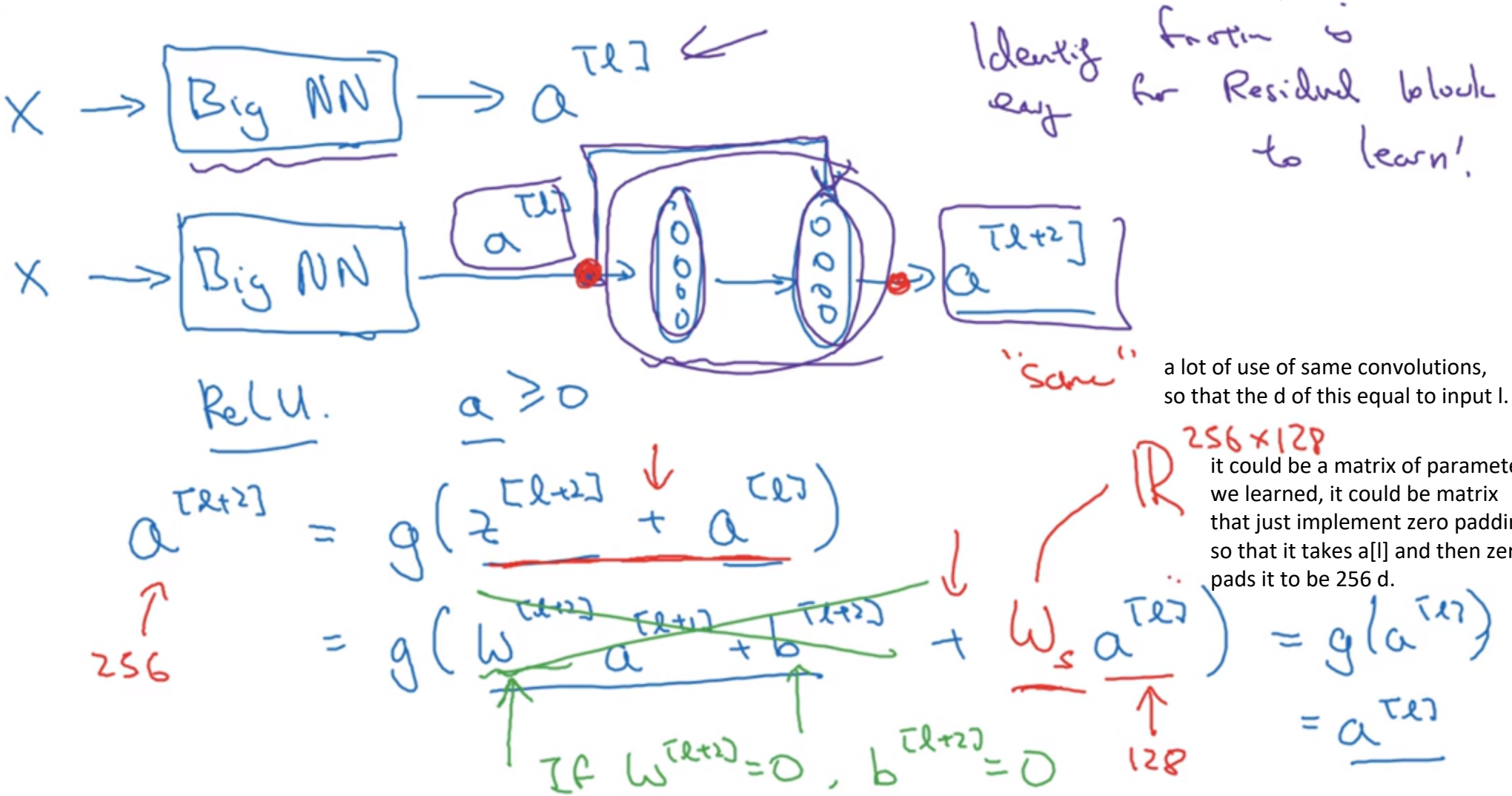
Why do residual networks work?

网易云课堂



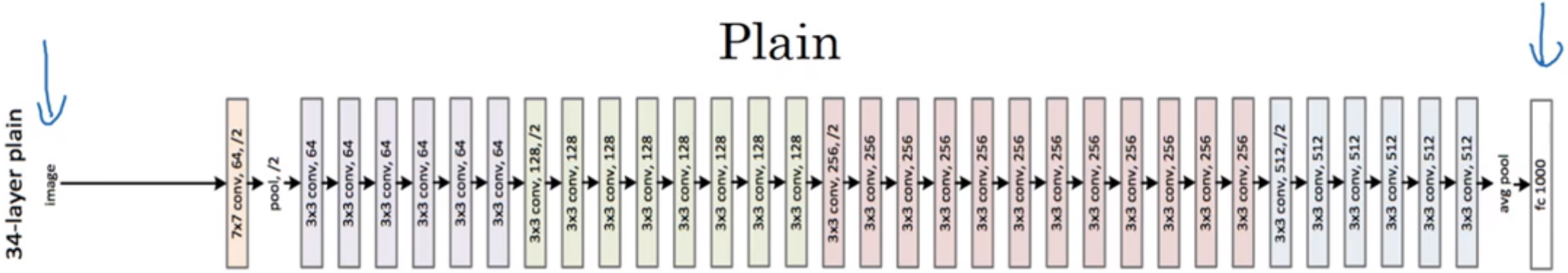
a lot of use of same convolutions, so that the d of this equal to input l.

it could be a matrix of parameters we learned, it could be matrix that just implement zero padding, so that it takes a[l] and then zero pads it to be 256 d.

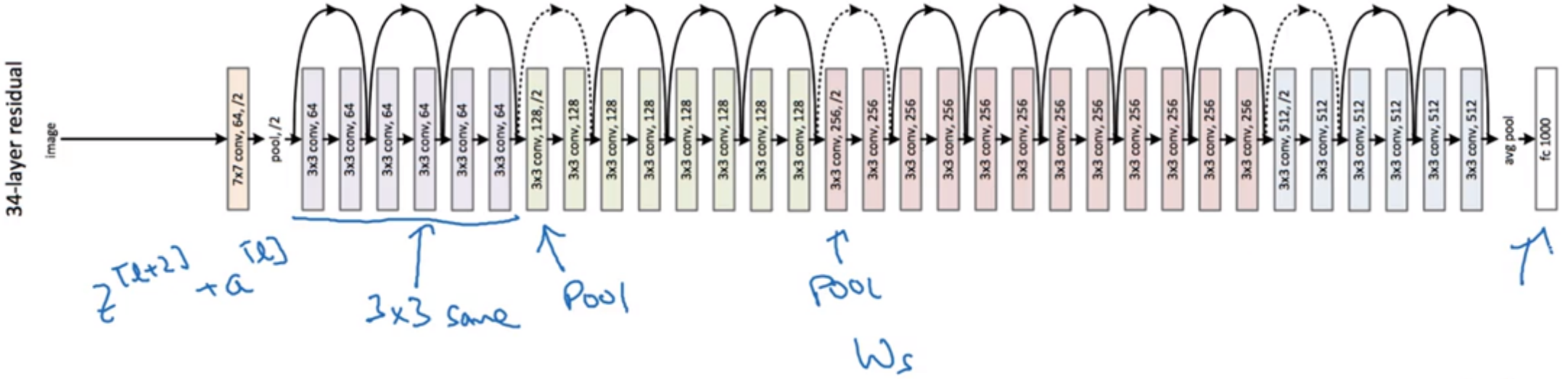
ResNet

网易云课堂

Plain



ResNet



[He et al., 2015. Deep residual networks for image recognition]

Andrew Ng