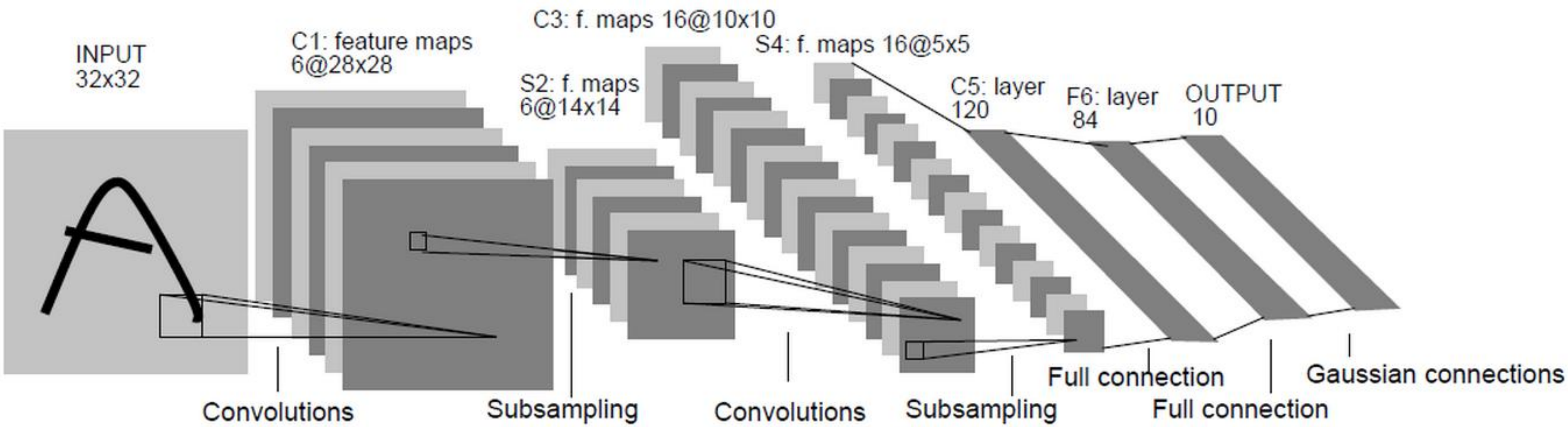
Some Popular Models of CNN

Reporter : Yihan Liu

Time : 2017.1.19

LeNet *Gradient-based learning applied to document recognition 1998*



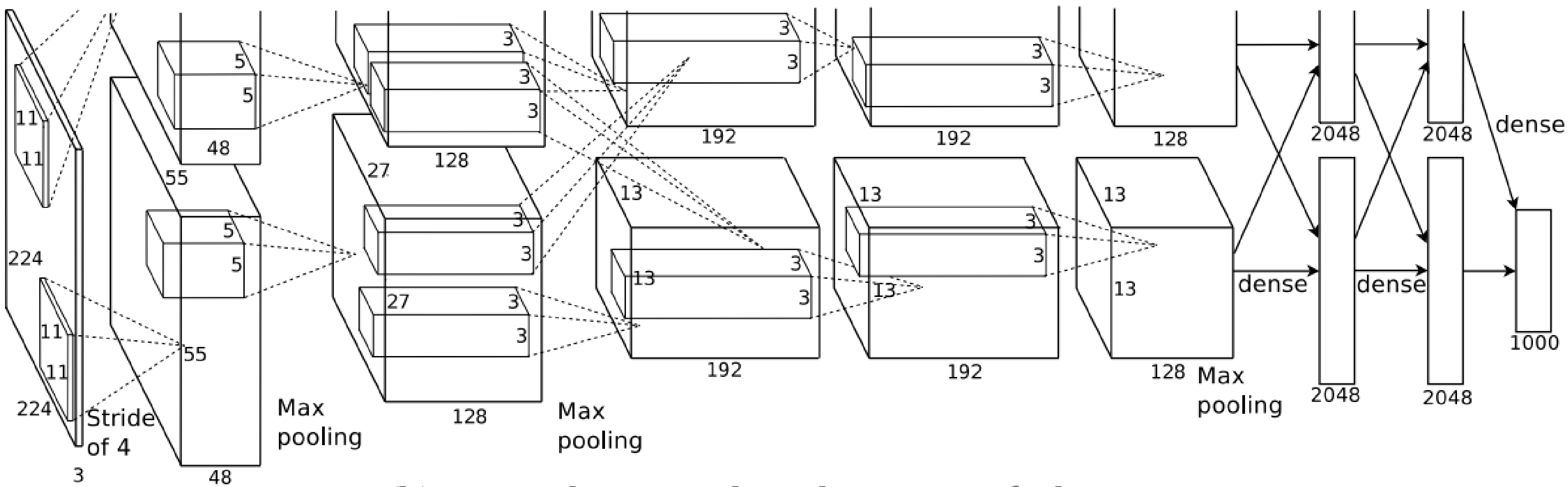
(a) Example network architecture of LeNet

- ✓ The birth of CNN : Y. LeCun, 1998
- ✓ 2 convolutional layers + 3 full-connected layers(including the last softmax layer)
- ✓ Handwriting-recognition model, used in bank
- ✓ Some blogs are available about compute the number of training parameters and connection

Comparison

Models	AlexNet	VGG	GoogLe Net	ResNet
Time	2012	2014	2014	2015
Layer_num	8	19	22	152
Top-5 Error	16.4%	7.3%	6.7%	3.57%
Data Augmentation	+	+	+	+
Inception	-	-	+	-
ConvLayer_num	5	16	21	151
Kernel_size	11,5,3	3	7,1,3,5	7,1,3,5
Dropout	+	+	+	+
LRN	+	-	+	-

AlexNet *ImageNet Classification with Deep Convolutional Networks 2012*



(b) Example network architecture of AlexNet

- ✓ The champion of ILSVRC-2012 image classification task
- ✓ 5 convolutional layers + 3 full-connected layers
- ✓ Max pooling layers are followed by conv1, conv2, conv5, and dropout are existed after the last two full-connected layers
- ✓ Has the maximum parameters(60M)
- ✓ To reduce overfitting : data augmentation, dropout
- ✓ Success factors : ReLU, multi-GPUs parallelization, LRN, overlapping pooling

VGG Net *Very Deep Convolutional Networks For Large-Scale Image Recognition 2014*

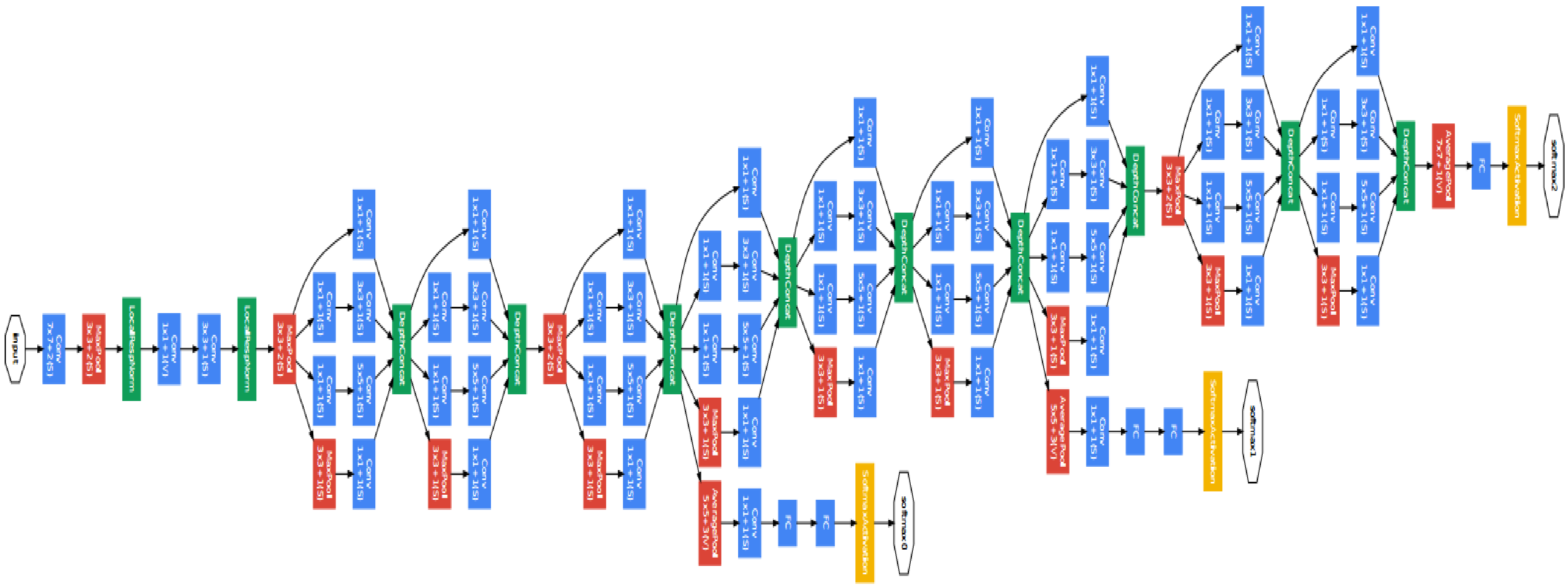
ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	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					

(c) Example network architecture of VGG Net

- ✓ Champion of ILSVRC-2014 object localization task and runner-up of classification task
- ✓ Has A-E 6 versions and the deepest has 19 weight layers
- ✓ Pre-training : train A until stabilize and then train B based on A, and so on
- ✓ Not the deeper the better :
  D achieves the best performance

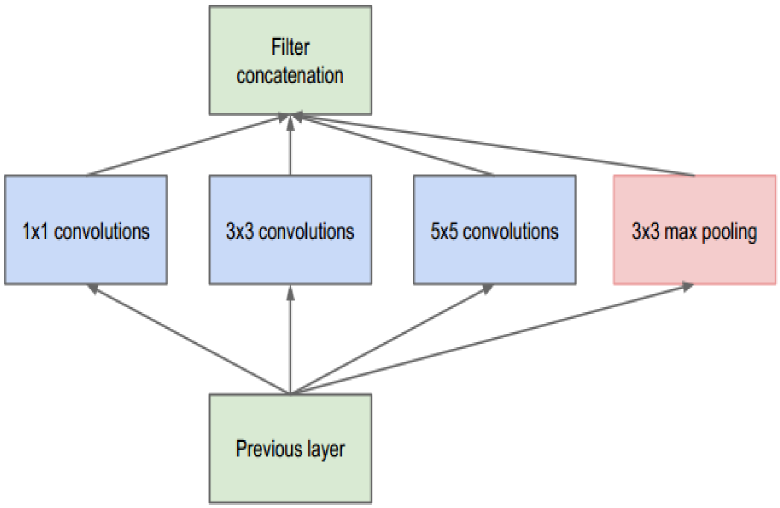
conv3-512 : 3 → 3*3 receptive filed ;
512 → 512*512 feature map

GoogLeNet *Going Deeper with Convolutions 2014*

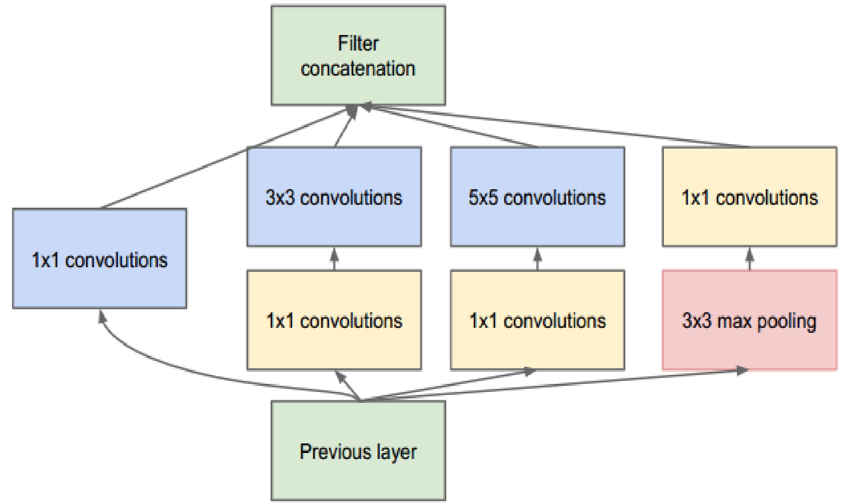


(d) Example network architecture of GoogLeNet

GoogLeNet *Going Deeper with Convolutions 2014*



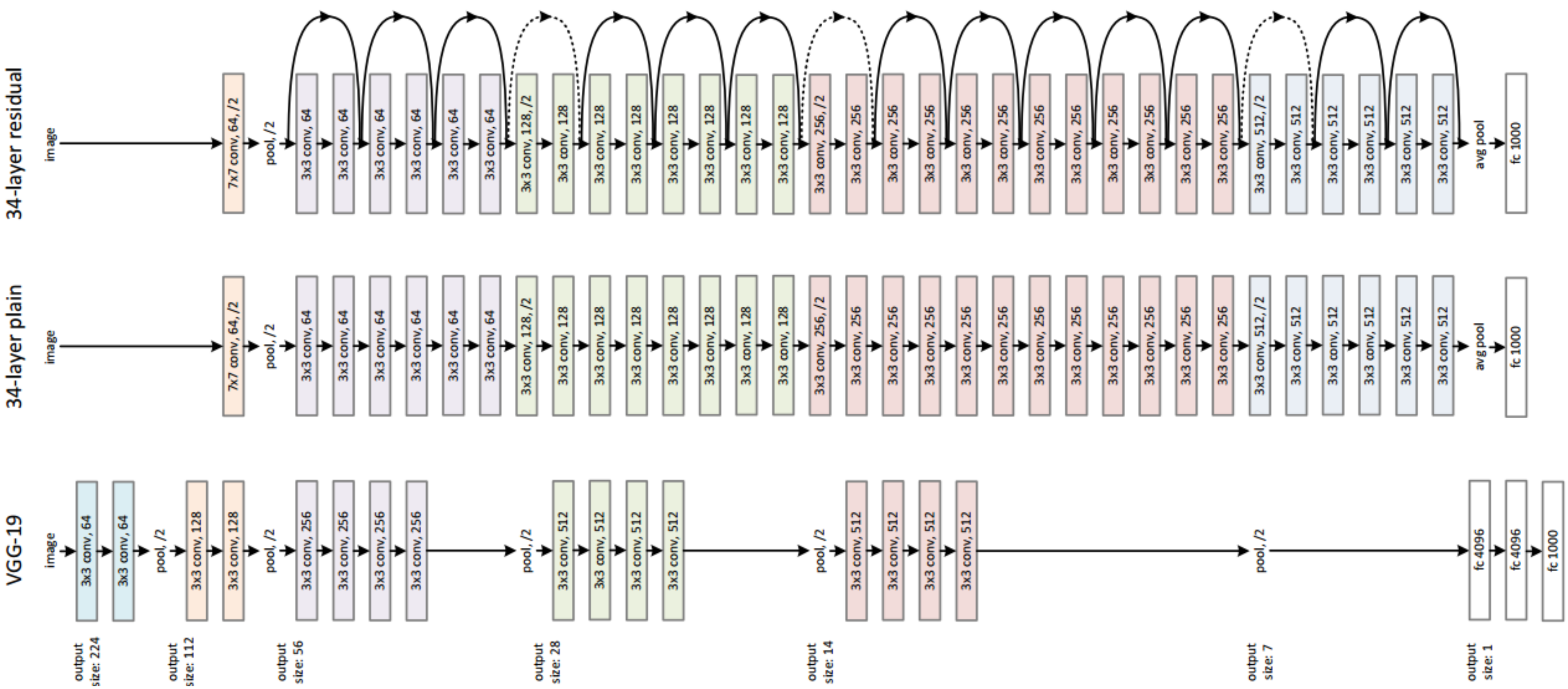
(e) Inception module, na ïve version



(f) Inception module with dimensionality reduction

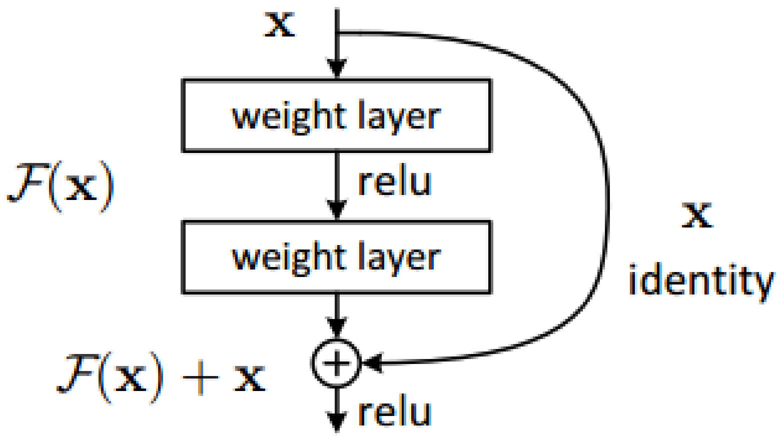
- ✓ The champion of ILSVRC-2014 classification and detection task
- ✓ Depth : more deeper(22 layers), add two loss to overcome gradient vanish
- ✓ Width : based on the intuition of multi-scale processing, GoogLeNet adds kernels that size are 1, 3, 5(e). And uses inception decrease thickness of fusion feature map(f).

ResNet Deep Residual Learning fro Image Recognition 2015



(g) Example network architecture of ResNet

ResNet *Deep Residual Learning fro Image Recognition 2015*



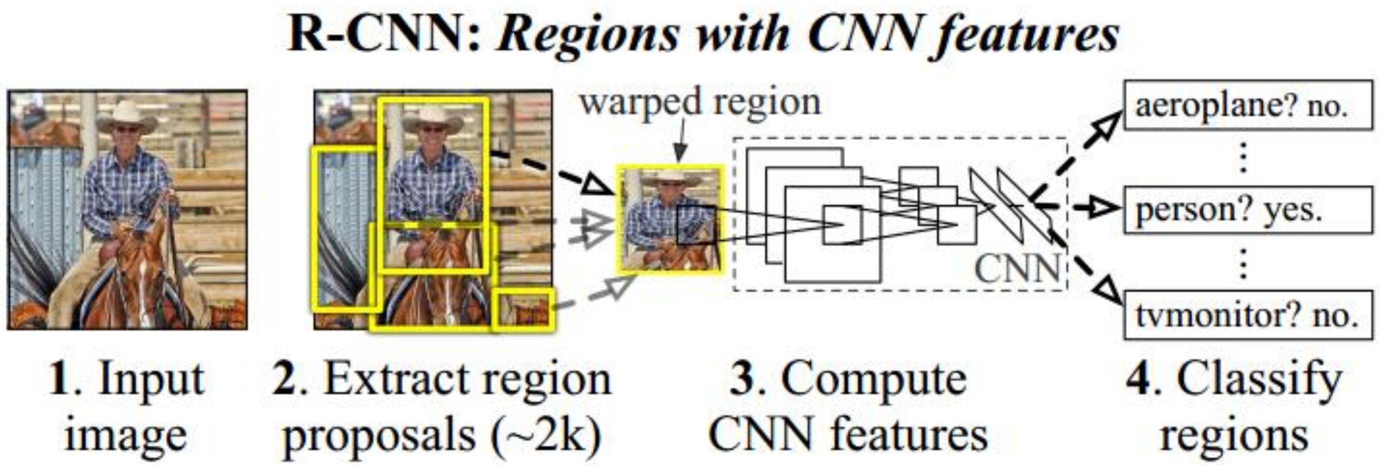
(h) Residual learning: a building block

model	top-1 err.	top-5 err.
VGG-16 [41]	28.07	9.33
GoogLeNet [44]	-	9.15
PReLU-net [13]	24.27	7.38
plain-34	28.54	10.02
ResNet-34 A	25.03	7.76
ResNet-34 B	24.52	7.46
ResNet-34 C	24.19	7.40
ResNet-50	22.85	6.71
ResNet-101	21.75	6.05
ResNet-152	**21.43**	**5.71**

(i) Error rates on ImageNet validation

- ✓ The champion of ILSVRC-2015 classification task
- ✓ The deepest network – 152 layers
- ✓ During deep ConvNet going deeper, it occurs underfitting. For this reason, ResNet puts forward skipping connection which consists of residual learning. It turns out to be useful when compared plain-34 with ResNet-34, referring to (i)

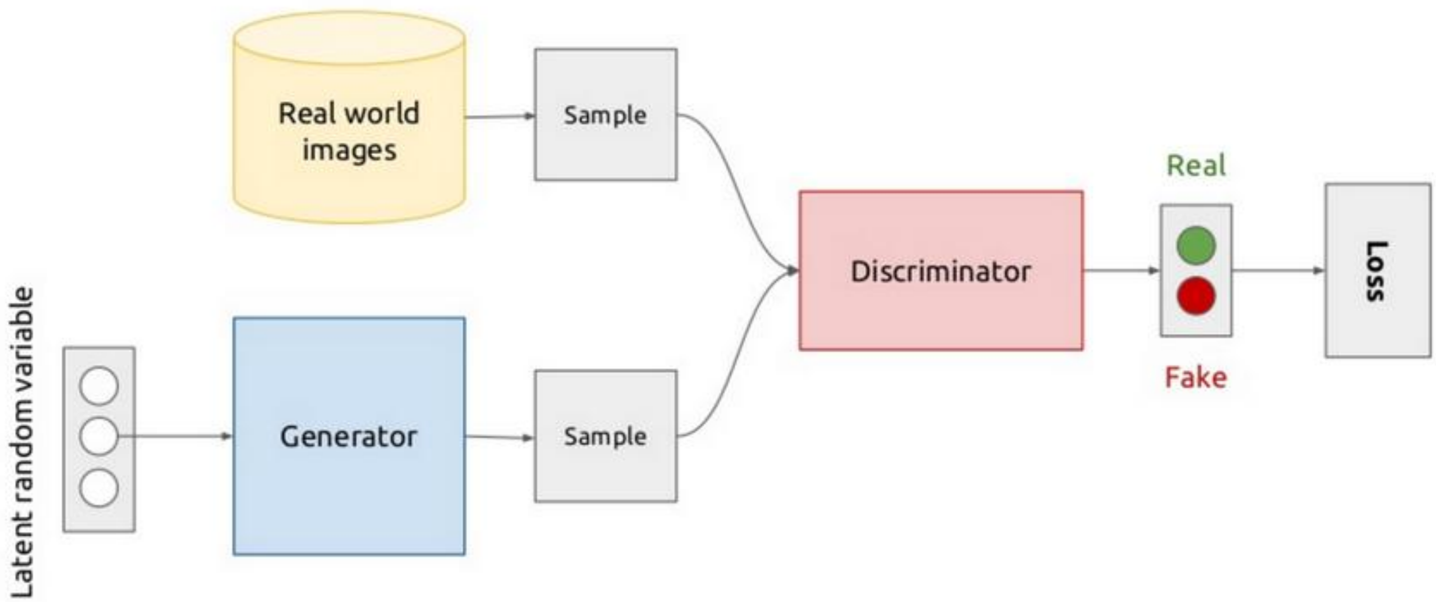
R-CNN *Rich feature hierarchies for accurate object detection and semantic segmentation 2014*



(j) Object detection system overview

- ✓ Solves object detection : location(regression) + classification
- ✓ Location → extract region proposals → selective search
- ✓ Use AlexNet to compute CNN features for each region proposal, then these region proposals compose training dataset. The dataset is separated as positive and negative based on IoU between ground-truth box.
- ✓ Evolution:
  - R. Girshick, "Fast R-CNN," in IEEE International Conference on Computer Vision, 2015
  - S.Ren, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", in Neural Information Processing Systems (NIPS), 2015

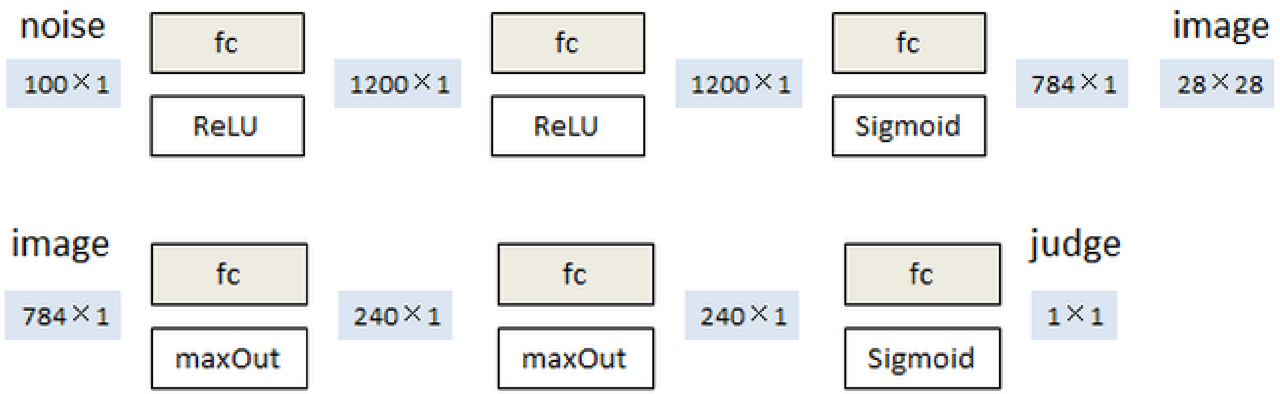
GAN Generative Adversarial Nets 2014



(k) Example network architecture of GAN

- ✓ Train generator and discriminator network at the same time : maximize distinguish error when training generator network and minimize distinguish error when it's discriminator network

GAN *Generative Adversarial Nets 2014*



(l) Example network architecture of GoogLeNet

- ✓ The generator and discriminator network consist of full-connected and convolutional layers. (l) is a example for MNIST dataset :
- ✓ Evolution:
  - Conditional Generative Adversarial Nets
  - Deep Generative Image Models using a Laplacian Pyramid of Adversarial Networks
  - Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks

Thank You !