可以按照graph boykov\_kolmogorov min s-t cut 提供的方法，先把一个神经网络切割割成两部分，可以在分割之后的网络中继续应用这个算法，然后分别部署在不同的边缘设备上。还有老师说的，最短路径问题，可以在一个设备中部署多个分割后的网络层，用 isinstance判断接下来要经过的卷积层与自己本身要经过的卷积层是否一致