채널 셀렉션을 한다는게 얼추 이해가 갈 것 같음.

Conv2d(3, 64, 3, 1, 1) 연산을 해주면 filter 개수가 64개 나오고 output feature map의 shape은

(64, 32, 32)가 된다.(dataset – cifar10 기준)

이때 batchNorm layer의 weight의 shape를 보면 (64, )이다.

저 output feature map의 채널들이 예쁜 분포를 갖도록 내부적으로 처리해주나 보다..

그러니까 이 batch norm weight에 L1 regularization 처리를 해줘서 train해서 sparsity 하게 만들어서 채널 셀렉션이 가능하도록 하는듯!