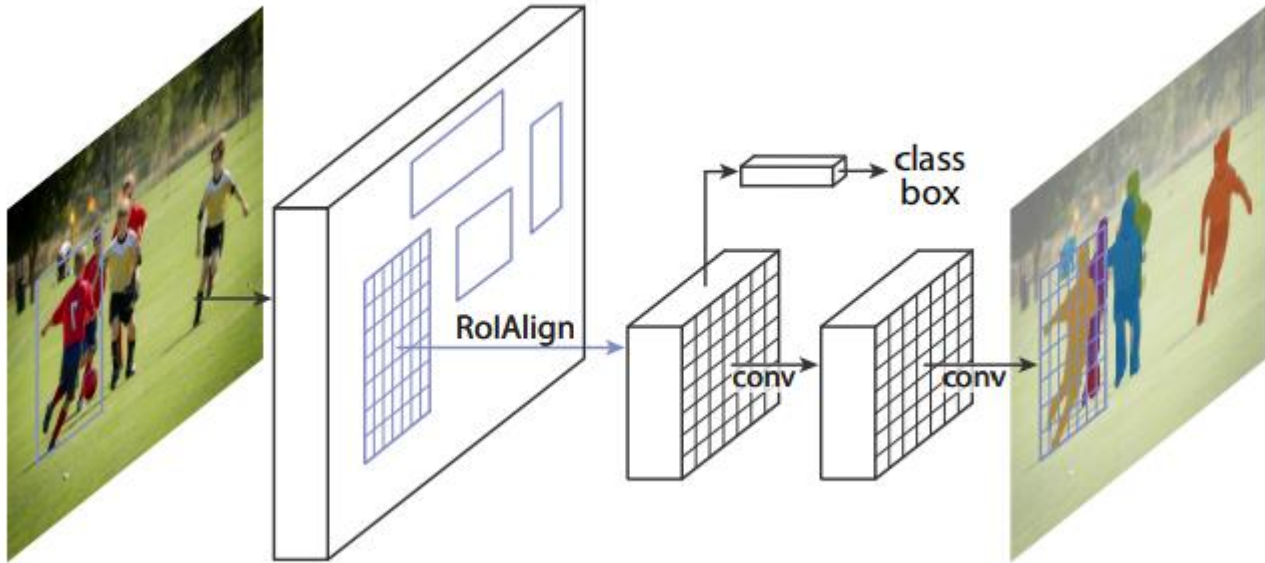
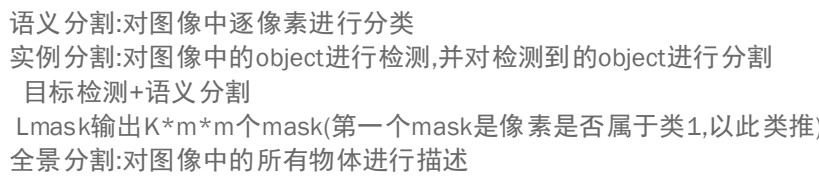
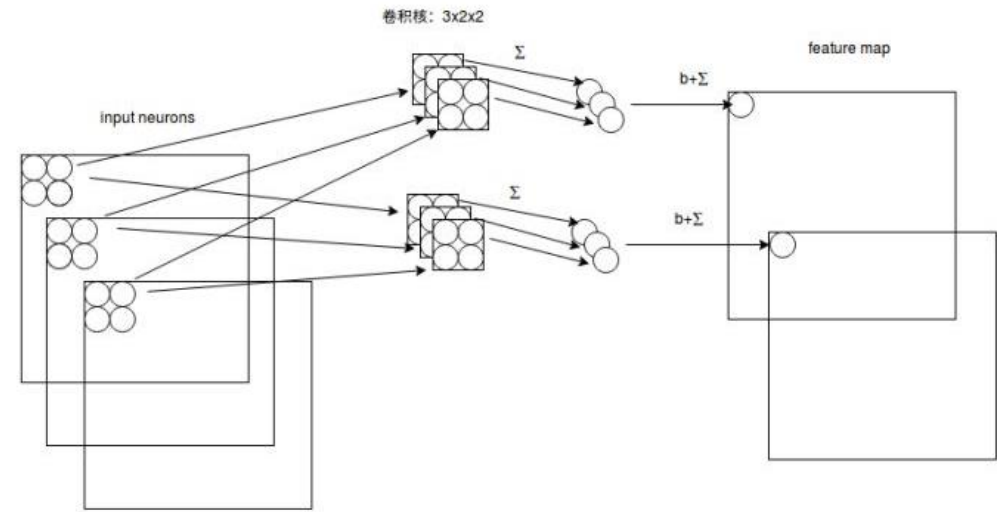


The diagram illustrates a multi-head attention block. It consists of three parallel processing paths. Each path takes an input tensor of size $K \times K \times S \times P \times P$ and applies a $\text{conv} + \text{relu}(K \times S \times S \times P \times P)$ operation. The resulting tensor size is $m \times n$, where $m = \frac{1}{k} \times ((M+2p) \times s)$ and $n = \frac{1}{k} \times ((N+2p) \times s)$. Each path then applies a RELU activation. The outputs of the three parallel paths are combined to produce the final output tensor of size $m \times n$.



Mask RCNN是Faster RCNN的扩展
Faster RCNN的每个Proposal Box都要使用FCN进行语义分割,分割任务与定位,分类任务是同时进行的