



**BRICS**  
Business Council



# 2023

## 金砖国家职业技能大赛 (金砖国家未来技能挑战赛)

### 省级选拔赛样题 (第二套)

BRICS-FS-26\_人工智能计算机视觉应用

2023 年 5 月

目录

1、 参赛形式.....3

2、 竞赛内容.....3

3、 项目模块.....3

    3.1 项目模块 .....3

    3.2 任务内容 .....3

模块一：人工智能环境搭建（15 分） .....4

    第 1 题：图像处理开发环境搭建 .....4

    第 2 题：深度学习环境搭建 .....5

模块二：数据处理与分析（20 分） .....6

    第 3 题：图像数据清洗 .....6

    第 4 题：图像预处理 .....8

模块三：机器学习（30 分） .....9

    第 5 题：视频网站用户活跃量预测 .....9

    第 6 题：视频网站每日广告收益预测 .....11

模块四：计算机视觉应用（35 分） .....12

    第 7 题：图像分类之猫狗分类 .....12

    第 8 题：图像分割之细胞分割 .....14

## 1、参赛形式

个人赛（1 人）。

## 2、竞赛内容

竞赛内容由 4 个模块组成，模块之间相互独立，模块错误不传递。

竞赛内容包含以下任务模块：

- 1) 人工智能环境搭建
- 2) 数据处理与分析
- 3) 机器学习
- 4) 计算机视觉应用

如果参赛选手不遵守职业健康安全环境要求，或使自己和其他选手面临危险，他们可能会被取消比赛资格。

参赛者完成竞赛后，由裁判组组织对选手提交结果进行评分。

## 3、项目模块

### 3.1 项目模块

人工智能计算机视觉应用赛项共 4 个模块，要求选手在 4 个小时内完成。具体项目模块名称和时间要求参照下表

模块	模块名称	权重	竞赛时长
模块一	人工智能环境搭建	15%	4h
模块二	数据处理与分析	20%	
模块三	机器学习	30%	
模块四	计算机视觉应用	35%	

### 3.2 任务内容

## 模块一：人工智能环境搭建（15 分）

### 第 1 题：图像处理开发环境搭建

#### 【功能说明】

在 Ubuntu 虚拟机中搭建 Python 所需基础开发环境，虚拟机中已安装 python 环境和包管理工具 pip，相关软件及依赖库的安装包放在虚拟机“/opt”目录下。

#### 【任务要求】

本环节需要在 Ubuntu 系统中安装 Python 基础开发环境，安装完成后需根据答题区要求截图，并将截图保存到答题区对应位置。环境搭建具体要求如下：

1、搭建 OpenCV 开发环境，根据提供的安装包安装 Python 图像处理第三方库 OpenCV，安装完成后在终端模拟器 Terminal 中验证并显示 OpenCV 版本信息。

2、搭建 Pillow 开发环境，根据提供的安装包安装 Python 图像处理第三方库 Pillow，安装完成后在终端模拟器 Terminal 中验证并显示 Pillow 版本信息。

3、搭建 scikit-Image 开发环境，根据提供的安装包安装 Python 图像处理第三方库 scikit-Image，安装完成后在终端模拟器 Terminal 中验证并显示 scikit-Image 版本信息。

4、搭建 SciPy 开发环境，根据提供的安装包安装 Python 图像处理第三方库 SciPy，安装完成后在终端模拟器 Terminal 中验证并显示

SciPy 版本信息。

### 【操作说明】

1、所有实操均在平台的虚拟机（Ubuntu 系统）中进行，虚拟机内为断网环境。

2、本环节需要用到的安装包及依赖库都存放在虚拟机“/opt”目录下，自行判断依赖库完整顺序完成环境搭建。

3、完成任务后，根据平台“答题区”要求，完成答题报告并点击“保存”按钮实时保存。

## 第 2 题：深度学习环境搭建

### 【功能说明】

在虚拟机中搭建常用的深度学习主流框架：TensorFlow、Keras、PyTorch，安装过程中所需的安装包与依赖库都存放在虚拟机“/opt”目录下，根据安装过程中的提示信息完成深度学习环境搭建。

### 【任务要求】

1、搭建 TensorFlow 深度学习框架，通过提供的安装包与依赖库，完成 TensorFlow 环境搭建，在 Terminal 终端模拟器中验证并显示 TensorFlow 版本信息。

2、搭建 Keras 深度学习框架，通过提供的安装包与依赖库，完成 Keras 环境搭建，在 Terminal 终端模拟器中验证并显示 Keras 版本信息。

2、搭建 PyTorch 深度学习框架，通过提供的安装包与依赖库，完成 PyTorch 环境搭建，在 Terminal 终端模拟器中验证并显示 PyTorch 版本信息。

### 【操作说明】

- 1、所有实操均在平台的虚拟机（Ubuntu 系统）中进行，虚拟机内为断网环境。
- 2、所需的安装包和依赖包存放在虚拟机“/opt”目录下。
- 3、完成任务后，根据平台“答题区”要求，完成答题报告并点击“保存”按钮实时保存。

## 模块二：数据处理与分析（20 分）

### 第 3 题：图像数据清洗

#### 【功能说明】

数据的质量对于深度学习算法模型的效果至关重要。通常在对采集到的大量数据进行标注前需要做一些数据清洗工作。现有一个图片数据集，其中包含了一些已损坏图片，请你根据任务要求统计出数据信息并去除这些损坏图片。

#### 【任务要求】

数据中混有已损坏图片、模糊图像，请你根据每个函数的功能描述编写“data\_cleaning.py”文件中 get\_data\_info() 和

cleaning\_bad() 两个函数中的内容。需要实现的功能如下：

- 1、编写 get\_data\_info() 函数，实现读取数据功能，统计数据集中模糊图片、损坏图片的数量、总数量，绘制柱状图展示出来，图中必须有柱子的标签以及高度值，其它参数可根据自己的喜好进行指定；
- 2、将图表通过 Django 渲染展示。
- 3、编写 cleaning\_bad() 函数用来去除已损坏图片，并将已损坏的图像保存到桌面“/提交文件/02\_数据处理与分析/03”目录。



图 1：模糊图片示意图

### 【操作说明】

- 1、所有实操均在平台的虚拟机（Ubuntu 系统）中进行，虚拟机内为断网环境。
- 2、从虚拟机桌面“赛题/02\_数据处理与分析/03\_图像数据清洗”路径下获取相关资料，结果保存至虚拟机桌面“提交文档/02\_数据处理与分析/03 ”文件夹中。
- 3、完成任务后，根据平台“答题区”要求，完成答题报告并点击“保存”按钮实时保存。

## 第 4 题：图像预处理

### 【功能说明】

深度神经网络的效果通常需要有充足的样本数量，一般来说样本数量越多，训练出来的模型效果越好，模型的泛化能力越强。但实际情况中，往往存在样本数量不足或者样本质量不好的情况，通过进行数据增强操作，增加我训练模型的数据量，提高模型泛化能力；同时在增强过程中为数据增加噪声，提升模型的鲁棒性。

使用 tensorflow 库中 keras 包的图像处理 API，将提供的数据集进行数据增强操作。

### 【任务要求】

编写 data\_preprocessing.py 文件，实现对图片数据集进行数据增强操作，图像数据存放在虚拟机桌面“赛题/02\_数据处理与可视化/04\_图像数据预处理”目录下，具体要求如下：

1、对图片进行归一化操作，保存为 JPG 格式，并在图片原名的基础上添加前缀“norm\_”。

2、将图片随机旋转 45 度，保存为 JPG 格式，并在图片原名的基础上添加前缀“rotate\_”。

3、将图片水平平移，平移幅度设置为 0.2，保存为 JPG 格式，并在图片原名的基础上添加前缀“hor\_”。

4、将图片进行垂直平移，平移幅度为 0.2，保存为 JPG 格式，并在图片原名的基础上添加前缀“ver\_”。

5、将图片进行随机水平翻转和垂直翻转，保存为 JPG 格式，并



在图片原名的基础上添加前缀“flip\_”。

6、图片保存到虚拟机桌面“/提交文件/02\_数据处理与分析/04”目录。

### 【操作说明】

1、所有实操均在平台的虚拟机（Ubuntu 系统）中进行，虚拟机内为断网环境。

2、从虚拟机桌面“赛题/02 数据处理/04\_图像预处理”路径下获取相关资料，结果保存至桌面“提交文档/02\_数据处理与分析/04 ”文件夹中。

## 模块三：机器学习（30 分）

### 第 5 题：视频网站用户活跃量预测

#### 【功能说明】

使用某视频网站近 2 年的用户行为日志数据，选择合适的机器学习算法，对该网站的用户活跃量进行预测。

#### 【任务要求】

在 machine\_learning\_iris.py 文件中实现用户活跃量预测功能，其中 load\_data() 函数用于加载数据集，具体要求如下：

1. 从数据库中读取用户日志数据；
2. 查看数据内容，对数据进行必要的清洗和处理，将处理后的数据保存为 CSV 数据；

3. 根据任务要求使用 Pandas 读取 CSV 数据进行特征工程；
4. 对数据集进行划分，训练集和测试集划分比例设置为 7:3；
5. 根据任务要求，选择合适的机器学习算法，构建机器学习模型；
6. 编写模型训练相关代码，完成模型训练；
7. 使用可视化库对测试数据的预测结果和真实结果进行可视化对比，并使用 Django 在前端页面中渲染展示；

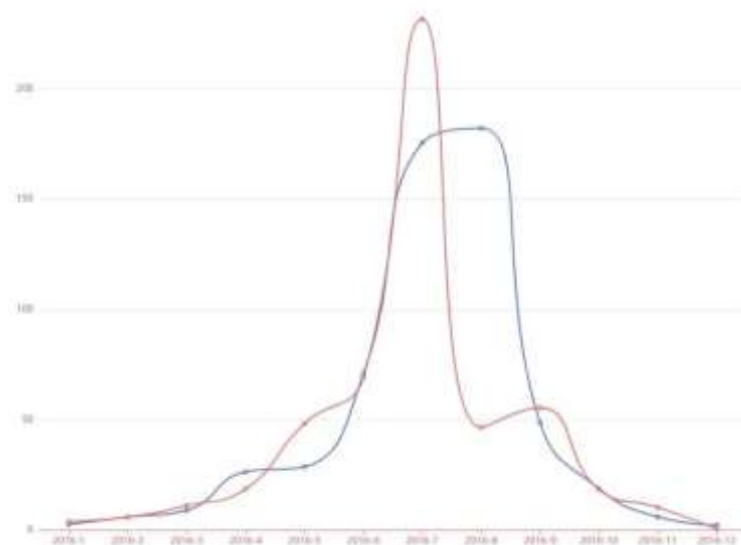


图 2：用户活跃量预测示意图

8. 将训练好的模型保存到虚拟机桌面“提交文档/02\_数据处理与分析/05”

### 【操作说明】

- 1、所有实操均在平台的虚拟机（Ubuntu 系统）中进行，虚拟机内为断网环境。
- 2、从虚拟机桌面“赛题/03\_机器学习/05\_活跃度预测”路径下获取相关资料，结果保存至虚拟机桌面“提交文档/03\_机器学习/05”

文件夹中。

3、完成任务后，根据平台“答题区”要求，完成答题报告并点击“保存”按钮实时保存。

## 第 6 题：视频网站每日广告收益预测

### 【功能说明】

根据视频网站历史广告收益数据，使用机器学习框架 sklearn 编写程序，实现预测该网站每日广告收益功能。

### 【任务要求】

在 machine\_learning\_price.py 文件中实现广告收益预测功能，其中历史广告收益数据存放在 income.csv，具体要求如下：

- 1、从 income.csv 表中读取数据，根据数据字段进行分析，进行必要的清洗、处理工作，处理方式不限。
- 2、从数据中进行特征工作，
- 3、基于机器学习框架 sklearn 进行数据划分，训练集和测试集划分比例为 7:3。
- 4、判读并选择合适的算法进行机器学习建模。
- 5、编写模型训练相关代码，完成模型训练；
- 6、使用可视化库对测试数据的预测结果和真实结果进行可视化对比，并使用 Django 在前端页面中渲染展示；



图 3：每日广告收益预测示例图

### 【操作说明】

- 1、所有实操均在平台的虚拟机（Ubuntu 系统）中进行，虚拟机内为断网环境。
- 2、从虚拟机桌面“赛题/03\_机器学习/06\_收益预测”路径下获取相关资料，结果保存至虚拟机桌面“提交文档/03\_机器学习/06 ”文件夹中。
- 3、完成任务后，根据平台“答题区”要求，完成答题报告并点击“保存”按钮实时保存。

## 模块四：计算机视觉应用（35 分）

### 第 7 题：图像分类之猫狗分类

#### 【功能说明】

使用 TensorFlow 深度学习框架搭建模型，对数据进行数据清洗，

数据预处理等操作，然后利用处理后的数据，完成模型的训练，实现猫狗分类项目。

### 【任务要求】

在 `cat_dog_train.py` 文件中编写代码，实现猫狗分类任务，其中数据集存放在虚拟机桌面“赛题/04\_计算机视觉应用/07\_图像分类/data”路径，具体要求如下：

1、编写 `load_data()` 函数，实现数据读取、数据增强、数据集划分操作：其中数据增强方式不限，数据训练集与测试集划分比例为 8:2。

2、编写 `build_model()` 函数构建深度学习模型和编译，要求模型在训练集的准确率不低于 75%，并将最终模型保存为 “`model.h5`”。

3、编写 `plt_train()` 函数，使用可视化库 Matplotlib 对训练过程进行可视化展示，观察损失值变化、准确率变化。并将可视化图表通过 Django 渲染展示。

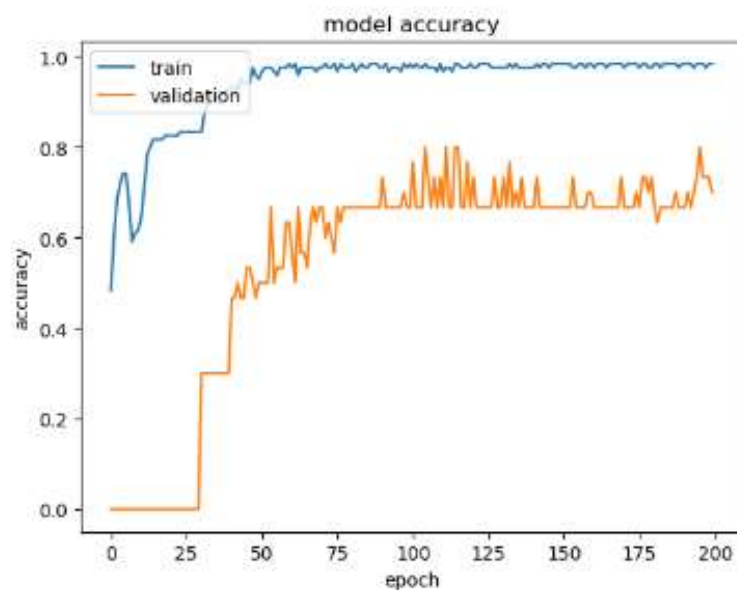


图 4：准确率变化示例图

4、将训练好的模型保存到虚拟机桌面“提交文件/04\_计算机视

觉应用/07\_图像分类/model”中。

5、调用保存好的模型，通过 Django 框架在前端页面展示模型识别效果。



图 5：猫狗分类示例图

### 【操作说明】

1、所有实操均在平台的虚拟机（Ubuntu 系统）中进行，虚拟机内为断网环境。

2、从虚拟机桌面“赛题/04\_计算机视觉应用/07\_图像识别”路径下获取相关资料，结果保存至虚拟机桌面“提交文档/04/07”文件夹中。

3、完成任务后，根据平台“答题区”要求，完成答题报告并点击“保存”按钮实时保存。

## 第 8 题：图像分割之细胞分割

### 【功能说明】

在智慧医疗场景中，细胞分割一直是视觉研究的重点和难度，应用计算机视觉技术辅助医学影像诊断，使用 TensorFlow、Keras 或

PyTorch 构建深度学习模型实现细胞壁分割任务。

### 【任务要求】

在 `cells_train.py` 中编写代码，实现细胞分割功能，数据集存放在“赛题/04\_计算机视觉应用/08\_图像分割/data”目录，具体要求如下：

- 1、从“data”路径加载数据集，根据数据情况对数据进行增强、降噪操作，数据处理方式不限。
- 2、数据集训练集、测试集划分比例设置为 8:2。
- 3、选择适合的深度学习算法，构建深度学习模型，定义模型超参数、选择合适的模型评估函数、优化算法、激活函数等。
- 4、进行模型训练保存模型。
- 5、构建模型预测函数，输入测试集查看模型的分割结果，将结果保存到“提交文档/04\_计算机视觉应用/08”，并通过 Django 渲染结果展示。

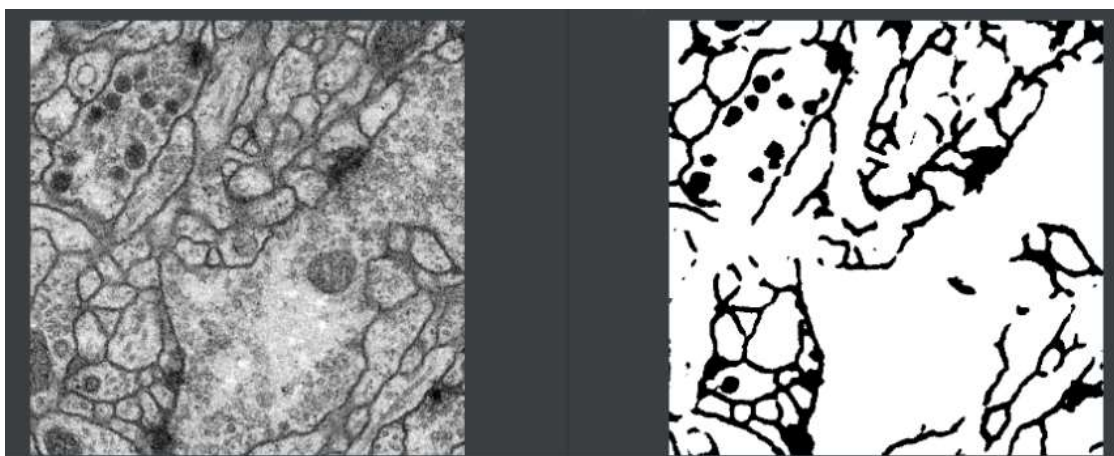


图 5：细胞分割示意图

### 【操作说明】

- 1、所有实操均在平台的虚拟机（Ubuntu 系统）中进行，虚拟机

内为断网环境。

2、从虚拟机桌面“赛题/04\_计算机视觉应用/08\_图像分割”路径下获取相关资料，结果保存至虚拟机桌面“提交文档/04\_计算机视觉应用/08 ”文件夹中。

3、完成任务后，根据平台“答题区”要求，完成答题报告并点击“保存”按钮实时保存。



# 2023

## 金砖国家职业技能大赛 (金砖国家未来技能挑战赛)



金砖职赛微信号