What is the vanishing gradient problem in machine learning?

It's a difficulty that arises when using gradient based learning and backpropagation to train artificial neural networks. Gradients are used to update the weights in a network in backpropagation. However, the gradient can sometimes become vanishingly small, thereby preventing the weights from changing their values. As a result, the network stops training since the same values are transmitted again and no relevant work is done. Residual neural networks are used to solve such challenges.

What is a residual neural network?

Residual neural networks, or ResNets, are a type of neural network that uses identity mapping to solve problems. This means that the input to one layer is sent to another layer directly or as a shortcut, essentially the most important thing in a residual neural network is a skip connection. The skip connection basically maps the identity of a previous layer to another, for a visualization look at figure one.

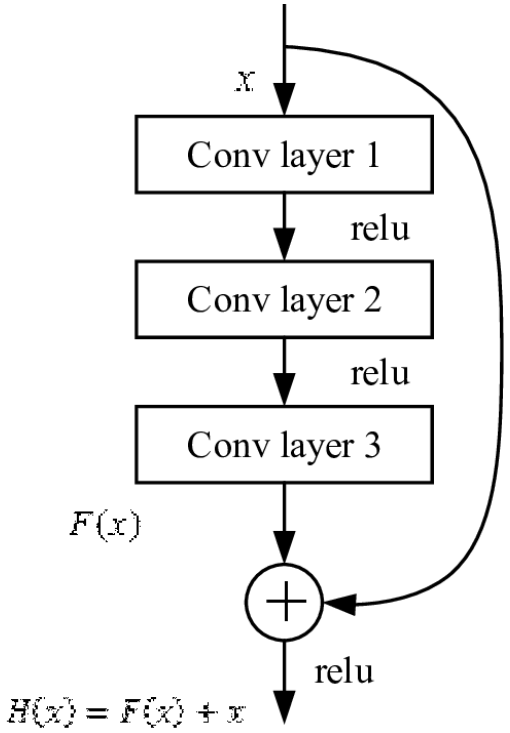

Fig 1; skip connection ResNet