ction is the Fourier transform of the target function.

gradient of the loss function of a DNN is insensitive to the

istribution. The weights are initialized by a Gaussian distribution. The weights are initialized by a Gaussian distrib

e weights is not a good measure of the quality of the network.
The proof is in the appendix.

ster for high-frequency dominant functions, and it is due to the
al guarantees for training deep neural networks
es are combined to form higher level features. The higher level features are then used to build even higher level f