

VANISHING GRADIENT PROBLEM

When the gradients of the loss function with respect to the parameters in the early layers of a network are very small. Causes slow learning and since many gradients are tiny, they don't contribute much to learning and can lead to poor performance

Chris Albon