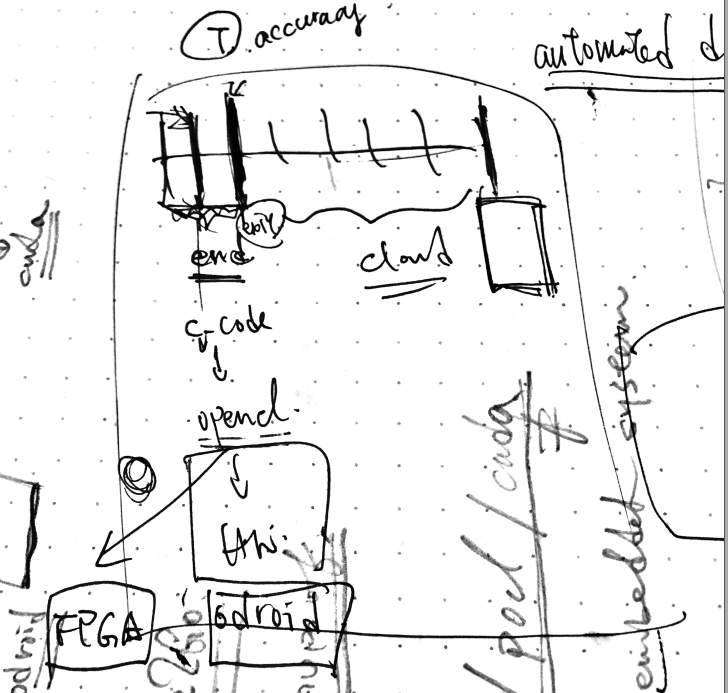
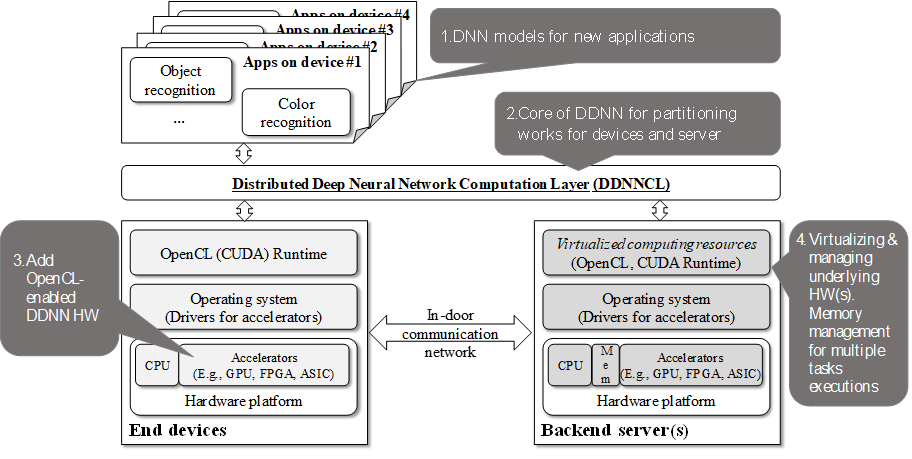
**DDNNCL构成：把ddnn 和brachynet垂直和横向的架构的东西拿出来，理论保留，  
并加上related work来进行延伸**

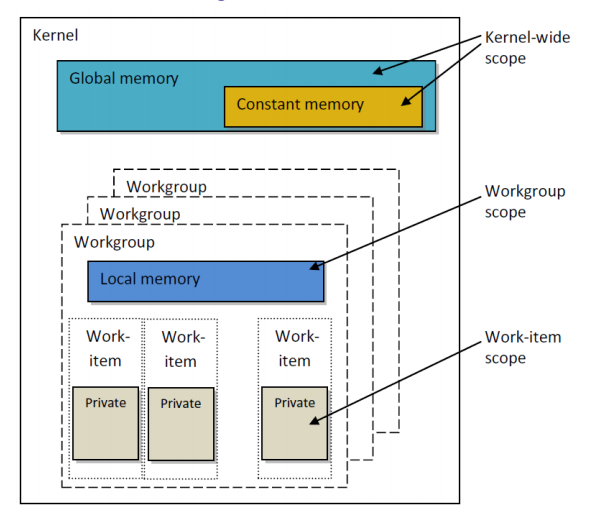


**思考：**

把proposal和目前的结果放上去（ddnncl部分放大，剩下的部分缩小，ddnncl有自动化的机制找好threshold，方便end device中准确率不够的数据动态的往server丢）



1. end device的东西怎么传给server，怎么做比较有效，（也就是说整个model的layers那些在前端，那些在后端，已经train好的参数要怎么给到后面去）
2. loop optimization (compiler generate code)——> loop fusion(一次loop但loop体较大）
3. memory optimization(Opencl)



数据存放使用Global的时候效能可能最差

1. 使用MQTT来传递消息，即从end device向server传递经过convolution之后的message（QOS support）

<https://zhuanlan.zhihu.com/p/20888181>

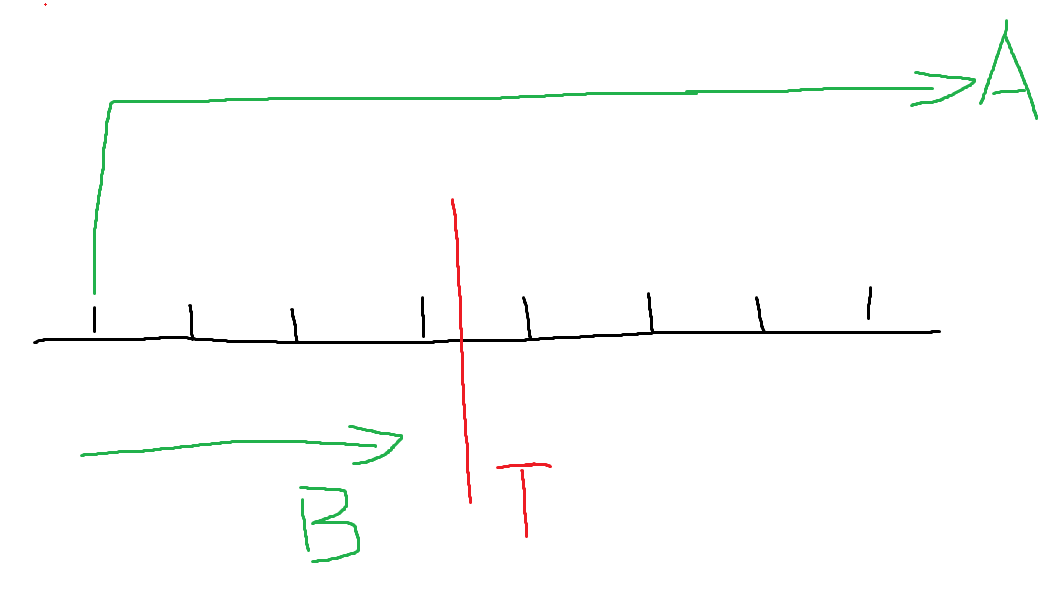
1. Ijij

**发散：**

1.考虑整个model进行inference过程中的power comsumption，来决定哪些work需要交给device，哪些需要给server

2.end device中有一些layer，可以再前面的layer加一层，通过判断此时的准确率来预测当end device所有layer做完之后其准确率是否符合threshold，如果不行就直接丢给后端server去做。

3.end device中convolution做完（数据有降维）之后，直接丢给后端server，然后前后端同时对其进行分析，如果前端准确率满足就直接结束，否则server会返回其计算结果



也就是说A,B同时进行，如果B没有满足threshold，A将计算结果直接给B

4. 目前的可移植性比较低——————跨framework（再说）