在本篇论文中，采用模型压缩的方式进行轻量化设计，同时采用Dw的方式进行卷积，，主要的创新点在于，通过逐块量化的方式，将部分卷积层进行三值化（即-1，0，1），三值化的方法如下：

1. 首先对全精度的模型进行训练，得到模型之后，对每一层的参数进行归一化
2. 之后对选取的进行归一化的层通过阈值法进行初步的阈值化，之后在通过