一、硬件设计中应该考虑哪些共性问题？

1. （1）核心/关键计算单元的共同特点

（2）位置特性（指不同的方法或网络中，核心计算所处datapath的不同位置），或者理解为data re-use，

以上两点是否能够决定建立的基础核心计算是可靠的？

1. 是否有必要建立异常function单元/额外storage，用于cover不同的ML方法？（使用寄存器存储）

**答：**异常function单元一般指特殊算法的特殊需求计算，例如：残差网络中的cutshort、或者多种的激活单元，建立异常单元需要根据可支持网络决定。

另外，异常storage（网络结构改变时，用于存储中间数据的storage）采用register的方式。采用register的主要原因是，虽然网络的计算无法完全pipeline，实现pixel-by-pixel，但是在一定的时间计算阶段内，计算仍然可以pipeline，因此添加额外storage有助于实现阶段性data流水。

1. 针对可重构、可编程coprocessor，如何简化control-path的复杂度？架构角度考虑，应该采用冯诺依曼架构还是使用通用外设逻辑控制（FSM）方式，或者依赖数据置信度/统计特性等等特点，实现的自适应内部控制？

**答：**目前，他们采用指令集架构的方式。采用指令的方式可以比较好的调控data-path前后关系等。

另外，数据置信度/统计性等控制整个通路的方式目前没有使用。在最后的报告，可微分神经计算中采用的方式就是数据置信度可控的方式。

1. 卷积运算在CNN中的物理意义是什么?特别是在各个卷积层的意义？

**答：**卷积的物理意义，以及在CNN中各个层的意义属于黑盒范畴。

理解这个问题可以从特征构成和特征提取的角度考虑，

特征构成的过程包括：

1. pixel-像素集特征
2. edge-边缘性特征
3. object-part-部分目标特征
4. object-整体目标特征

因此从结构化特征的角度理解，高层次特征由一些基本特征组成。CNN中各层卷积的物理意义应该是通过滤波操作提取不同层次的object特征，从而识别object。

CNN网络相关问题

1. CNN网络的input数据格式要求，以及性能如何？

**答：**CNN的input数据一般以RGB为主，或者单独使用Y也可以。分类的能力这两种格式的分类能力基本差别不大。检测能力不确定，带有颜色的检测有利于目标跟踪和位置判断等。

直接使用RAW数据的分类能力可能会比较差。

B、卷积层N与卷积层N-1互联时，N层的某一个feature map是否应该由N-1所有的feature map 得到？或者从N-1层中选择几个feature map，采用加权的方式计算？

**答：**目前采用的方式是第N层某一个feature map的某一个元素与上一层N-1所有的feature中的对应位置元素卷积后累加的方式，没有采用加权的方式。

1. pooling中max和mean的方式选择依据？

**答：**max和mean的选择属于经验性、实验性结论。

D、卷积核大小的选择依据是什么？

**答：**卷积核的大小依赖于长期训练和实验的结果。

E、如何解决classify和detect对于位置敏感的需求冲突？

**答：**在CNN中网络本身具有平移不变性，主要是由于pooling层的作用。因此在平移不会对classify产生影响。

1. 有哪些技术可以解决或缓解CNN的数据复用问题？

**答：**数据复用包括两部分，参数复用和input data复用。

参数复用的方式是，参数全部存储在DDR中，load一组参数竟可能完成当前帧所需的全部计算。

Input data复用方式使用SRAM，存储需要重复使用的data。

1. 如何应对大网络（parameters较大，eg.VGG）所需的超大weights对数据带宽的需求。

**答：1、**对于weight过大的问题，所有的网络全部采用将parameter放在DDR中的方式，根据网络实际运行的阶段，从DDR中load相关参数。

**2、**采用权重量化的方式，可以量化为16bit、8bit甚至二值化的定点数。

1. 进入CNN网络的数据需要经过什么样的操作？具体过程？

**答：**在使用框架时，需要将数据转换为对应的格式，对数据本身的大小需要做scaler等。