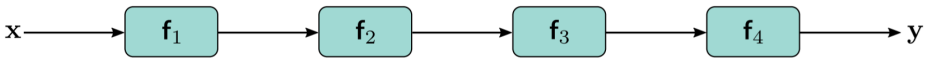
# Residual Networks (Part 1)

Dr. Alireza Aghamohammadi

# Motivation

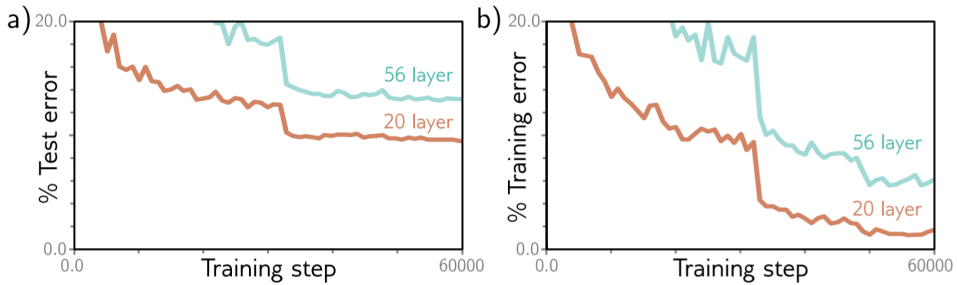
- This lecture presents the rationale behind the use of residual blocks.
- In standard neural networks, the output from each layer is directly fed into the subsequent layer[1].




[1]Reference: "Understanding Deep Learning"

# Problem Statement

- The VGG network, which comprises eighteen layers, exhibits superior performance compared to the eight-layered AlexNet.
- However, as more layers are added beyond this point, the performance in image classification begins to decline.
- This decline is observed in both the training and testing sets, suggesting that the challenge lies in training deeper networks, rather than in the capacity of these deeper networks to generalize[2].




[2]Reference: "Understanding Deep Learning"

# Idea Behind Residual Networks

- Each layer in a network computes an **additive function** for training. However, this leads to an exponential increase in the magnitude of activations during initialization.
- To counteract this, residual blocks incorporate **batch normalization**. This technique adjusts the activations at each layer by re-centering and rescaling them.
- By using residual blocks with batch normalization, **we can train much deeper networks**. These deeper networks enhance performance across a wide range of tasks.