第十二届“麒麟杯”全国开源应用软件开发大赛

手写和语音识别模块

负责人：

联系方式：

# 引言

随着智能手机、电脑等移动信息工具的普及和应用，手写和语音识别也随之发展，手写和语音无疑是最贴合、最方便、最自然的输入方式。其中手写输入常见于银行手写签名确认和日常生活中对生僻字的手写输入；而语音输入能极大解放人类的双手，如果能更为精准识别人声，将会大大提高输入速度。

因此在当今高速发展的社会，一款兼备二者输入功能的模块显得尤为重要。本技术说明旨在提供基于百度AI平台的“手写和语音识别”实施方法，帮助开发者理解、使用和改进该功能。

该功能模块包括手写识别（包括手写汉字、英文和数字）和语音识别，希望能借此更好地向用户介绍此模块的使用方式，也为开发者提供详实的技术描述，以便未来该功能的完善和改进。

针对手写和语音识别，市面上，市面上有不少类似的功能模块，已经能够极大地提高用户输入，比如搜狗的手写和语音识别模块、讯飞的语音识别模块。

但是这类功能模块较为单一，比如无法提供手写英文或数字、无法智能识别方言等缺陷，且难以兼容openKylin操作系统。而该项目将基于百度AI开放平台中手写和语音识别的引擎和openKylin操作系统进行开发一款简易但能弥补上述缺陷的模块。

# 功能介绍

## 2.1 获取手写字图像

**相关源代码文件**：main.py

画布工具可以使得用户通过鼠标、触控板和触摸屏等设备进行手写输入，画布工具包含撤销、删除、清屏等功能可以供用户对所写字进行更改；而保存功能可以直接显示用户所写的文字，用户可以根据结果进行判断是否是所想要的结果进行自行修改；针对老年群体，该画布工具还可以调整笔触大小，为老年人提供便利。

（1）onLeftButtonDown、onLeftButtonMove

左键，画图函数，判断用户是否写入手写字。

（2）onRightButtonUp

右键，对所画图像进行操作的菜单栏。

（3）Open、Clear、Back、

对所绘制图像进行操作，Open可以打开一幅图片（本模块还能进行识别印刷体图片上的文字），Clear可以将所绘制的图片清空，Back可以撤销上一步所绘制的笔画。

（4）Save

保存用户所绘制的图像，用于下一步的图像识别。

（5）drawText

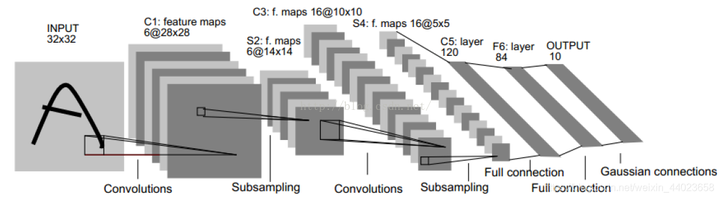
绘制文本框文字。

## 2.2 识别手写数字（测试模块）

**相关源代码文件**：test\_digits.py，train.py

采用LeNet-5模型

1989年，Yann Lecun在论文 Gradient-Based Learning Applied to Document Recognition 中提出了LeNet-5网络模型（现已改进为LeNet-6），它是一种用于手写体字符识别的非常有效的神经网络。LeNet-5 网络架构的成功应用，极大推动着深度学习领域的发展。



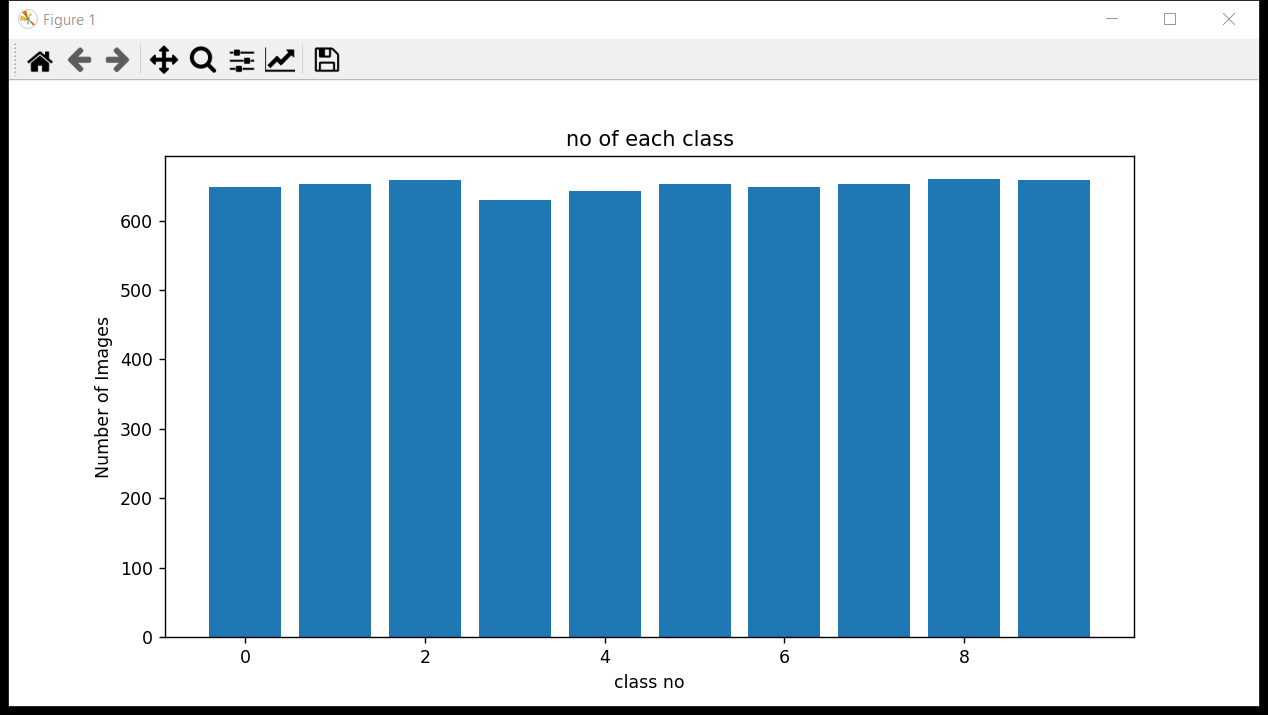
上图是该模型应用的简易原理图

(1) train.py

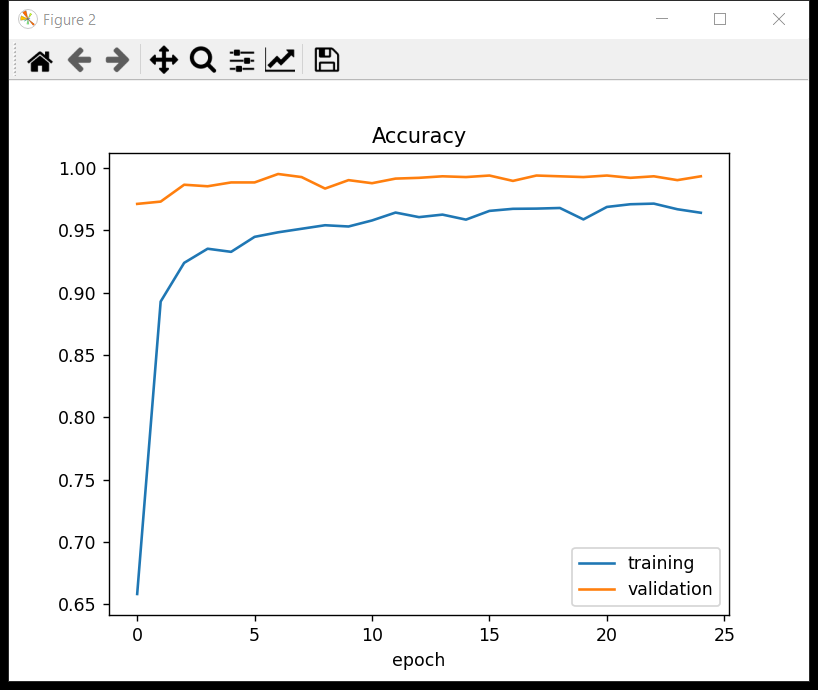
读取数据包，将数据划分为训练集、测试集和验证集，并分别读取数据集中的图片，进行图片预处理。建立LeNet-5模型，并开始训练，得到模型结果。

其中数据集选用的是MINST数据集。选该数据集的好处是，对训练来说，MNIST数据集中数据更适合训练出相同的模型。使得训练的模型能够更好的识别手写体数字。

其训练集如下：



测试准确性：



(2) test\_digits.py

可在画板上绘制手写数字保存为test\_img.jpg进行测试。

## 2.3 识别手写汉字和英文

**相关源代码文件**：main.py

针对手写汉字和英文，这里采用开源或商业的手写识别引擎和语音识别引擎，即百度AI中手写汉字识别Api进行识别。

（1）Get\_Picture

采集在画图Gui界面的文字图片

（2）Get\_str\_from\_img

将采集获得的图片转换成文字形式

## 2.4 识别语音

**相关源代码文件**：voice.py

针对短音频，这里同样采用开源或商业的手写识别引擎和语音识别引擎，即百度AI中手写汉字识别Api进行识别。

（1）get\_file\_content

获取音频文件

（2）Get\_str\_from\_voice

将采集语音音频转换成文字形式

# 运行方式：

## 3.1编译程序

编译需要安装第三方库：百度Api库，Opencv库，tkinter库，PIL库，numpy库，keras库等第三方库

## 3.2运行程序

手写识别模块直接点击main.py即可运行，对于语音识别模块需将音频文件命名为test.wav才可在voice.py上进行识别（如果需要单独测试数字识别模块可以将手写数字保存为test\_img，运行test\_digits.py进行数字识别）。

# 设计思路

## 4.1 获取手写字图像

**相关源代码文件**：main.py

画布工具可以使得用户通过鼠标、触控板和触摸屏等设备进行手写输入，画布工具包含撤销、删除、清屏等功能可以供用户对所写字进行更改；而保存功能可以直接显示用户所写的文字，用户可以根据结果进行判断是否是所想要的结果进行自行修改；针对老年群体，该画布工具还可以调整笔触大小，为老年人提供便利。

（1）onLeftButtonDown、onLeftButtonMove

左键，画图函数，判断用户是否写入手写字。

（2）onRightButtonUp

右键，对所画图像进行操作的菜单栏。

（3）Open、Clear、Back、

对所绘制图像进行操作，Open可以打开一幅图片（本模块还能进行识别印刷体图片上的文字），Clear可以将所绘制的图片清空，Back可以撤销上一步所绘制的笔画。

（4）Save

保存用户所绘制的图像，用于下一步的图像识别。

（5）drawText

绘制文本框文字。

## 4.2 获取手写数字（测试模块）

**相关源代码文件**：test\_digits.py，train.py

(1) train.py

读取数据包，将数据划分为训练集、测试集和验证集，并分别读取数据集中的图片，进行图片预处理。建立LeNet-5模型，并开始训练，得到模型结果。

其中数据集选用的是MINST数据集。选该数据集的好处是，对训练来说，MNIST数据集中数据更适合训练出相同的模型。使得训练的模型能够更好的识别手写体数字。

(2) test\_digits.py

可在画板上绘制手写数字，或通过截图保存为test\_img.jpg进行测试。

## 4.3 识别手写汉字和英文

**相关源代码文件**：main.py

针对手写汉字和英文，这里采用开源或商业的手写识别引擎和语音识别引擎，即百度AI中手写汉字识别Api进行识别。

（1）Get\_Picture

使用类Canvas的画板方式，截图保存在画图Gui界面的文字图片。

（2）Get\_str\_from\_img

将采集获得的图片转换成文字形式，并在终端显示语音输入结果。

## 4.4 识别语音

**相关源代码文件**：voice.py

针对短音频，这里同样采用开源或商业的手写识别引擎和语音识别引擎，即百度AI中手写汉字识别Api进行识别。

（1）get\_file\_content

先通过电脑麦克风录制语音，获取音频文件。

（2）Get\_str\_from\_voice

将采集语音音频转换成文字形式。

# 结论

随着技术进步增长和实际应用高效性，保证高效输入和解放双手成为了每一个输入模块的前提条件。本项目中开发的“手写和语音识别模块”提供了一个强大且有效的方法来提高用户输入方式。具体功能涵盖获取手写图片、读取和识别手写数字、识别手写中文英文和识别短音频语音。

此工具基于Kylin系统，并利用了现有的库，如百度Api库，Opencv库，tkinter库，PIL库等，确保了信息识别的准确性和可靠性。同时，技术说明为开发者提供了详细的实现细节，从而为未来的扩展和优化提供了基础。此外，工具的实际应用已经经过了严格的测试，保证了在实际环境中的稳定运行。

总的来说，此“手写和语音识别”不仅对开发者有价值，也为维护者和用户提供了一个实用且易于使用的工具模块，确保了输入的高效性和便捷性。在现今高速发展、快节奏的时代，这样的工具无疑是在用户实际办公中不可或缺的一部分。