## Transformer

### Architecture：

1.残差网络，主要目的是通过增加网络层数，提高准确率，避免出现网络退化问题

图示

描述已自动生成

2.通过stride 2的conv层，进行降采样

3.网络结构

图示

描述已自动生成

图示, 箱线图

描述已自动生成

4.模型拓展调整

表格

描述已自动生成

### Train&Evaluation：

### Skill&Knowledge：

1. 网络变深会容易产生梯度消失/爆炸，通过NL解决
2. 盲目增加网络深度，并不一定能够提升准确率

### Questions:

### Improve&Validation：

1. 没有使用dropout，可以用一下对比
2. 使用SGD，目前的优化器已经迭代了很多版本，尝试换一下

## Extension