

面向选手的说明（新）

一、发布文件说明

1. 发布包含 3 个文件或文件夹：

- (1) 训练集
- (2) 测试集
- (3) 面向选手的说明

2. 说明

训练集使用方法请见训练集中的说明文档；测试集和样例说明请见下文。

3. 本次竞赛主要分为两部分，一部分是对不定点场景中的图片进行一个空气质量等级评估，分别分为优，良，轻度污染，中度污染，重度污染，严重污染 6 个等级；另一部分，是对定点场景的图片进行 pm2.5 值的预测，单位 $\mu\text{g}/\text{m}^3$ ，这部分可以借助空气湿度辅助判断。

二、输出格式，样例及评分说明

1. 测试规模与参数说明

(1) 测试集包含若干个用于提交测试结果的'.jpg'文件，定点测试集中还包含一个湿度说明文档。最后请选手将测试结果提交为一个'submission.txt'文件（后文简称 a.txt），它将与我们作为答案文件的'Groundtruth.txt'文件（后文简称 b.txt）进行对比判定分数。

(2) a.txt 应该含有若干行，它们以'a_'开头，这些表示“空气质量等级”的测试结果，每行包括两部分，以','分隔，前边表示图片 id，即测试图片名称（包含后缀名，例如图片叫 a_1.jpg，那么对应的图片 id 是 a_1.jpg），后边表示测试结果，测试结果 result 须满足 $\text{result} \in \{0, 1, 2, 3, 4, 5\}$ ，它们分别与空气质量等级中的“优，良，轻度污染，中度污染，重度污染，严重污染”一一对应。

(3) a.txt 还应该含有另外若干行，它们以'b_'开头，这些表示“pm2.5 值”的测试结果，每行包括两部分，以','分隔，前边表示图片 id，即测试图片名称（包含后缀名，例如图片叫 b_1.jpg，那么对应的图片 id 是 b_1.jpg），后边表示测试结果，测试结果 result 须满足 $\{\text{result} | 0 \leq \text{result} \leq 1000, \text{result} \in \mathbb{R}\}$ ，它表示 pm2.5 的具体数值，单位 $\mu\text{g}/\text{m}^3$ 。这部分最终测试时可以使用湿度辅助判断，湿度信息包含于测试集中的说明文档中。

2. 评分细则

(1) 提交文件中不可包含中文字符

(2) 若出现重复的图片 id，则这一个测试点记 0 分

(3) 若 b.txt 的某个图片 id 在 a.txt 中没有找到对应的图片 id，则这一个测试点记 0 分

(4) 在图片 id 符合要求的前提下，采用以下方式评分

[1] 空气质量等级结果评测：将相差的空气质量等级命名为差异度 (GOD)，差异度的计算公式如下：

$$\text{GOD}(i) = |p_i - t_i| \quad (2-1)$$

i 表示第 i 个结果， p_i 表示第 i 张图片的预测结果， t_i 表示第 i 张图片的金标准， $p_i, t_i \in \{0, 1, 2, 3, 4, 5\}$ 。若 $\text{GOD}_i=0$ 得 2 分， $\text{GOD}_i=1$ 得 1 分，其它情况得 0 分，特别注意， p_i 超出范围不得分

[2] pm2.5 值的评测：采用平均绝对误差 (MAE)，假设预测结果数值为 A，金标准为 B，其中 $\{(A, B) | 0 \leq A, B \leq 1000, A \in \mathbb{R}, B \in \mathbb{R}\}$ 。那么误差 E 符合如下公式：

$$E = |A - B| \quad (2-2)$$

考虑到实际需要，我们希望误差不要大于 10, 所以 $E > 10$ ，计 0 分， $10 \geq E > 1$ ，计 1 分， $E \leq 1$ ，计 3 分，例如 $E=1.01$ 计 1 分， $E=0.99$ 计 3 分，特别注意，A 超出范围不得分

[3]对两个项目分别进行评测排名，每个项目满分 100 分，采用加权平均，对于第一个项目空气质量等级评测，其公式如下：

$$\text{score1} = \frac{\sum_{i=1}^{N1} s_i}{2 \times N1} \quad (2-3)$$

其中 $N1$ 是第一个项目的测试图片数量， s_i 是每张图片评测的单次得分。

对于第二个项目 $\text{pm}_{2.5}$ 值评测，其公式如下：

$$\text{score2} = \frac{\sum_{j=1}^{N2} s_j}{3 \times N2} \quad (2-4)$$

其中 $N2$ 是第二个项目的测试图片数量， s_j 是每张图片评测的单次得分

[4]最后以总成绩为准，总成绩为 $(\text{score1} + \text{score2}) / 2$

3. 测试样例及说明

金标准样例：

a_2018_1_1.jpg, 0
a_2018_1_2.jpg, 1
a_2018_1_3.jpg, 2
a_2018_1_4.jpg, 3
a_2018_1_5.jpg, 4
a_2018_1_6.jpg, 5
b_2018_2_1.jpg, 30
b_2018_2_2.jpg, 40
b_2018_2_3.jpg, 50
b_2018_2_4.jpg, 60
b_2018_2_5.jpg, 70
b_2018_2_6.jpg, 80

提交文件样例：

b_2018_2_6.jpg, 79
a_2018_1_1.jpg, 0
a_2018_1_2.jpg, dsd
sdf, 3
b_2018_2_6.jpg, 20
a_2018_1_4.jpg
a_2018_1_5.jpg, 3
a_2018_1_6.jpg, 6
b_2018_2_1.jpg, 41
b_2018_2_4.jpg, 62.94
b_2018_2_3.jpg, 49.32
b_2018_2_2.jpg, 41.01
b_2018_2_5.jpg, 73.05
b_2018_2_6.jpg, 71.03

以下为说明：

这组样例的得分细节如下表：

图片 id	得分	说明
b_2018_2_6. jpg	0	id 重复出现不得分
a_2018_1_1. jpg	2	正常结果
a_2018_1_2. jpg	0	结果不在范围内
sdf	0	不在金标准内的图片 id
a_2018_1_4. jpg	0	格式无效，未用','分隔
a_2018_1_5. jpg	1	正常结果
a_2018_1_6. jpg	0	结果不在范围内
b_2018_2_1. jpg	0	正常结果
b_2018_2_4. jpg	2	正常结果
b_2018_2_3. jpg	3	正常结果
b_2018_2_2. jpg	0	格式无效，未用','分隔
b_2018_2_5. jpg	1	正常结果

所以 a 类样例得分为 $(2+1)/(6*2)=25.00$ ， b 类样例得分为 $(2+3+1)/(6*3)=33.33$ ， 总分为 $(25.00+33.33)/2=29.17$

4. 其它说明

- (1) 图片 id 的输出顺序不做规定，但需要符合输出规范
- (2) 支持提交在 Linux 或 Windows 下编辑的 '.txt' 文档
- (3) 最终提交的结果应与“介绍—>数据说明与下载—>点击下载初赛阶段**测试集**”的测试结果相对应，样例中的数据不会出现在最终的测试结果中
- (4) 由于之前的数据存在一些问题，现对数据集和测试集进行了修改，请选手们重新下载，评分规则不变，对带给选手们的不便表示由衷的歉意！
- (5) 由于测试集规模较小，我们决定将选手的提交次数修改为上限每天 2 次。
- (6) 如有疑问，欢迎与主办方联系！