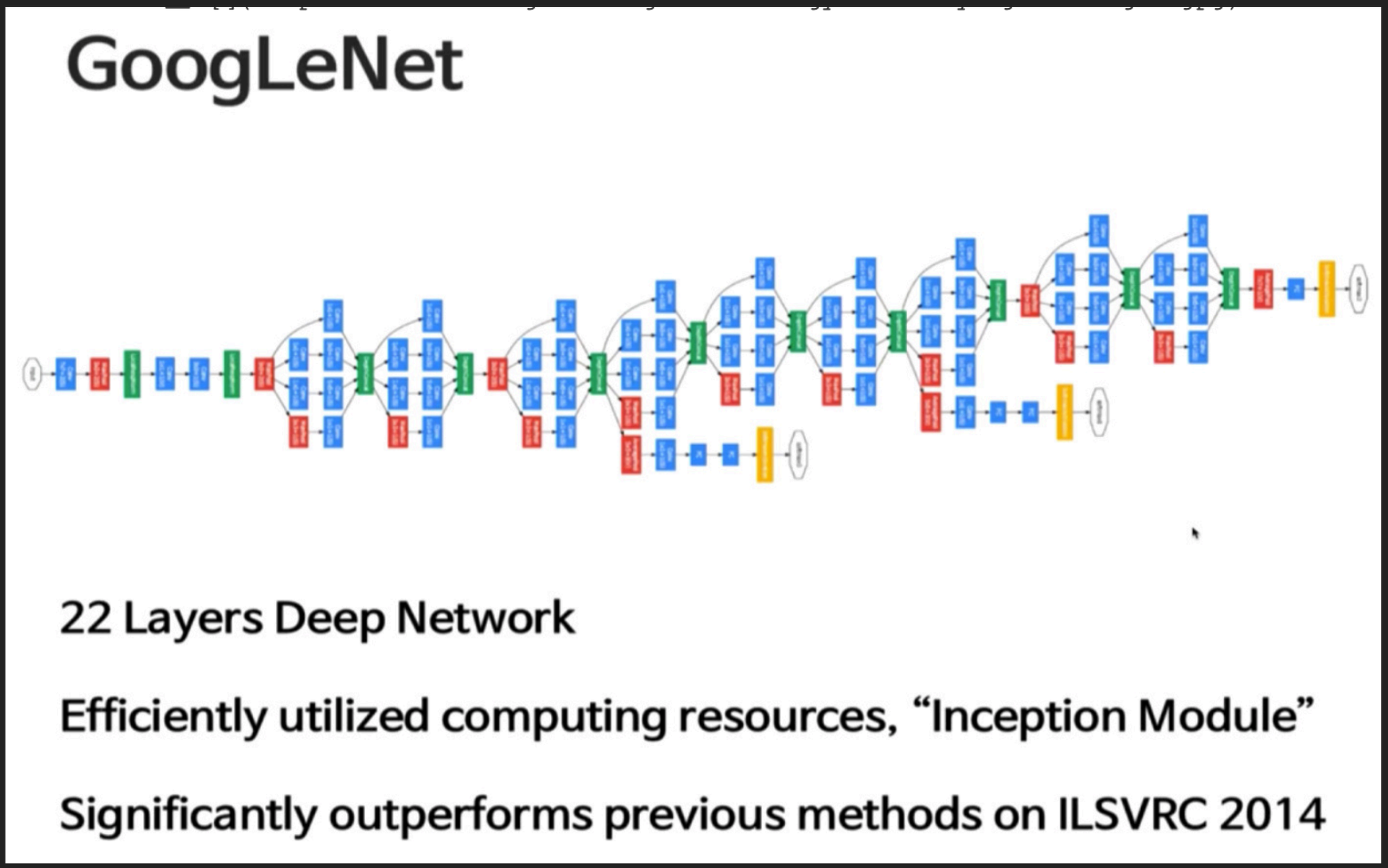
Vincent Vanhoucke
Google Inc.

Andrew Rabinovich
Google Inc.

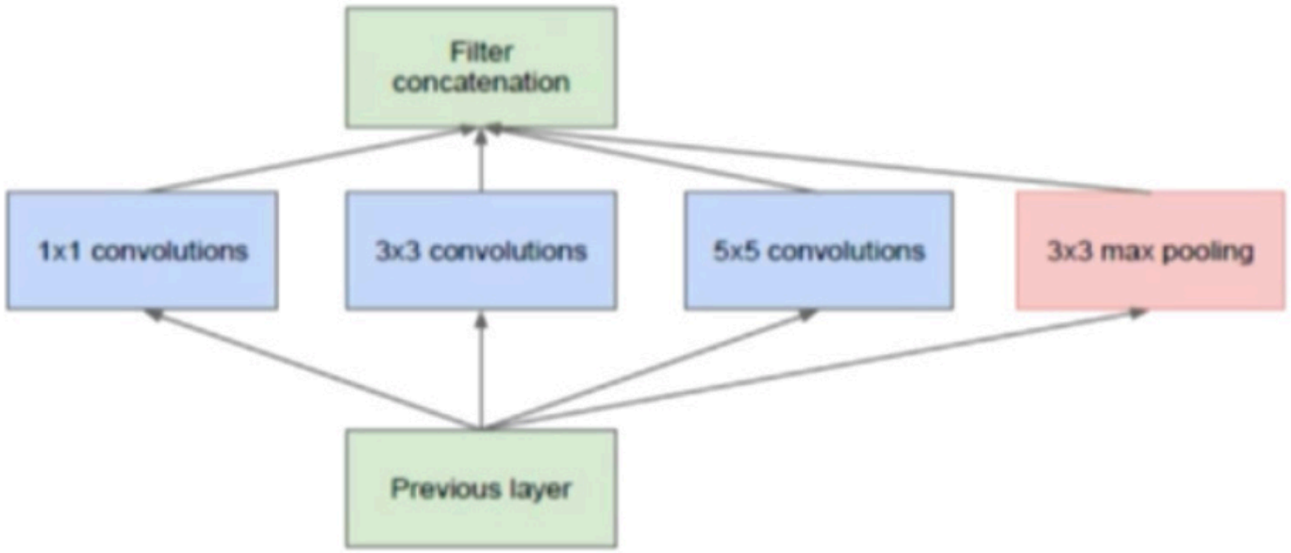
Abstract

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.

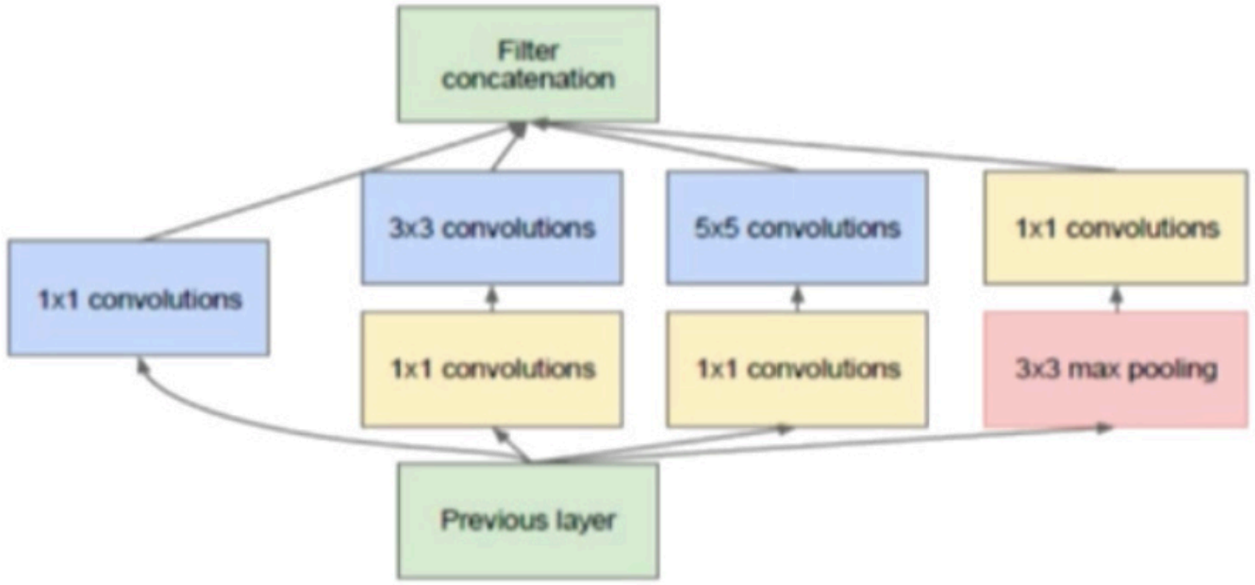
GOOGLENET


GoogLeNet
22 Layers Deep Network
Efficiently utilized computing resources, “Inception Module”
Significantly outperforms previous methods on ILSVRC 2014

Inception module

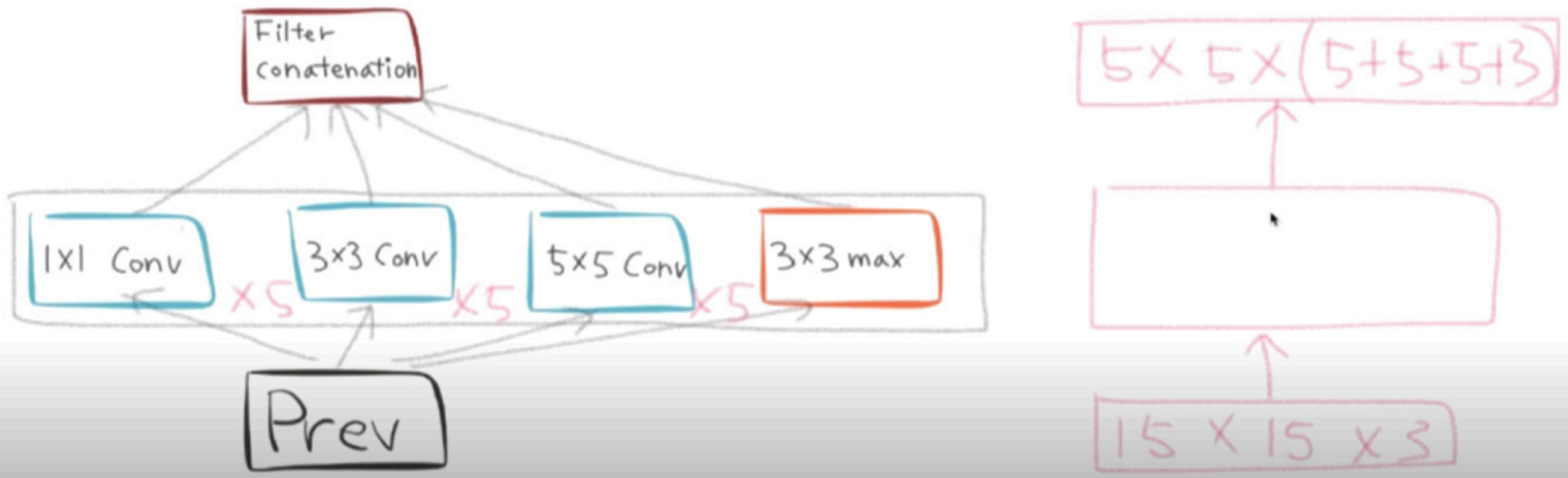


(a) Inception module, naïve version

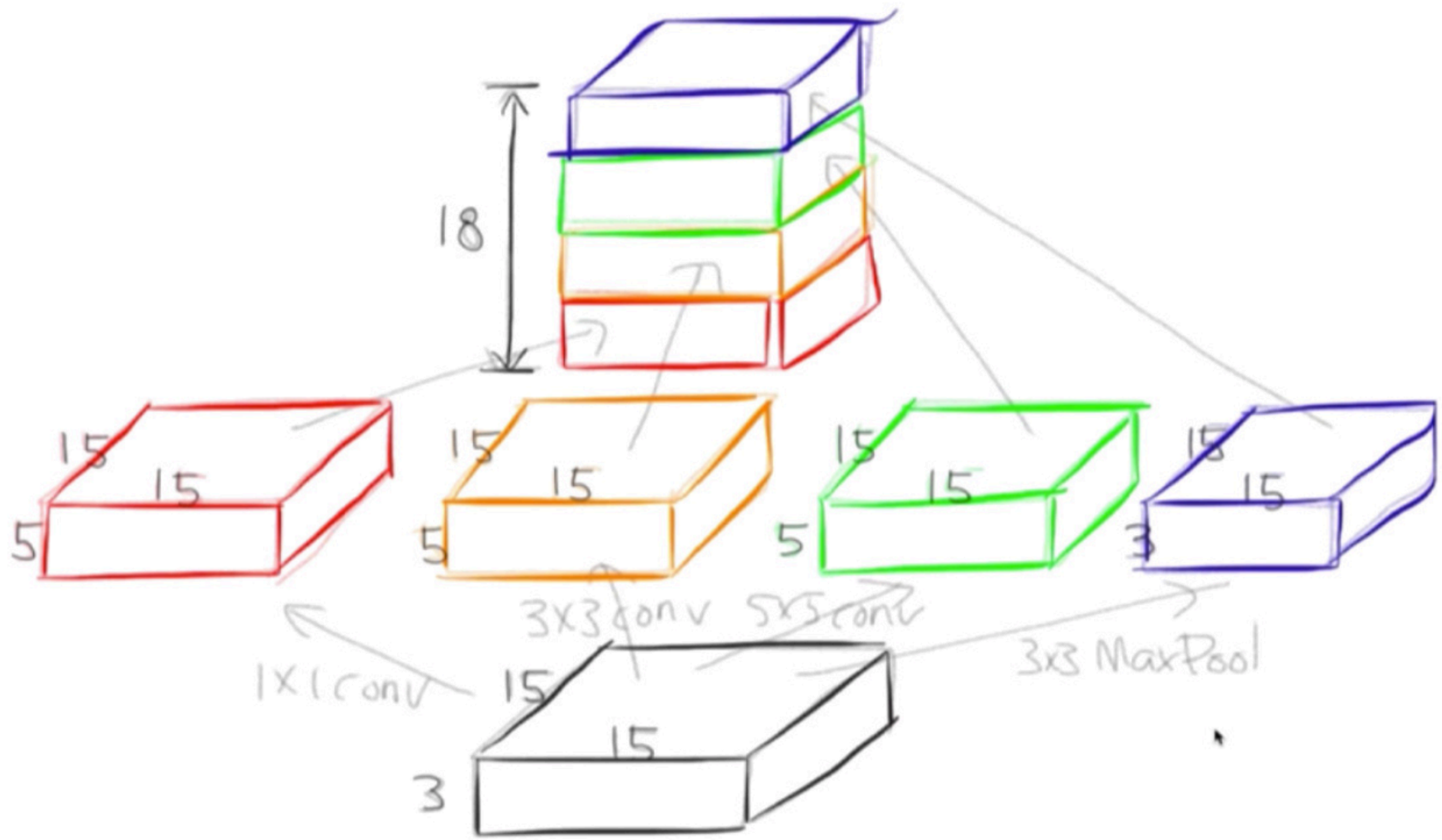


(b) Inception module with dimensionality reduction

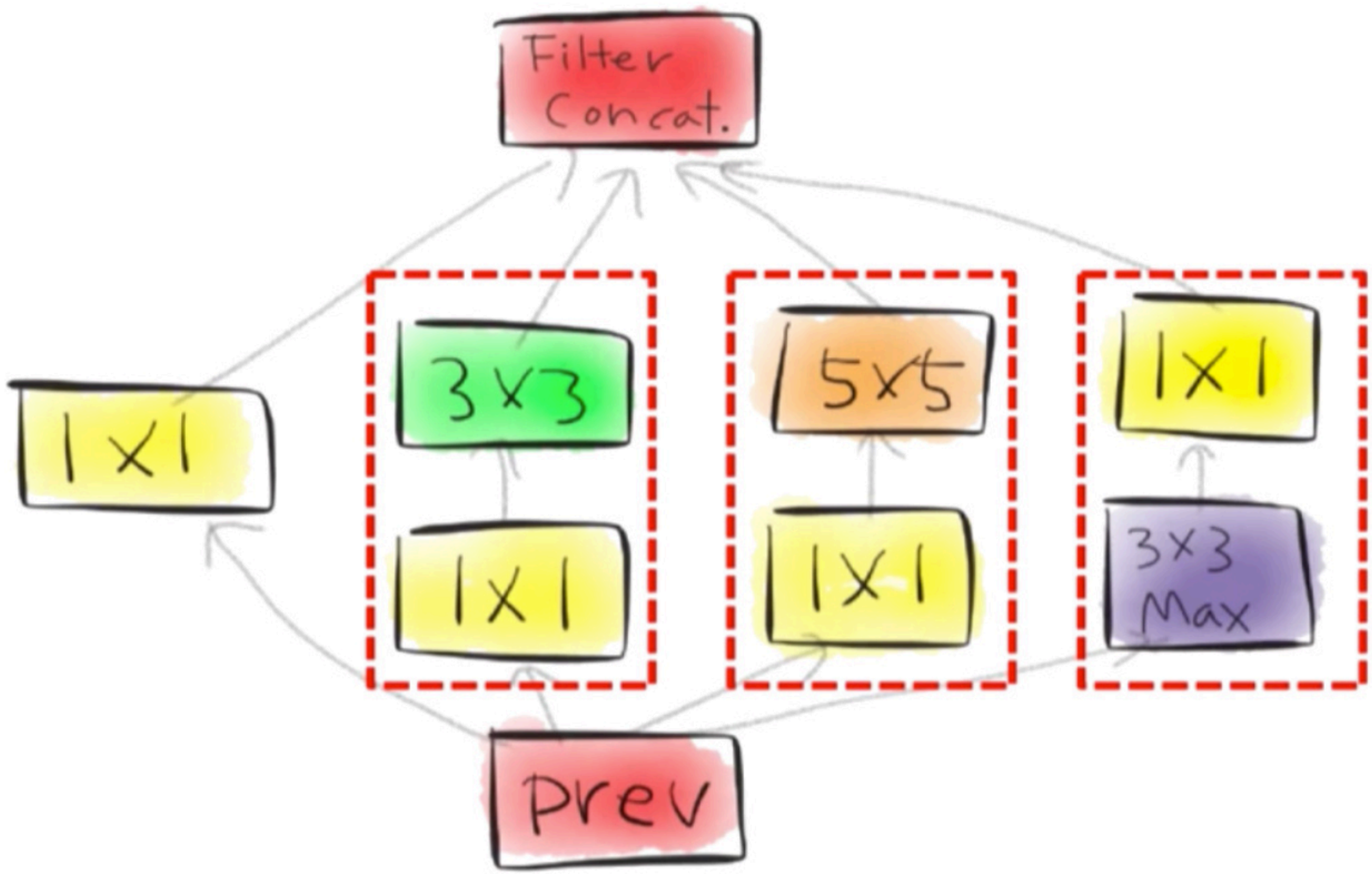
Naïve inception module



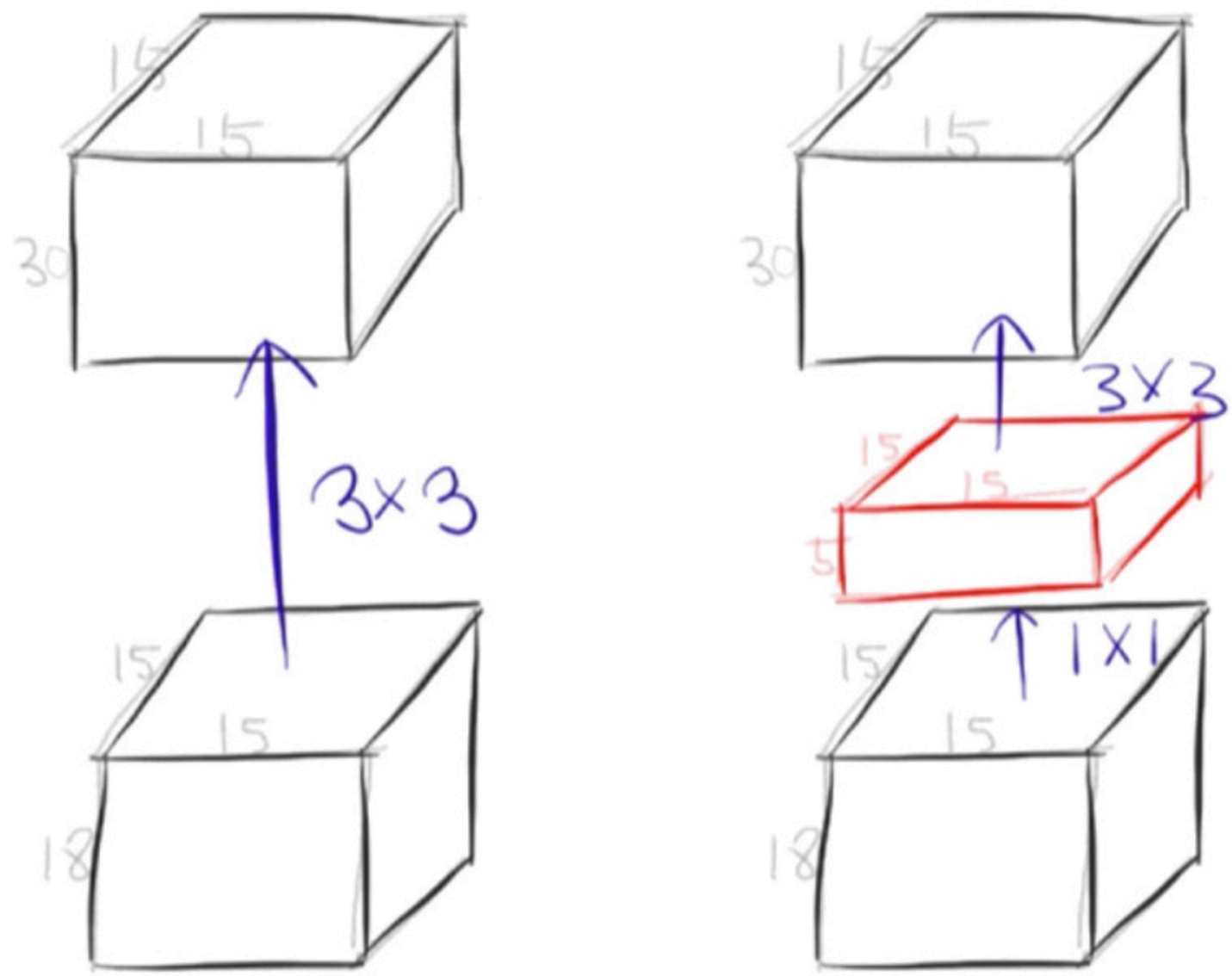
Naïve inception module



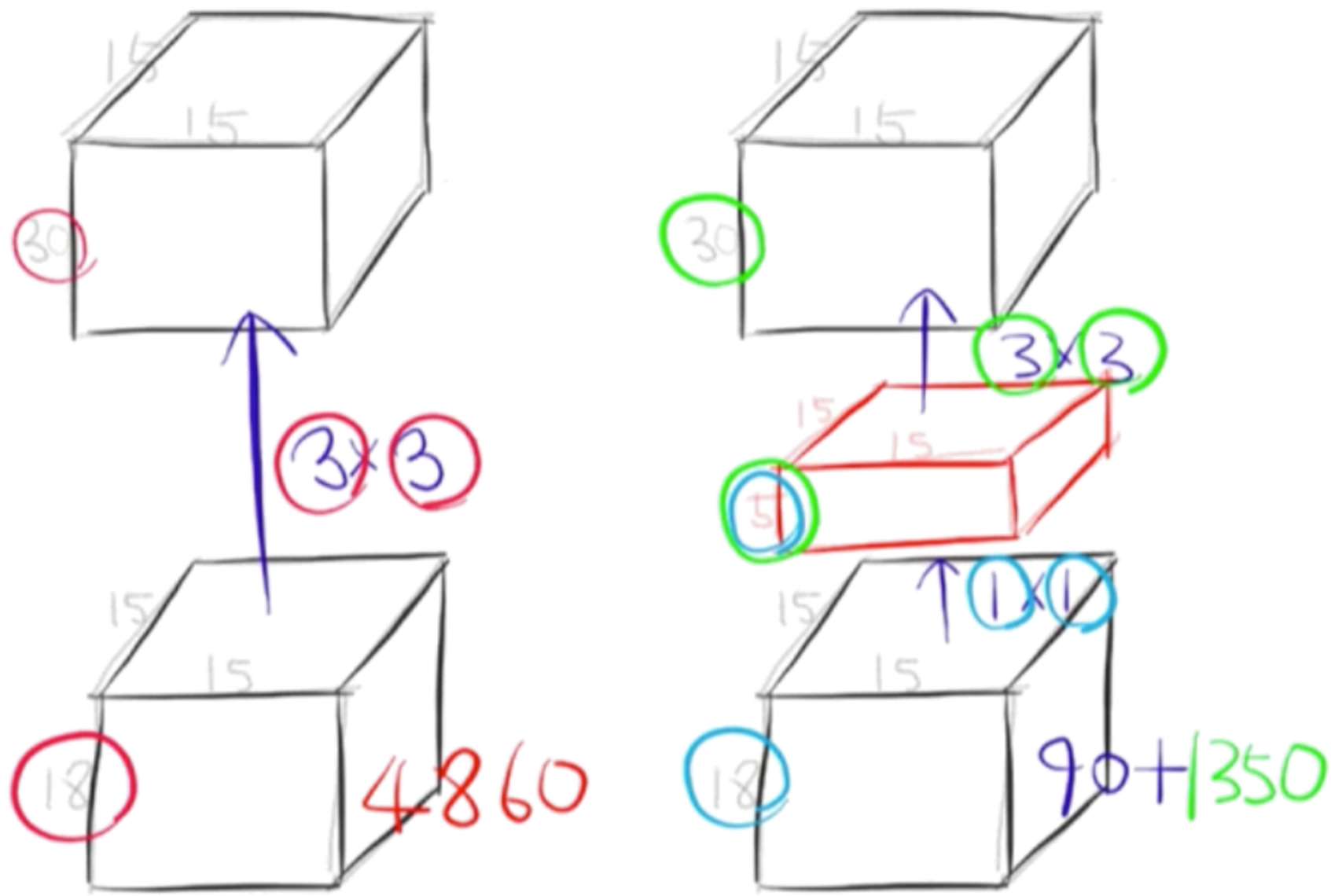
Actual inception module



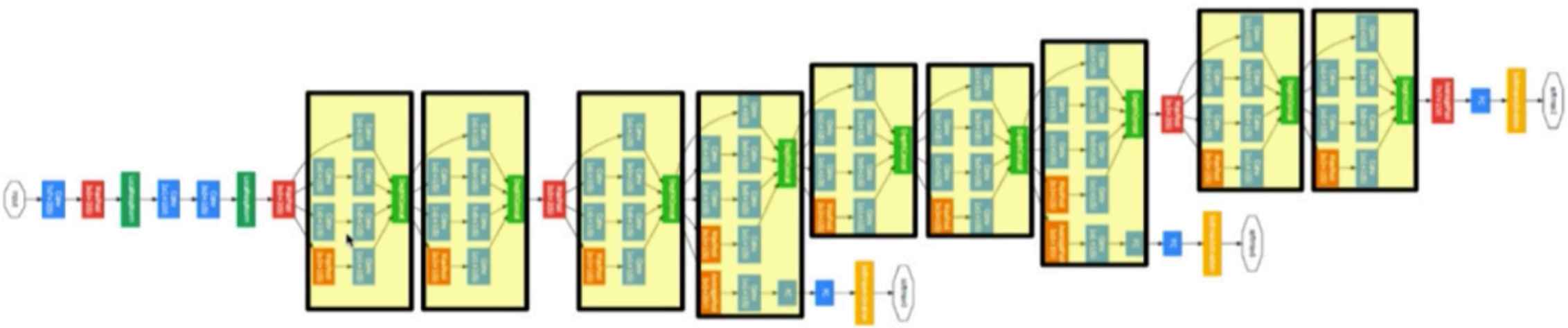
One by one convolution



One by one convolution

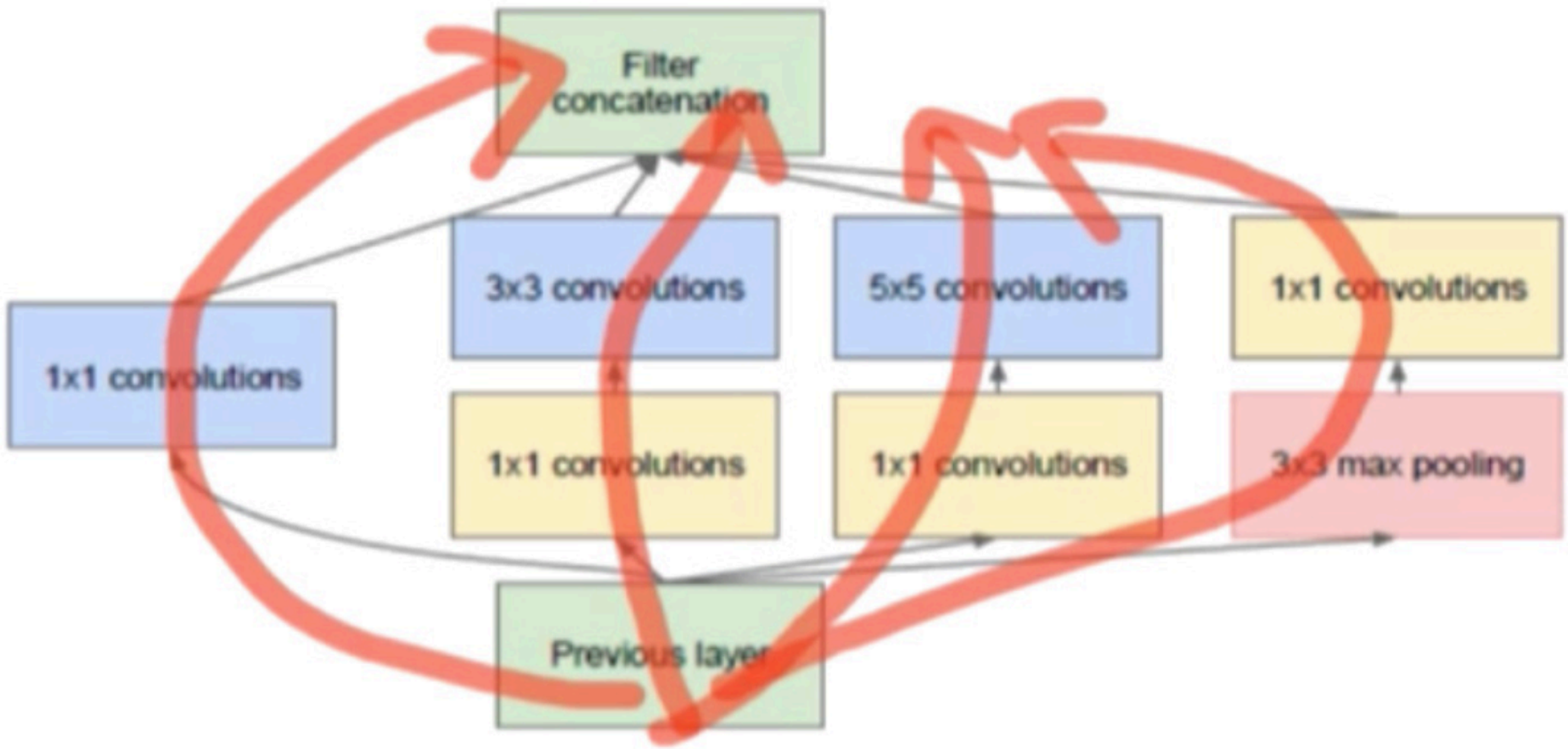


GoogLeNet



Network in Network!

GOOGLENET



(b) Inception module with dimensionality reduction

NEXT TIME

- ResNet
- Overfitting
- Regularization