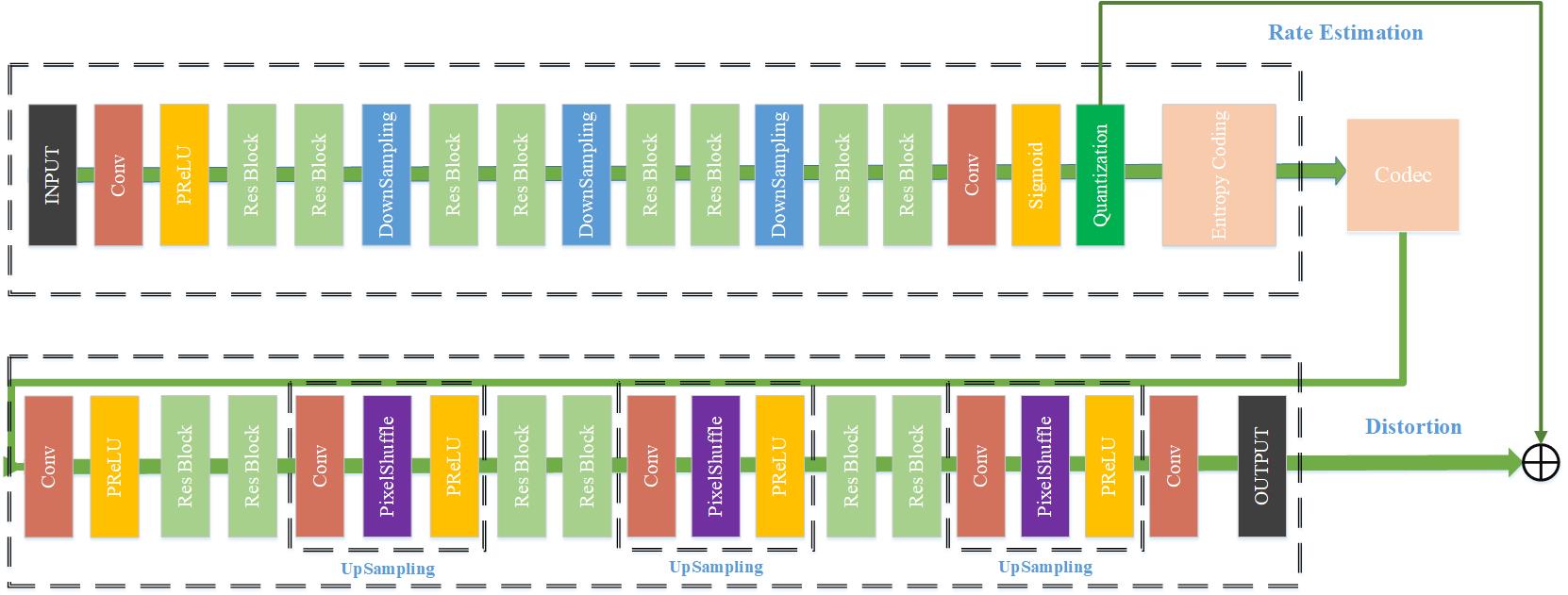
我们目前的Coded Vision框架主要分为两个部分，一是组里另两位大佬原创的deep coder框架，另一个是我现在正在着手做的coded vision的vision task engine部分。当前的状态是在coding和classification的融合上有了一些初步的结果，但精度还不是很高。我们计划先以分类为核心，辐射其他基于CNN的CV应用，拓宽该框架的适用领域，而后再针对有应用潜力的编码/视觉融合任务优化算法实现和精度。

1. **Deep Coder**

Deep Coder的工作旨在将传统的压缩编码方法整合到deep learning的框架中，使用学习的方法来消除图像/视频中的冗余信息。该方法总体上采用了Auto encoder的网络结构，将encoder最后一层的得到的feature maps(fMaps)作为压缩后的图像信息，在对该fMap进行量化和熵编码（算术编码）后，将其传输至decoder端；随后decoder端进行算数解码，将二进制码流恢复成fMap；Auto encoder的decoder网络则负责将fMap还原成输入图像。如下图所示，上半部分为encoder，下半部分为decoder。



在网络实现的上，deep coder主要采用了残差网络（residual block）作为网络的基本结构；考虑到要在压缩编码的同时，兼顾图像的恢复质量，网络仅对图像进行了3次pooling，且该pooling是通过stride=2的卷积层实现的，这样可以一定程度上避免max pooling/avgpooling带来的直接信息损失。

从整体上来看，deep coder的框架与传统编码最大的差别在于，传统编码通过一系列的数学变换与量化编码，将图像信息压缩为不具备可读性的码流，而deep coder则是采用提取图像特征的方式进行压缩，在简单的算数解码后，码流就具备很高的可读性与一定的可视化能力，这为我们后续的coded vison工作奠定了重要的基础。

1. **Vision Task Engine**

Coded Vision的vision task engine部分旨在从deep coder的码流出发，直接进行相关的视觉任务，跳过传统方法所必须的解码部分，直接提取得到图像的高层信息。进一步的，该工作致力于如何渐进地、一体化地提取图像不同层次的特征，以高效的适应于不同的图像处理与视觉应用。

在先前的文章中，我们采用了固定deep coder，在encoder后接分类网络，并且只训练分类网络的方法对编码和分类应用做了初步的融合。在该方法中，编码性能已被优化到相对较高的水平，并固定不变。此基础上，我们进一步单独优化了分类性能，相关的结果可见PCM的论文。在选用的分类网络上，我们在后期选用了预训练过并且微调过网络结构的Resnet18作为分类网络，可以节省分类网络的训练时间，其分能性能也更有保证。总体来说，在固定的encoder后接classifier网络是相对比较低效的，encoder所需的feature和classifier所需的浅层feature较为相似，但仍有相当的差距。

我们后续所采用的是joint training的训练框架来整合编码与分类应用（见下图），在这个框架中，网络由分类loss、重建loss和码率控制三者进行联合优化，而非单独的优化某一任务。在联合优化实现上，重建质量和分类性能的优先级更高，因此采取了先行联合优化performance loss 和 distortion，待两者接近收敛后，再加入rate control的部分，从而实现三者的联合优化

