Self-Attention

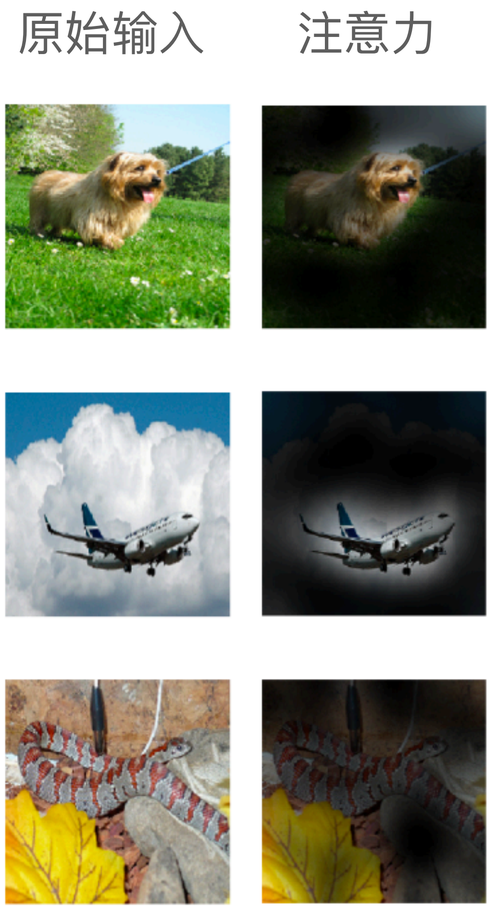

图 4　注意力可视化

- 注意力可视化结果显示，模型倾向于将注意力集中于原始输入图像中、与分类信息相关性较高的区域

MLP Head

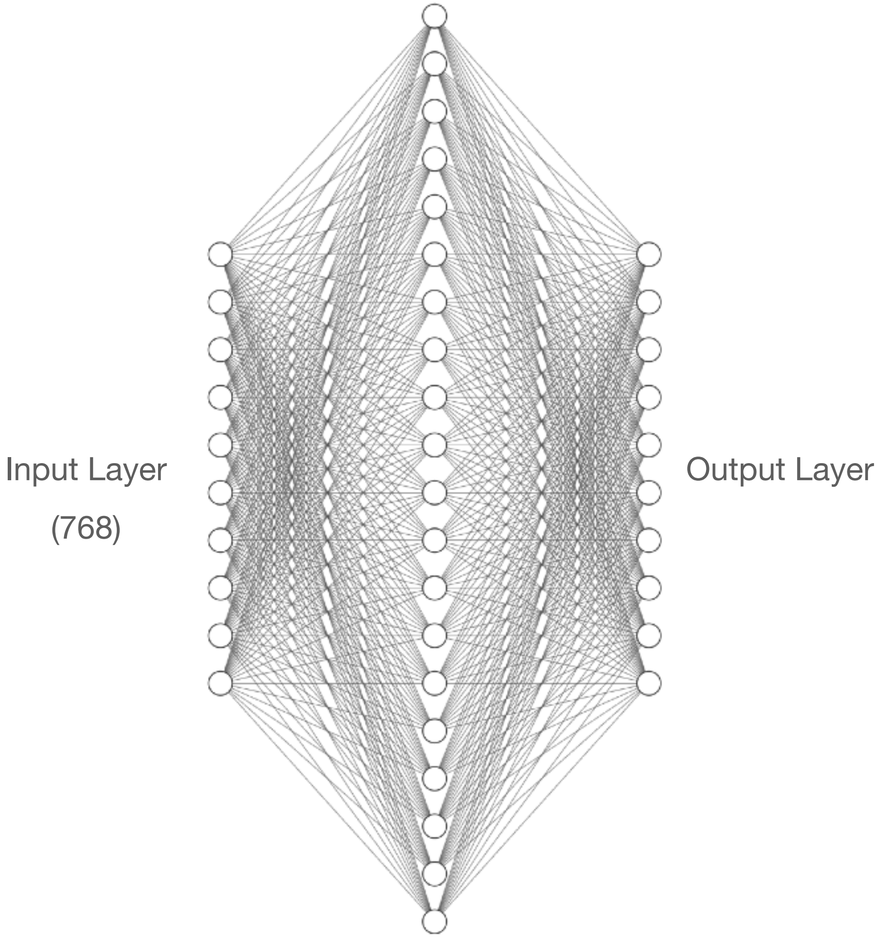

图 5　MLP Head结构

- MLP Head 将 Transformer Encoder 输出中 token 为 [class] 的向量作为图像整体特征信息作为输入，经过一层隐藏层，最终输出分类判断