NPTEL ONLINE CERTIFICATION COURSES

Course Name: Deep Learning

Faculty Name: Prof. P. K. Biswas

Department : E & ECE, IIT Kharagpur

Topic

Lecture 40: Popular CNN Models IV

CONCEPTS COVERED

Concepts Covered:

- CNN
  - AlexNet
  - VGG Net
  - Transfer Learning
  - Challenges in Deep Learning
  - GoogLeNet
  - ResNet
  - etc.

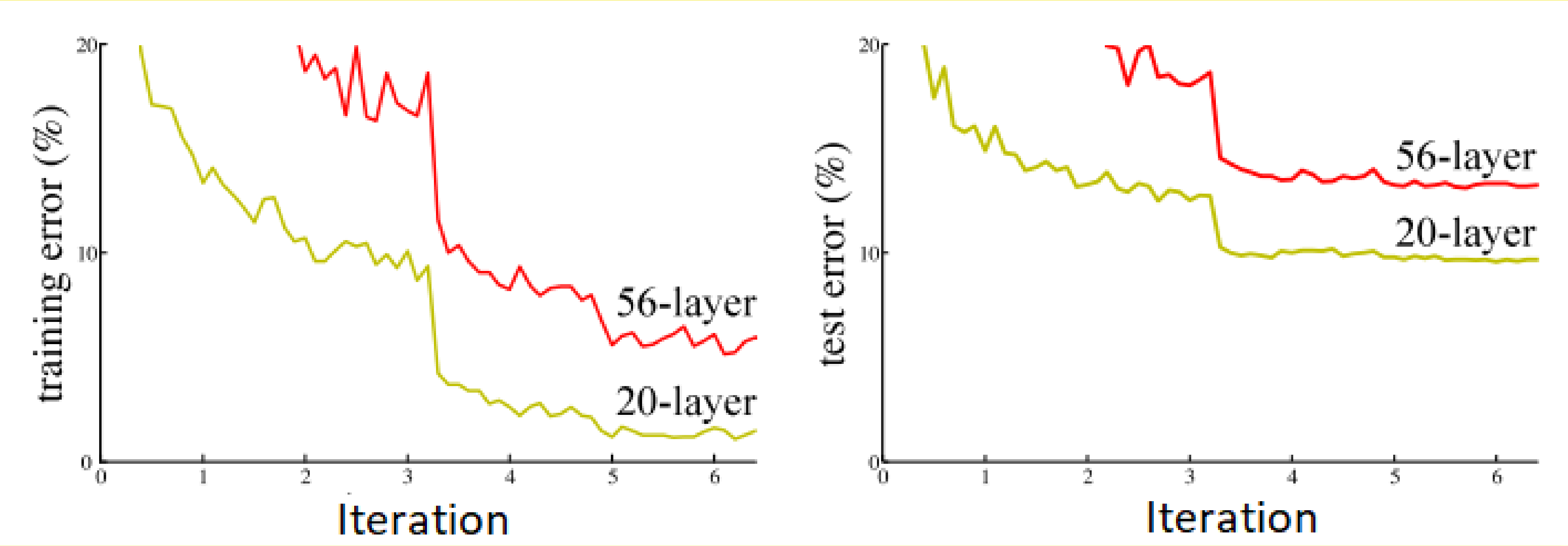
Deep Learning Challenges

Challenges

- ❑ Deep learning is data hungry.
- ❑ Overfitting or lack of generalization.
- ❑ Vanishing/Exploding Gradient Problem.
- ❑ Appropriate Learning Rate.
- ❑ Covariate Shift.
- ❑ Effective training.

Vanishing Gradient

Vanishing Gradient Problem



https://towardsdatascience.com/an-overview-of-resnet-and-its-variants-5281e2f56035

Vanishing Gradient Problem

$$X \xrightarrow{W_1} f_1 \xrightarrow{W_2} f_2 \xrightarrow{W_3} f_3 \xrightarrow{W_4} f_4 \longrightarrow O$$

$$O = f_4(W_4 f_3(W_3 f_2(W_2 f_1(W_1 X))))$$

Vanishing Gradient Problem

$$O = f_4\overbrace{(W_4 f_3 \underbrace{(W_3 f_2 (\overbrace{W_2 f_1(\underbrace{W_1 X}_{\theta_1})}^{\theta_2}))}_{\theta_3})}^{\theta_4}$$

Vanishing Gradient Problem

$$O = f_4(\theta_4) \quad \theta_4 = W_4 f_3(\theta_3) \quad \theta_3 = W_3 f_2(\theta_2) \quad \theta_2 = W_2 f_1(\theta_1) \quad \theta_1 = W_1 X$$

$$\frac{\partial O}{\partial W_1} = \frac{\partial O}{\partial \theta_4}.\frac{\partial \theta_4}{\partial f_3}.\frac{\partial f_3}{\partial \theta_3}.\frac{\partial \theta_3}{\partial f_2}.\frac{\partial f_2}{\partial \theta_2}.\frac{\partial \theta_2}{\partial f_1}.\frac{\partial f_1}{\partial \theta_1}.\frac{\partial \theta_1}{\partial W_1} = X.f_1'.W_2.f_2'.W_3.f_3'.W_4.\frac{\partial O}{\partial \theta_4}$$

$$\frac{\partial O}{\partial W_2} = \frac{\partial O}{\partial \theta_4}.\frac{\partial \theta_4}{\partial f_3}.\frac{\partial f_3}{\partial \theta_3}.\frac{\partial \theta_3}{\partial f_2}.\frac{\partial f_2}{\partial \theta_2}.\frac{\partial \theta_2}{\partial W_2} = f_1.f_2'.W_3.f_3'.W_4.\frac{\partial O}{\partial \theta_4}$$

Vanishing Gradient Problem

- Choice of activation function: ReLU instead of Sigmoid.
- Appropriate initialization of weights.
- Intelligent Back Propagation Learning Algorithm.

NPTEL ONLINE CERTIFICATION COURSES

Thank you