

2. CNN

Dong-Gyu, Lee

Dept. of Statistics, KU

2020, Jan 21

Contents

1 Analysis Goal

2 MLP

3 CNN

4 Advanced CNN

5 Next Time

6 Reference

7 Terminology

Today's Goal

- We will compare the performance of the imbalanced and balanced image data to the neural network model.
- Especially, the "Convolutional Neural Networks(CNN)" models are known for their high performance in image classification.
- Put imbalanced and balanced data aside, in this time, we will look at the various CNN models and their details.

MLP

- Before understanding the CNN models, you should be familiar with the basic Neural Network model. It is called the "Multilayer Perceptron(MLP)", as one case of "Deep Neural Network(DNN)".

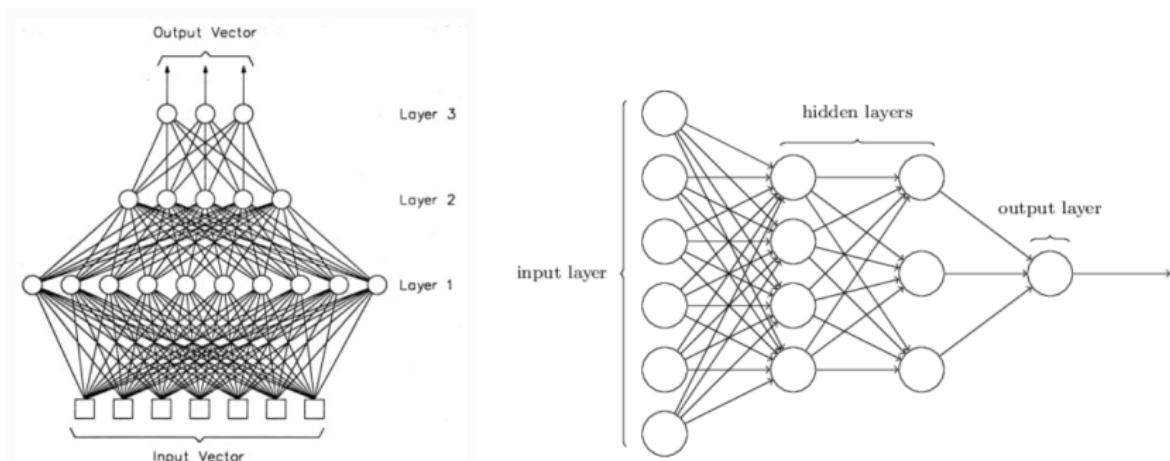
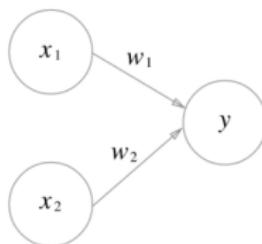


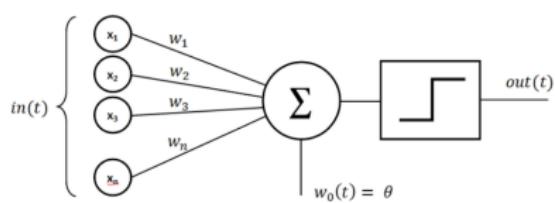
Figure 1: The MLP Sturctures

MLP

- The Perceptron proposed by Frank Rosenblatt in 1957 has the following structure:



(a) Perceptron1



(b) Perceptron2

Figure 2: Perceptron

- This can be considered as a matter of simply drawing a straight line and classifying it. And stacking them together(MLP) can solve problems like "XOR".

MLP

- The MLP model is also called "Feed-forward Deep Neural Network".
- Besides, the MLP structure's one layer in Figure 1 is called a "Fully Connected Layer" or "Dense Layer".
- Basically, the layers are consists of three parts : Input Layer, Hidden Layer, Output Layer.
- The circle shape is called the Node and the lines mean the Weights. And all of the weights are parameters for learning.
- Therefore, in the MLP structure, these lines(weights) are considered a model.
- The one thing to note is that the model complexity increases exponentially with the depth of the model.[3]

CNN: LeNet5

- The CNN model was contrived by "Yann Lecun" who was inspired by "Cat Thought Experiment" conducted by "David H. Hubel" and "Torsten Wiesel" in 1958.
- In 1998, professor Yann. invented LeNet5 which was the ancestor of CNN and used for mail classification.

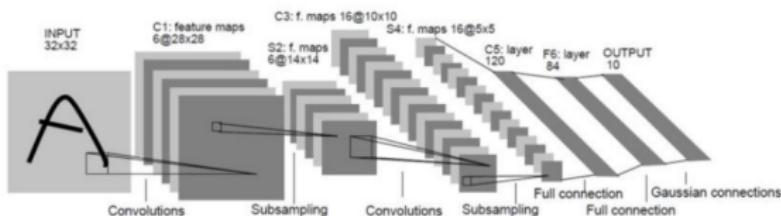


Figure 3: LeNet5

CNN: Model

- The following Figure 4 is the basic form of the CNN model.

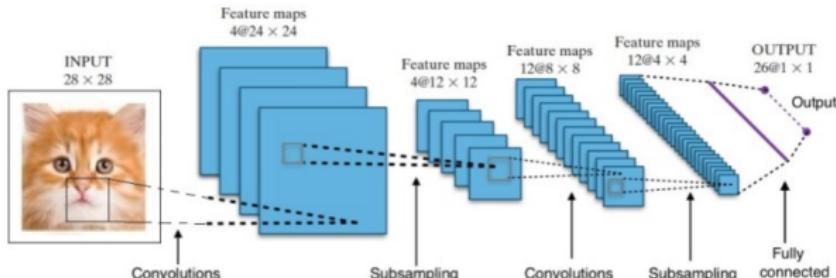


Figure 4: The CNN Model

- To understand the CNN well, you acquaint the following terms : **Convolution, Filter, Feature Map, Stride, Padding, Pooling, Activation Function, etc.**

CNN: 1. Convolution, Filter, Feature Map

- The convolution is the act of multiplying and adding numbers easily.
- Let's think of the green square as an image and the yellow square as a Filter. Then, the convolution operation proceeds like Figure 4 and the output is called Feature Map.

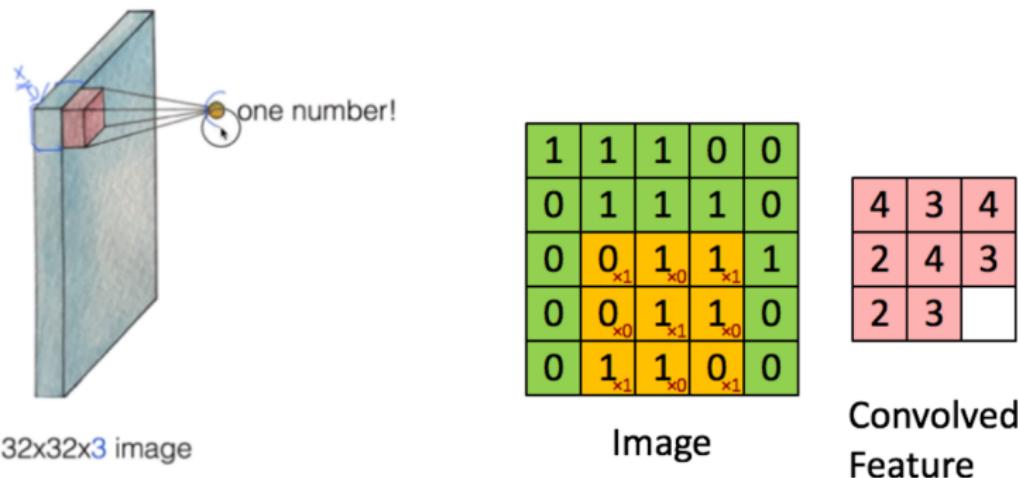
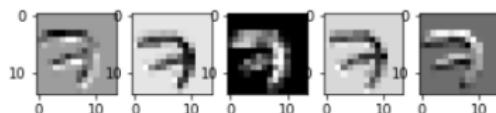


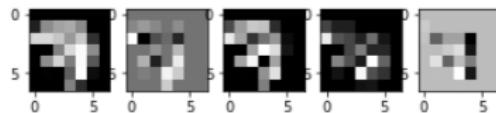
Figure 5: Convoluton operation

CNN: 1. Convolution, Filter, Feature Map

- The following Figure 5 shows the change of the image data(MNIST data) when the Convolution operation is performed one time.



(a) Before



(b) After

Figure 6: Image Change

CNN: 2. Stride

- In the following figure, you can see the Filter, Feature Map, and Stride: Filter = 3×3 ($\times 3$ if RGB), Stride = 2, Feature Map = 3×3 ($\times 3$ if RGB)

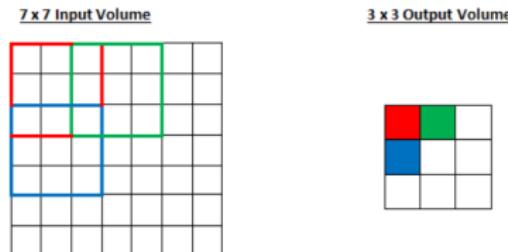


Figure 7: Stride with size 2

- Remember we are working on the one image.

CNN: 3. Padding

- Also, sometimes we want to make the same size feature map as the input or adjust it arbitrarily.
- In this case, we use the Padding(or Zero Padding) method.

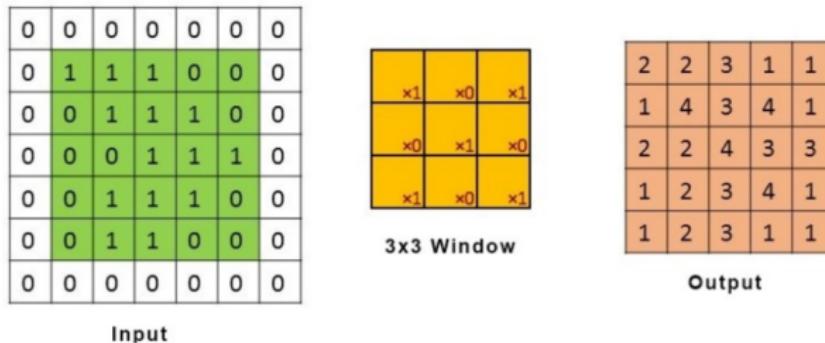


Figure 8: Zero Padding with size 1

- In the above, the Padding Size is 1

CNN: 4. Equation

- There is a relationship between Input(image), Output(Feature Map), Stride, Padding :

$$O = \frac{(I - F + 2P)}{S} + 1$$

O = Output, I = Input, F = Filter, P = Padding

- Also if you want to make the same size Feature Map as the input, use the following equation.

$$P = \frac{(F - 1)}{2}$$

P = Padding, F = Filter

- The above formulas are calculated for the horizontal and vertical directions respectively

CNN: 5. Pooling

- The Pooling is used to compress the information.
- This results in preventing the model from overfitting and avoiding the use of many parameters.
- There are Max Pooling, Mean Pooling methods and so on.

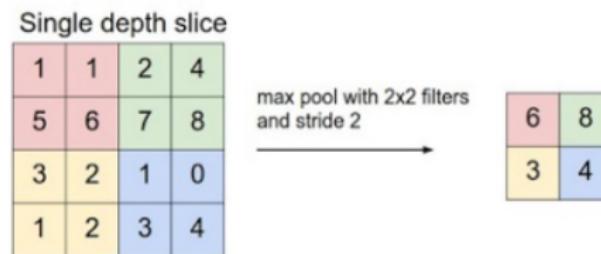


Figure 9: Max Pooling

CNN: 6. Activation Function

- The activation function makes the data nonlinear.
- The CNN model usually uses Rectified Linear Unit(ReLU) as an activation function for avoiding gradient vanishing.
- Of course, there are many other activation functions : Tanh, Maxout, LeakyReLU, ELU, and so forth.

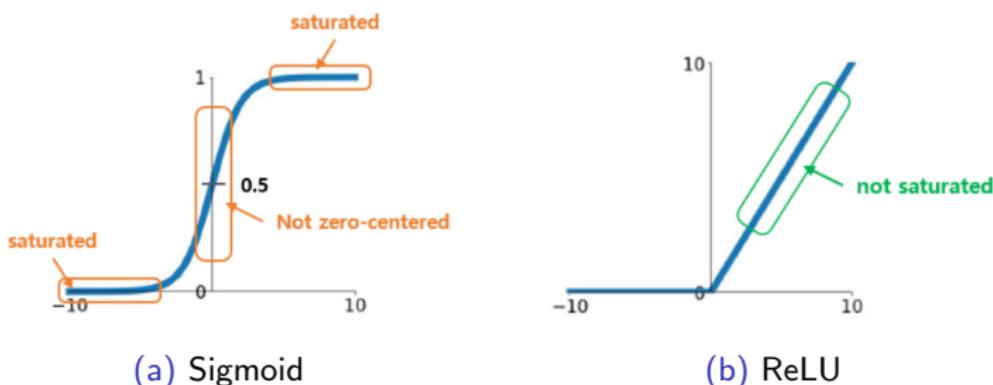


Figure 10: Activation Functions

CNN: 7. More

Let's take a look at the following terms simply.

- Cost Function
 - ① This is a function related to our goal. It is important to update this value low.
 - ② According to purpose, use the loss function appropriately and variously.
- Gradient Descent Algorithm
 - ① It is algorithm for updating parameters.
 - ② The basic form of Gradient Descent updating equation is $w_{k+1} = w_k - \lambda_k \nabla f(w_k)$ for each parameters. w is a parameter and f is a Cost Function(Convex, Differentiable).
- Backpropagation
 - ① This method is used to calculate the gradient values for which nodes affect the Cost Function.
 - ② Basically, it is calculated through the Chain Rule.

CNN: Summary

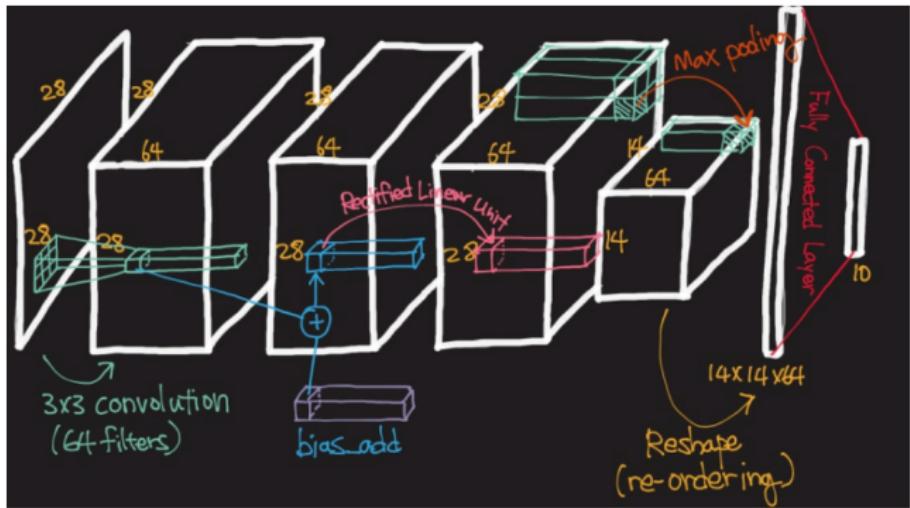


Figure 11: CNN Layer Structure

CNN: Summary

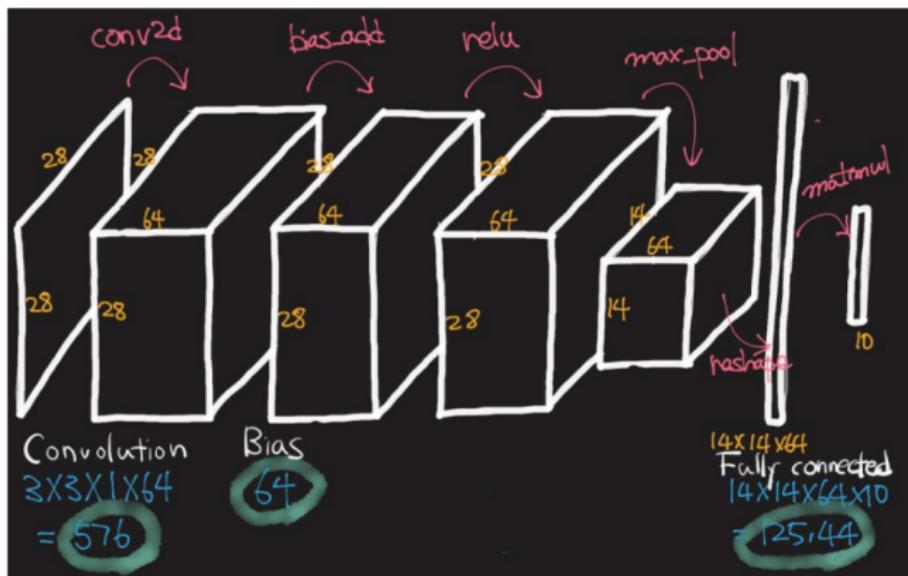


Figure 12: Parameters Calculation

CNN: Summary

Machine can do it!



Figure 13: Dogs and Cats

Advanced CNN: ILSVRC Contest

- ImageNet Large Scale Visual Recognition Challenge(ILSVRC), launched in 2010 first, is a contest to evaluate the performance of image classification algorithms using large amounts of image data (1000 categories, 1.4 million data).

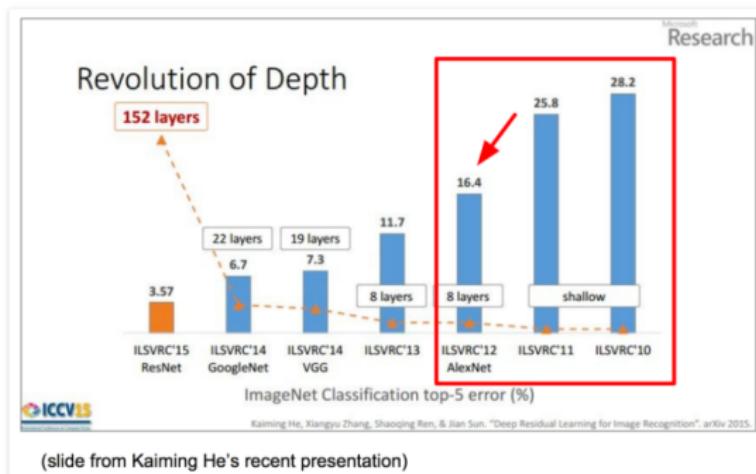


Figure 14: ILSVRC Models

Advanced CNN: ILSVRC Contest

- You should know the AlexNet, VGG19, GoogLeNet, and ResNet CNN models.
- We will focus on the AlexNet and VGG19 this time.

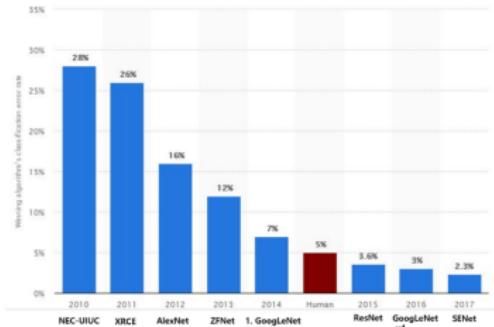
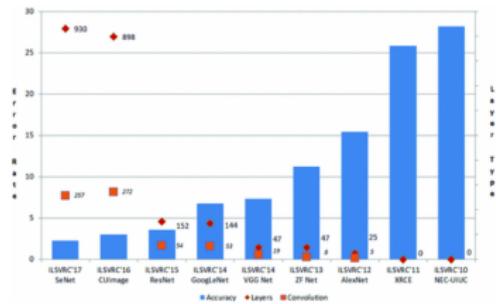


Figure 15: Error Rate until 2017

Advanced CNN: AlexNet

- The Alexnet[2] accelerated the development of the CNN model for image classification problems by using **ReLU**, **LRN**, **Data Augmentation**, **Dropout** and so on.
- Because the Alexnet used more deeper Network than other models at that time, people began to be increasingly interested in the depth of the overall Neural Network.

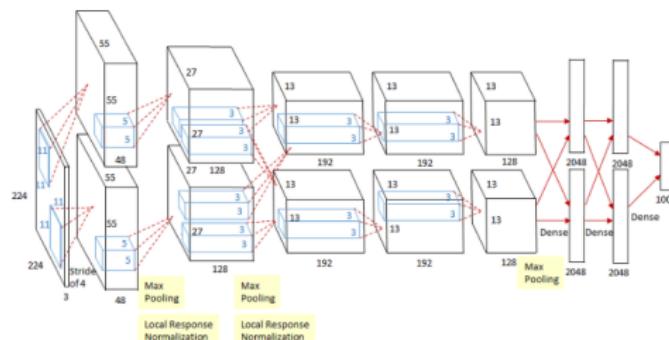


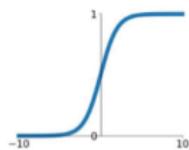
Figure 16: AlexNet Structure

Advanced CNN: AlexNet

- There are many Activation Functions. And we will focus on the ReLU.

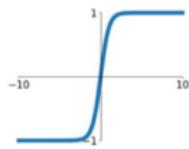
Sigmoid

$$\sigma(x) = \frac{1}{1+e^{-x}}$$



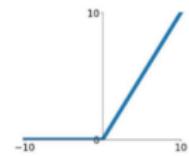
tanh

$$\tanh(x)$$



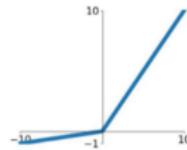
ReLU

$$\max(0, x)$$



Leaky ReLU

$$\max(0.1x, x)$$



Maxout

$$\max(w_1^T x + b_1, w_2^T x + b_2)$$

ELU

$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$

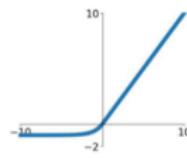


Figure 17: Activation Functions

Advanced CNN: AlexNet

Let's take a look at the following terms.

- Local Response Normalization(LRN)
 - ① The LRN is a kind of normalization method based on ideas from neurobiology. It makes a node more stronger when the node is relatively strong to be compared with its surroundings.
 - ② This method isn't used nowadays.
- Data Augmentation
 - ① Data Augmentation means an expansion of the number of data.
 - ② Cut the image appropriately and perform "Crop" or "Flip".
 - ③ The Color Variation(Jittering) is performed by fitting PCA.
- Dropout
 - ① Dropout is the act of randomly turning off a node during the training.
 - ② It is often used for prevent models from overfitting.

Advanced CNN: AlexNet

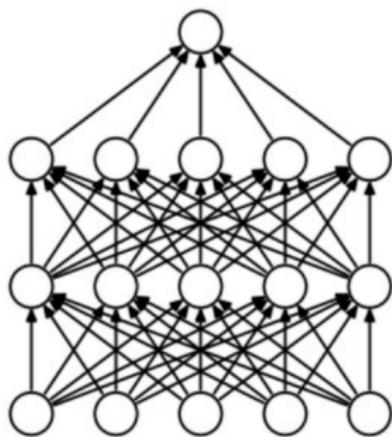
- LRN

$$b_{x,y}^i = a_{x,y}^i / \left(k + \alpha \sum_{j=\max(0,i-n/2)}^{\min(N-1,i+n/2)} (a_{x,y}^j)^2 \right)^\beta$$

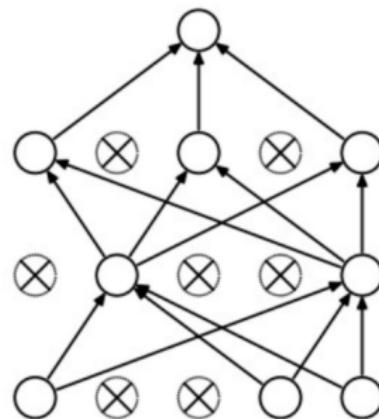
- ① $a_{x,y}^i$: The activity of a neuron computed by applying kernel i at position (x, y) .
- ② $b_{x,y}^i$: The response-normalized activity.
- ③ n : 'Adjacent' kernel maps at the same spatial position.
- ④ N : The total number of kernels in the layer.
- ⑤ k, n, α, β : Hyper parameters.
- ⑥ AlexNet used $k = 2$, $n = 5$, $\alpha = 10^{-4}$, and $\beta = 0.75$.

Advanced CNN: AlexNet

- Dropout



(a) Standard Neural Network



(b) Applying Dropout

Figure 18: Dropout: A Simple Way to Prevent Neural Networks from Overfitting

Advanced CNN: AlexNet

- The AlexNet consists of 5 Convolution layers and 3 Fully-connected layers.
- It has almost 63,000,000 parameters.

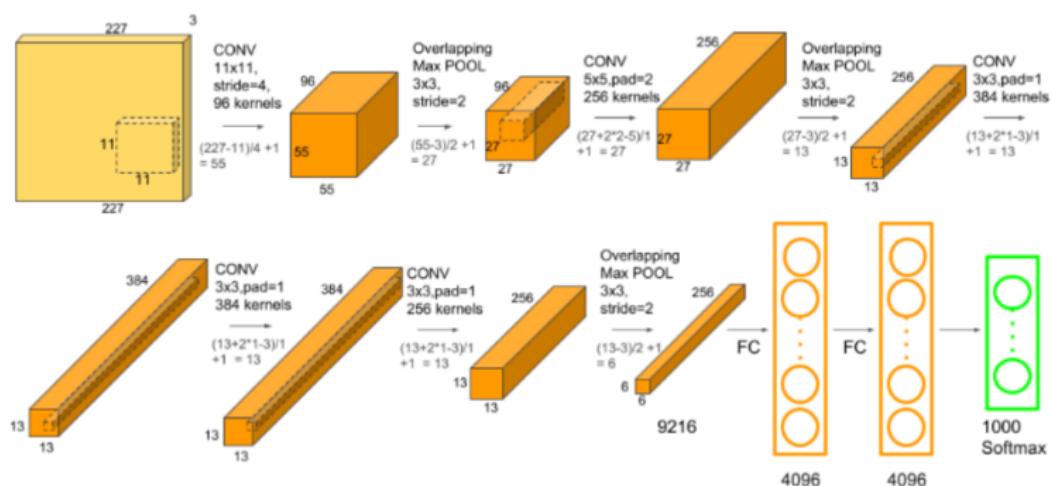


Figure 19: Specific AlexNet

Advanced CNN: VGG19

- Since the emergence of the Alexnet, people have begun to be concerned about how the depth of Neural Network models affects performance.
- the VGG19, which was runner-up at ILSVRC in 2014, is simpler than GoogLeNet.
- For this reason, properly modified VGG model [4] has been widely used.

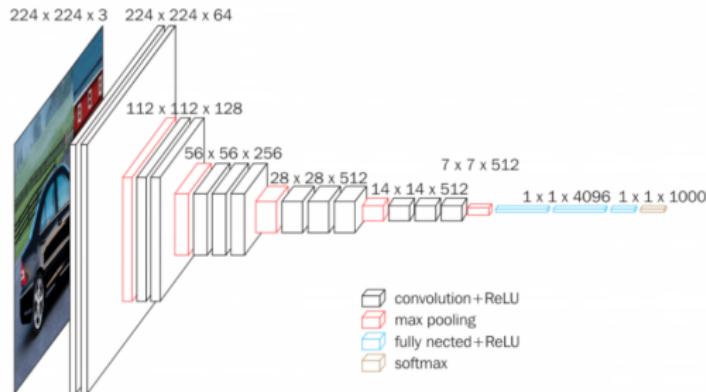


Figure 20: VGG19 Structure

Advanced CNN: VGG19

- VGG models are characterized by using only 3×3 Filters.
- That's because of 2 reasons : more nonlinearity and fewer parameters.

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					

Figure 21: VGG models

Advanced CNN: VGG19

- As VGG models becomes deeper, Gradient Vanishing and Overfitting problems become greater
- In the VGG models article, team VGG corrected this problem using the results obtained with prior-learning through the VGG11.

Advanced CNN: And..

- Moreover, there are relatively new models GoogLeNet and ResNet.

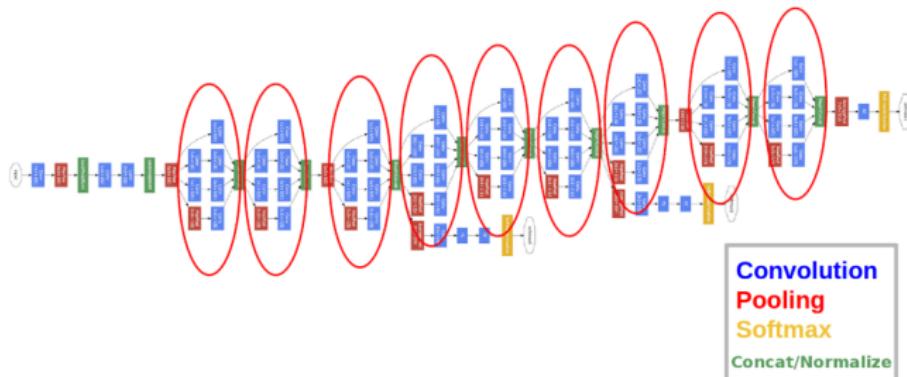


Figure 22: GoogLeNet

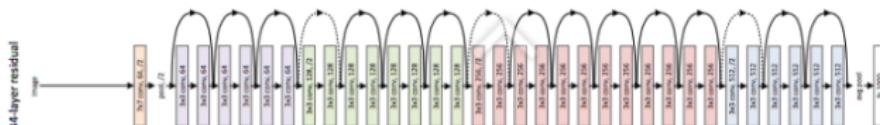


Figure 23: ResNet

Advanced CNN: And..

- Although not covered in this ppt, you can check each model through the following : GoogLeNet[5] , ResNet[1]

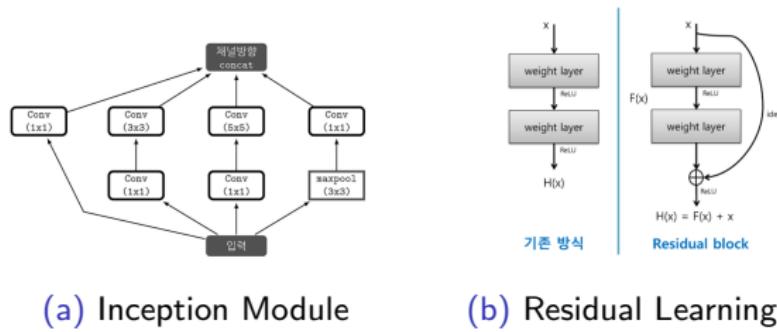


Figure 24: Key concepts in each model

- In preprocessing, data augmentation can be expressed in various ways :
 - ① Flip all data left and right for small number of labels.
 - ② Make an object(data) through the rotation or translation invariance (Crop).
 - ③ Use the ADASYN, SMOTE, GAN , etc.
 - ④ Balance the label with other data.
- Also we will cover the various model such as the Alexnet, VGG19, ResNet18 etc.
- And last, we will talk about the various ensembles like voting system, through the combination of above models.

Reference

- [1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [2] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [3] Maithra Raghu, Ben Poole, Jon Kleinberg, Surya Ganguli, and Jascha Sohl Dickstein. On the expressive power of deep neural networks. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 2847–2854. JMLR.org, 2017.
- [4] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [5] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

Terminology

- z : latent random vector