Skip connection:

Directly add input layer to output layer,m that means the output can be divided into input and a nonlinear transform of input

Ad: addressing to vanishing gradient problem, add a 1 an each original gradient, the network is more sensitive to the output; reduce weight degradation by eliminate the symmetry of the network.

Weight degradation: only a few unit can be changed when the input changed, the rank of the weight matrix is not high