



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

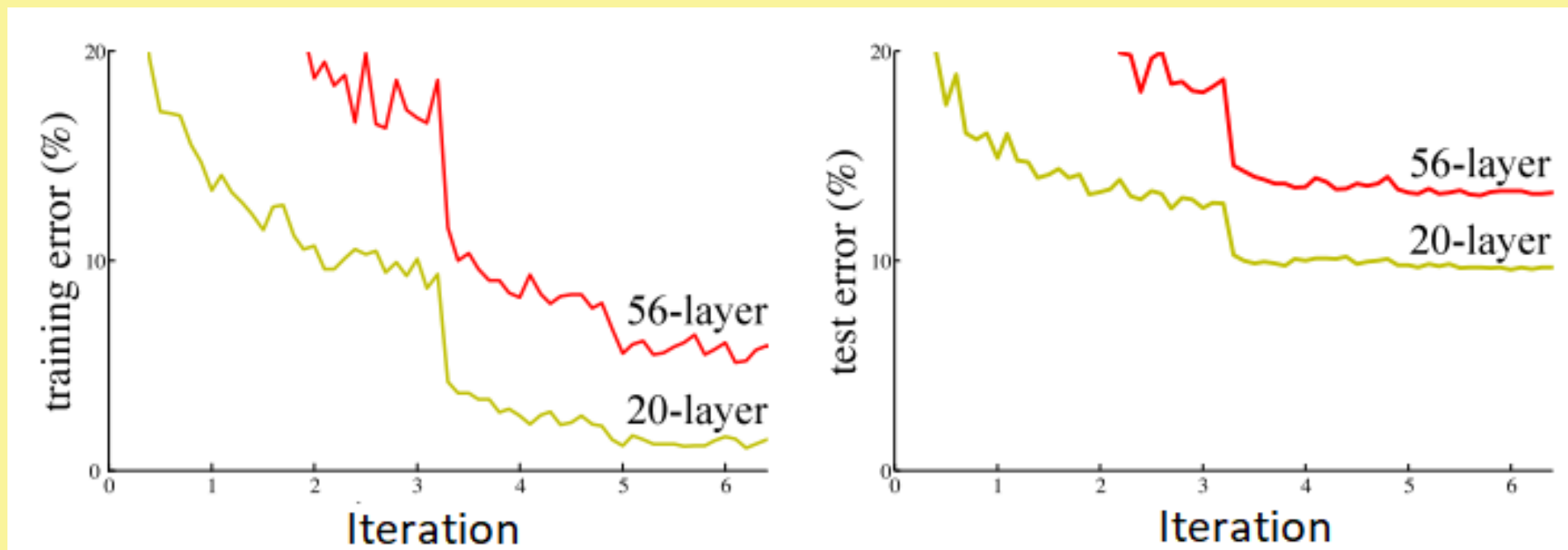
- ❑ Deep learning is data hungry.
- ❑ Overfitting or lack of generalization.
- ❑ Vanishing/Exploding Gradient Problem.
- ❑ Appropriate Learning Rate.
- ❑ Covariate Shift.
- ❑ Effective training.



Vanishing Gradient

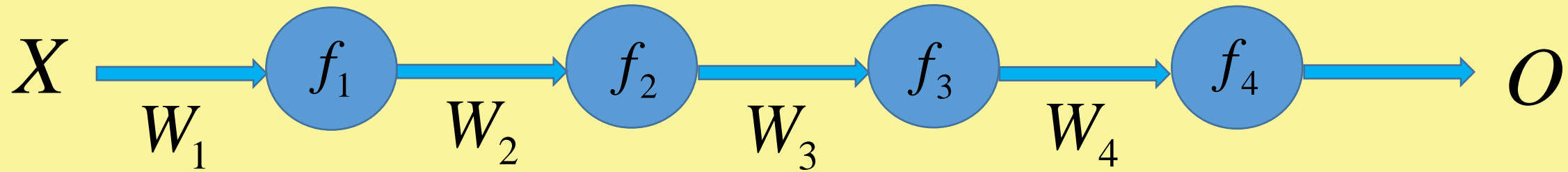


Vanishing Gradient Problem



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

Vanishing Gradient Problem



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



Vanishing Gradient Problem

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

The diagram illustrates the Vanishing Gradient Problem in a deep neural network. The equation $O = f_4(W_4 f_3(W_3 f_2(W_2 f_1(W_1 X))))$ is shown with four colored brackets above it, each labeled with a θ parameter. A green bracket labeled θ_4 spans the entire expression. A blue bracket labeled θ_2 spans $f_2(W_2 f_1(W_1 X))$. A pink bracket labeled θ_1 spans $f_1(W_1 X)$. A red bracket labeled θ_3 spans $f_3(W_3 f_2(W_2 f_1(W_1 X)))$.



Vanishing Gradient Problem

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

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

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



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*

