

# MTSVRC 2018 评测方案

## 算法误差评测:

本竞赛使用 Top-1 分类方法评估算法误差，具体测评方法如下：

对于每个测试视频，算法将按最高分类置信度输出预测标签  $l$ ，视频的真实标签为  $g$ 。则此视频的算法误差为：

$$e=d(l,g)$$

若  $x=y$ ，则  $d(x,y)=0$ 。若  $x \neq y$ ，则  $d(x,y)=1$ 。算法评估的整体误差为测试数据集中所有视频的平均误差。

## 算法运行时间评测:

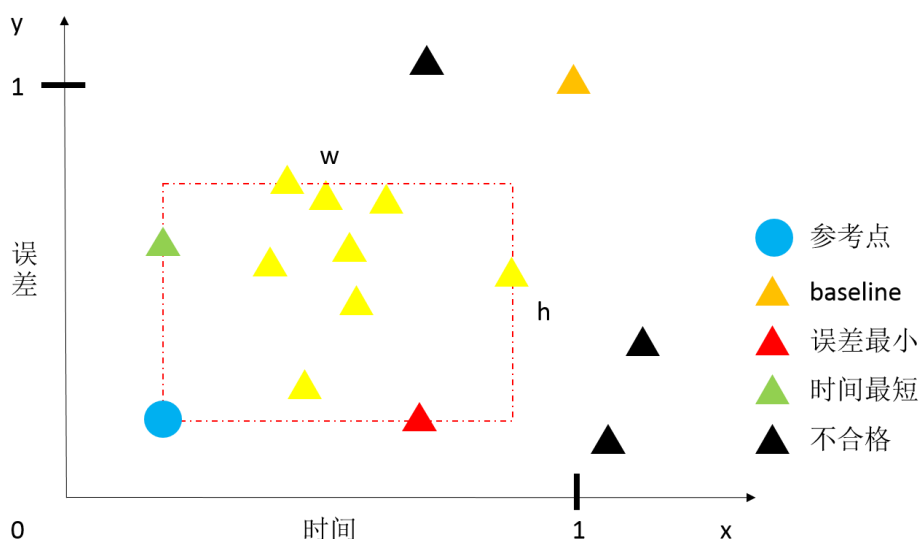
参赛队伍程序的输入数据为视频文件，输出为该视频的类别。单个视频运行时间包括所有预处理和推理时间，即为单个视频从视频输入到结果输出所需整体时间。最终度量值为  $M$  次处理测试集里  $N$  个视频的平均值  $T_{average}$ ，计算方式如下：

$$T_m = \frac{\sum_{n=1}^N T_n}{N}$$
$$T_{average} = \frac{\sum_{m=1}^M T_m}{M}$$

其中， $T_n$  为算法处理第  $n$  个视频花费的时间， $N$  为测试视频个数， $M$  为测试次数。

注意：竞赛组织方会首先测评参赛队伍的算法运行时间。若运行时间超出 baseline 指标，则视为本次提交无效。

## 算法综合测评:



最终冠军队伍将综合算法误差和算法实时性的结果确定，具体测评方法如下：

如上图所示，坐标系的  $x$  坐标由运行时间基准值归一化， $y$  坐标由误差基准值归一化，坐标值为(1,1)的橙色三角形是组织方公布的 baseline。若参赛团队提交了低于 baseline 的结果（黑色三角形）则不参与测评。同时优于误差和运行时间基准值的结果（红黄绿三角形）参与测评。

组织方根据各团队提交的结果确认一个算法评测参考点（蓝色圆形）。算法评测参考点的 x 坐标由最短时间结果（绿色三角形）确定，y 坐标由误差最小结果（红色三角形）确定。

为综合评估误差和时间的关系，所有参赛队伍提交的结果将被规范在红色虚线矩形中。组织方将综合时间和误差的关系权重，对提交结果与参考点计算加权距离：

$$D_i = \sqrt{(x_i - x_k)^2 + \alpha(y_i - y_k)^2}$$

其中，加权因子 $\alpha$ 为红色虚线矩形的宽高比平方，即 $\alpha = \left(\frac{w}{h}\right)^2$ 。

$D_i$ 为第 i 个结果的加权距离， $x_i$ 和 $y_i$ 分别为第 i 个结果的 x 和 y 值， $x_k$ 和 $y_k$ 分别为参考点的 x 和 y 值。加权距离越短，准确性和运行速度的综合效果越好。

随后组织方会根据距离排序确定参赛队伍名次。提交最短距离结果的参赛队伍为本次竞赛的冠军。

注意：为突出算法实时性，若多个提交结果距离相同，算法运行时间更短的队伍获胜。