中国大学生计算机设计大赛

作品编号：2023021538

作品名称：基于深度学习的盲人自然语言书籍识别阅读软件

作　　者：王云飞，覃浩平，李明哲

版本编号：

填写日期：2023.4.14

目录

第一章 需求分析

21 世纪，科技迅速发展，社会对弱势群体也更加关注。 盲人是如今弱势群体中最多的一类，我国目前有 700 万左右盲人，加上高度近视与视力障碍的将近 4000 万人，相当于每三十个人中就有一个人是盲人。与此同时，当下社会对于教育和知识的重视程度日益加强，但尽管是对普通人来说最简单的阅读，对盲人而言也不是一件简单的事。

根据《中华人民共和国残疾人保障法》第四十三条……（二）组织和扶持盲文读物、盲人有声读物及其他残疾人读物的编写和出版，根据盲人的实际需要，在公共图书馆设立盲文读物、盲人有声读物图书室；第五十条……国家鼓励和支持提供电信、广播电视服务的单位对盲人、听力残疾人、言语残疾人给予优惠。第五十四条……国家举办的各类升学考试、职业资格考试和任职考试，有盲人参加的，应当为盲人提供盲文试卷、电子试卷或者由专门的工作人员予以协助。这些条例表明国家重视盲人自主阅读的能力，并且也说明在现实应用场景中盲人自主阅读拥有非常大的潜在发展市场。

但目前即使是让盲人获取最简单的生活信息的方式都并不多，且大多都存在一定缺陷，无法满足盲人的基本生活和学习需求。

比如盲文的出现曾一度被看作是盲人的福音，但很快，盲文书籍高昂的造价和难以长时间保存的困境让盲文的发展陷入了泥潭。再有，现如今出现的一些有声书，同样可以让盲人“阅读”。但差强人意的是，因为有声书的录制成本越来越高，目前市场上的有声书几乎全部是能够获得足够利润的网络小说类型的书籍，而名家著作、经典名著等有声书的比例并不高，显然这并不能完全满足所有盲人的需求。

最重要的是，通过我们的了解，大多数盲人希望能够不借助于他人，自由自主的阅读他们想读的书，享受阅读。文明的书籍成果浩如烟海，盲人也想要和正常人一样，能够自如的阅读任何一本自己想了解的书籍。

本项目致力于通过智能系统关于解决盲人阅读的这一问题，让盲人能过自主的、不借助他人帮助地阅读面向普通人的自然语言书籍，为了知识的无障碍传播贡献力量。

本项目成果可用于盲人教育事业，辅助盲人学习课本知识，完成作业、考试等教育过程，促进盲人教育事业的发展。另外，盲人可以在本产品的帮助下自由的阅读任何一本自己感兴趣的实体书籍书籍，包括且不限于出版书籍、报纸、杂志等。不再受到盲文书籍发行迟缓、价格昂贵、涉及内容少的限制，不再局限于少数的盲文书籍，降低了盲人阅读成本的同时，让数量庞大无比的实体书籍不再对他们设防，浩如烟海的人类精神财富任他们汲取，并使得

盲人也能及时了解时事，跟上时代的前进步伐。

本产品对于促进盲人教育事业以及提升盲人社会参与感、培育盲人社会责任感、维护社会安定具有重大意义。

第二章 概要设计

“逸读”是通过采集文字图像及预处理，采用文字点字特征提取与识别技术从采集到的图像中提取识别文字，并通过 TTS 语音技术将识别到的文字转换为语音信息进行朗读，最后利用语音识别技术实现语音控制来实现基于人工智能的盲人自然语言书籍阅读技术。通过这四个模块的协同工作，组成集文字识别、语音控制、语音播报于一体的盲人助读器，盲人可以利用该工具实现针对所有书籍的阅读，而不再需要他人协助或者制作昂贵的盲文书籍。

为了能够让盲人能够轻松便捷地使用我们的产品进行阅读，项目团队根据盲人阅读的实际情况设计了使移动端软件和产品硬件能够相互配合的产品功能。

（1） 软件功能：对于用户所需要阅读的文章，“逸读”能够通过文字识别模型中的文字图像的采集及预处理功能对文章的文字进行采集并进行预处理以为后续的文字提取与识别做准备，再由文字识别模型中的文字点字特征提取和识别功能从文章中提取文字特征并与模型数据库的文字特征进行匹配。在完成对文章中文字的匹配后，“逸读”中的语音合成模型通过 TTS 语音合成技术，对文章的内容进行语音播报，使用户愉快轻松地聆听所需要阅读的文章。

（2） 硬件功能：为了能够让视障人士能够方便地对移动端阅读软件进行操作，产品配备了一个硬件支架。支架上装载有一个用于拍摄的摄像头。用户能够通过手机和摄像头的无线连接，在手机上使用“逸读”app 操控硬件支架上的摄像头对支架上的书籍进行拍摄。硬件支架上配有书籍固定装置，用户可根据需要对书籍进行固定。

（3） 产品操作流程：打开“逸读”软件，首先由用户使用数据线将摄像头和手机进行连接，当摄像头与手机连接成功时，“逸读”软件会对用户语音提示连接成功。此时“逸读”软件界面会显示摄像头所拍摄到的画面。摄像头能够智能对焦以使拍摄画面清晰。用户单击屏幕，“逸读”软件就会操控摄像头对硬件支架上的书籍进行拍摄识别。当文字识别成功时，“逸读”软件就会自动朗读书籍的左页文字。再次单击屏幕，“逸读”软件会立刻停止对左

页文字的朗读，并且开始朗读右页的文字。此时单击屏幕，“逸读”软件会立刻停止对右页文字的朗读，并跳转回摄像头拍摄的界面。

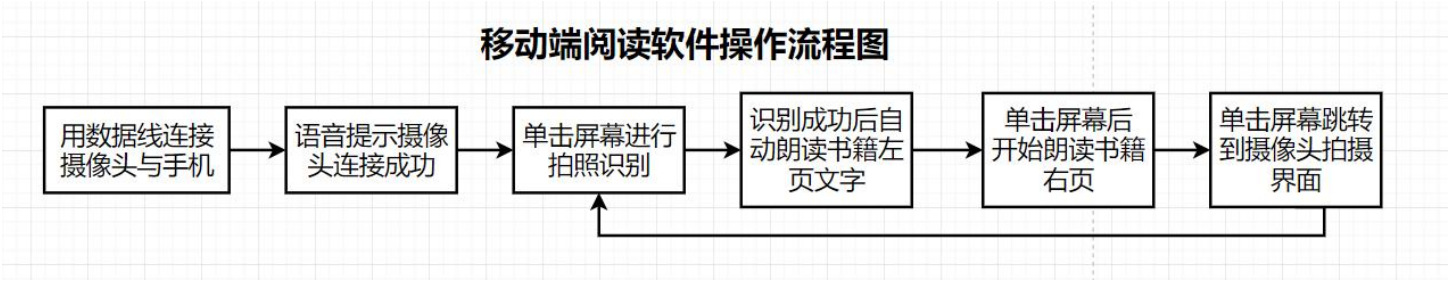

图 2-1“逸读”移动端阅读软件操作流程图

第三章 详细设计

1. 产品具体功能

（1） 连接外部设备

本软件在打开之前需要连接外部USB摄像头，否则无法运行软件。连接好外部USB摄像头后即可运行本软件。

（2） 拍摄图像

运行本软件后，用户可点击屏幕上的任意位置即可拍摄图像。拍照完成后，程序会自动将图像存至指定路径下，为软件实现后续功能做准备。

（3） 图像预处理

由于纸质书籍存在单双页问题，故软件会自动识别图像中纸质书籍为单页还是双页。如果为单页，则不进行预处理。如果为双页，则先将图像划分为两张单页图像。

（4） 加载并运行文字识别模型

软件得到单张或两张单页图像后，软件会先加载深度学习文字识别模型，再逐一对单页图像进行文字识别，并将已识别的图像和文字识别结果显示在屏幕上。

（5） 语音合成

在软件得到文字识别结果后，软件会利用语音合成模型将文字识别模型得到的结果从文字形式转为语音形式并输出。

（6） 退出

如果用户要离开当前软件，点击退出按钮，软件将自动关闭。

2. 产品研发路线

（1） 市场调查与资料收集

本项目从视障人士对书籍阅读的实际需求出发，在前期对盲人阅读的市场进行了调研，同时收集了相关的资料，初步了解了盲人对于书籍阅读的基本需求，论证了项目的可行性，明确了创新创业训练的目的、内容以及技术路线。

（2） 训练模型

在前期工作的基础上，项目小组收集了大量各种类型书籍的图片数据集，用于对文字识别模型的基础训练，并且补充了大量特殊的场景下训练数据（例如弯曲的书籍页面），用以提高模型的识别精度。在完成对文字识别模型训练的同时，项目小组也以相同的方式完成了

文字转语言模型的训练。

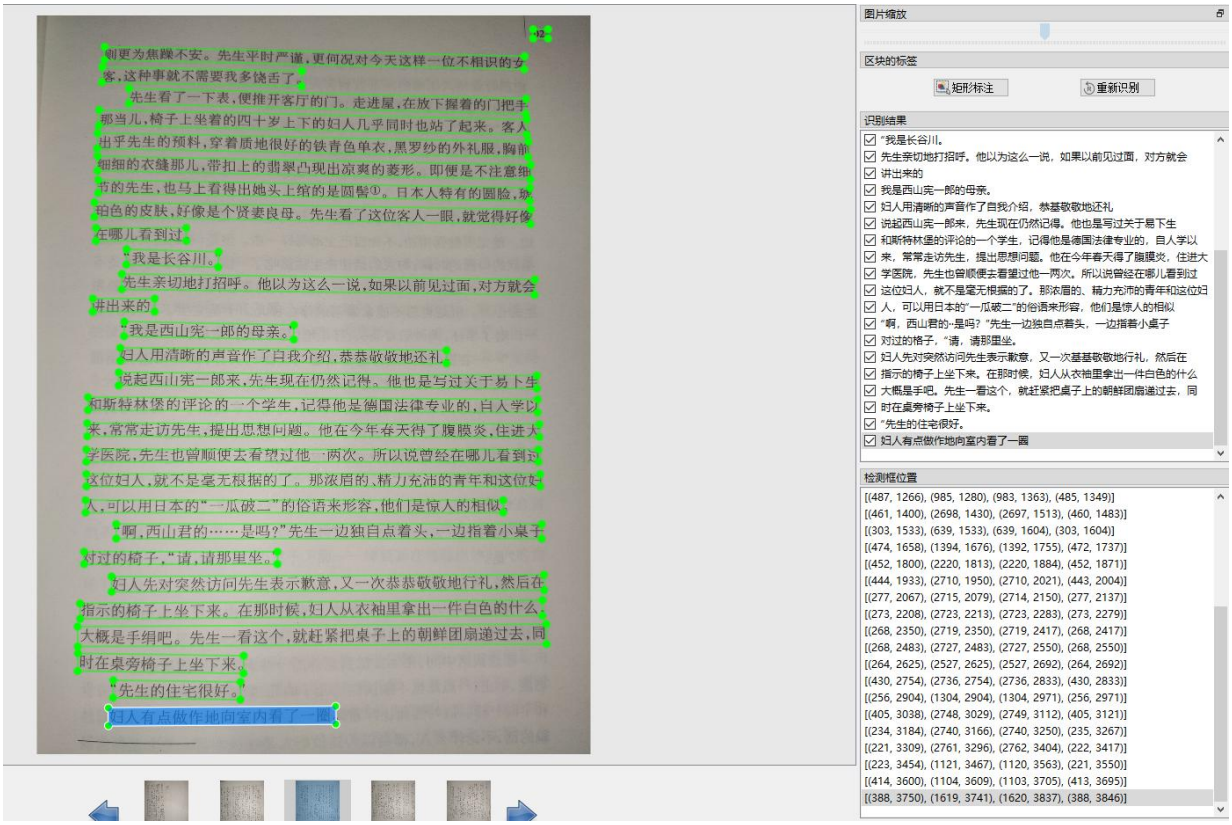

图 3-2-1 文字识别模型训练图

（3） 设计开发移动端阅读软件

在模型训练完成之后，小组成员在开始学习移动端软件的专业开发知识，参考文字识别模型和文字转语音模型移动端部署的相关资料，不仅完成了模型在移动端软件上的集成，并且成功设计出了一款能让视障人士能够便捷使用的移动端阅读软件。

（4） 设计阅读软件辅助设备

为了能够让视障人士能够方便地对移动端阅读软件进行操作，小组成员利用机械设计的基本知识，并学习机械绘图专业知识和软件工具，参考机械相关设计资料，完成软件辅助设备的结构设计，并使用 3D 打印技术完成了样品的制作。

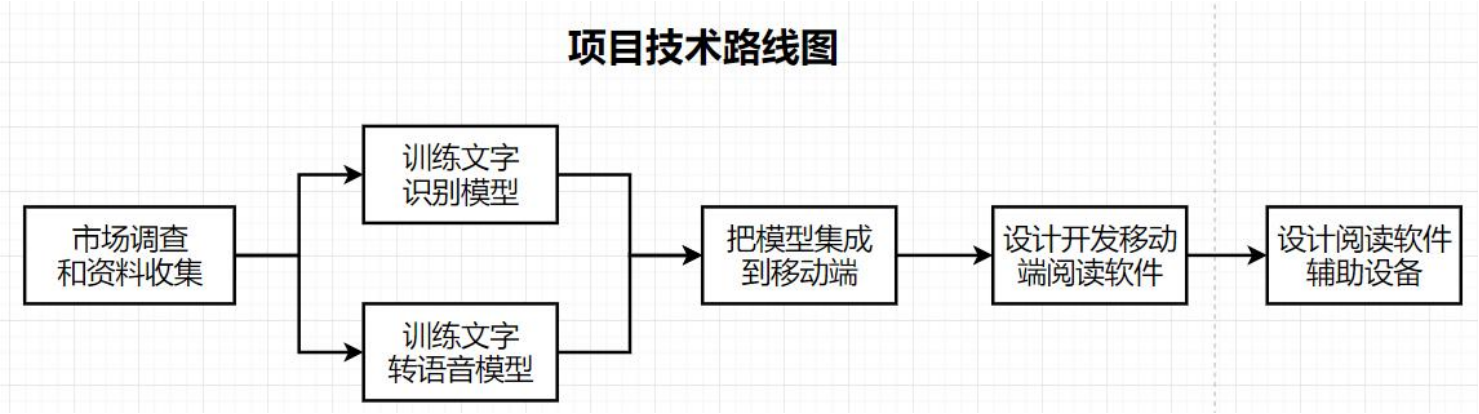

图 3-2-2 项目研发路线图

3. 关键技术

（1） 基于深度学习模型的 OCR 文字识别算法

传统 OCR 识别算法采用统计模式，包括图像的预处理、二值化、连通域分析、版面分析、行切分、字切分、单字符识别和后处理等步骤。传统的 OCR 识别算法存在着一系列缺陷，如处理流程较长、识别精度低、处理效率低下等问题，而深度学习算法可以有效地规避传统 OCR 识别的不足。深度学习算法通过组合低层特征形成更加抽象的高层表示属性类别或特征，挖掘数据的分布式特征表示。借助神经网络来模拟人脑进行分析、学习和训练，即模仿人脑机制来分析图像、声音和文本等数据。针对识别场景文字的问题，基于深度学习的识别方法在性能上远远超过传统的文本检测方法，体现出了应用深度学习的优势。

在不受约束的环境中识别文本仍然是当下面临的最具挑战的问题，通过研究者们的不懈努力，已经有了很多针对上述问题的基于深度学习的文字识别模型。如基于深度学习的 OCR 定位与识别通过卷积神经网络 CNN、循环神经网络 RNN、长短期记忆网络 LSTM 技术实现，可在灰度图像上实现文字区域的自动定位和整行文字的识别，解决了传统 OCR 技术中单字识别无法借助上下文来判断形似字的问题。此外，智能 OCR 识别技术在低质量图片的容忍能力和识别准确率方面得到了显著的提升，可在印刷体低分辨率与模糊字符识别、印刷体复杂或者非均匀背景识别、印刷体多语言混合识别、印刷体艺术字体识别、手写小写数字识别、手写大写金额识别、手写通用文本识别等场景下实现高效的识别和分类。基于深度学习的智能 OCR 识别技术支持移动设备拍摄的图像识别，可适用于对焦不准、高噪声、低分辨率、强光影等复杂背景。

对于普通用户来说，OCR识别的需求更多的发生在移动端，即用户更喜欢用手机拍照后即可识别。目前，基于深度学习的智能OCR技术综合了已有的信息化技术，可在各种移动

端实现适配。首先，基于轻量级深度学习技术，实现移动端的取图功能；其次，融合视频流识别技术，即从视频中识别出有效信息。深度学习网络可高效地学习到边缘情况，通过边缘的检测，得到物体的边缘轮廓，然后通过边缘跟踪合并，保障识别效果。移动端适配网络计算量很小，大多数的移动端设备均支持，即使透视变换很严重的图像也能很好地校正，保证移动端识别的准确率。

（2） PP-OCR 文字识别模型

我们选用 PaddlePaddle 的 PP-OCR 模型作为项目的文字识别模型。PP-OCR 自研的实用的超轻量 OCR 系统。在实现 OCR 前沿算法的基础上，考虑精度与速度的平衡，进行模型瘦身和深度优化。该系统由文本检测、检测框校正、文本识别这三个部分组成，系统架构如图 3-3-2-1 所示。其中文本检测算法选用可微分二值化处理算法，文本识别算法选用 IJCAI 2022 最新收录的文本识别算法 SVTR，并在检测和识别模块之间添加文本方向分类器，以应对不同方向的文本识别。

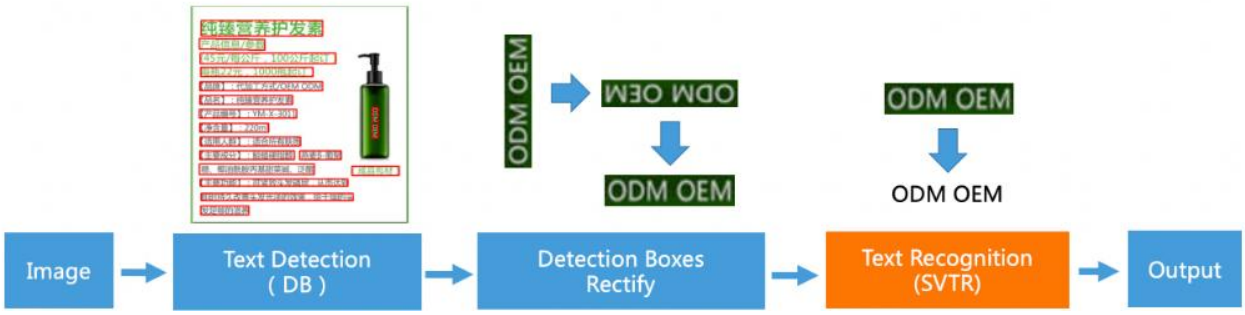

图 3-3-2-1 PP-OCR 系统架构

PP-OCR 的检测模型引入了 CML（Collaborative Mutual Learning）协同互学习文本检测蒸馏策略。如图 3-3-2-2 所示，在 CML 中，两个子学生模型使用 DML（Deep Mutual Learning）方法相互学习。同时，有一个教师模型来指导两个学生模型的学习。教师模型使用 ResNet18 作为主干，学生模型使用 Mobinetv3 大型模型作为主干。PP-OCR 分别针对教师模型和学生模型进行进一步效果优化。其中，PP-OCR 在对教师模型优化时，使用了大感受野的 PAN 结构 LK-PAN 和引入了 DML 蒸馏策略，可以有效提升文本检测模型的精度。；在对学生模型优化时，使用了残差注意力机制的 FPN 结构 RSE-FPN（Residual Squeeze-and-Excitation FPN），进一步提升特征图的表征能力。

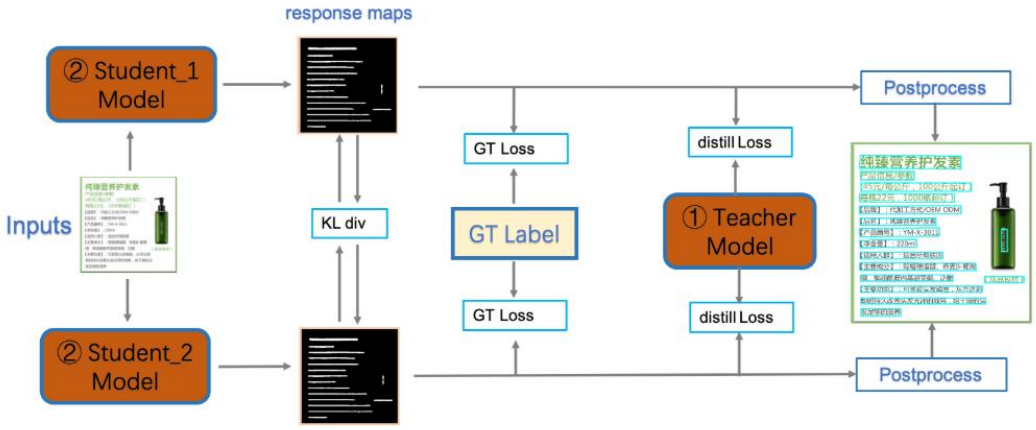

图 3-3-2-2 基于 CLM 的文本检测模型

PP-OCR 的识别模块是基于文本识别算法 SVTR 优化。如图 3-3-2-3 所示，SVTR 是一种文本定制的识别模型，它是一个三级逐步下采样的网络，引入了局部和全局混合块，分别提取笔划特征和字符间相关性，并结合多尺度 backbone，形成多粒度特征描述。SVRT 不再采用 RNN 结构，通过引入 Transformers 结构更加有效地挖掘文本行图像的上下文信息，从而提升文本识别能力。此外，PP-OCR 还采用 Attention 模块 CTC 训练，融合多种文本特征的表达，TextConAug，使用大量无标注的文本行数据、联合互学习策略、无标注数据挖掘方案等优化策略以提高模型精度并加快模型识别速度。

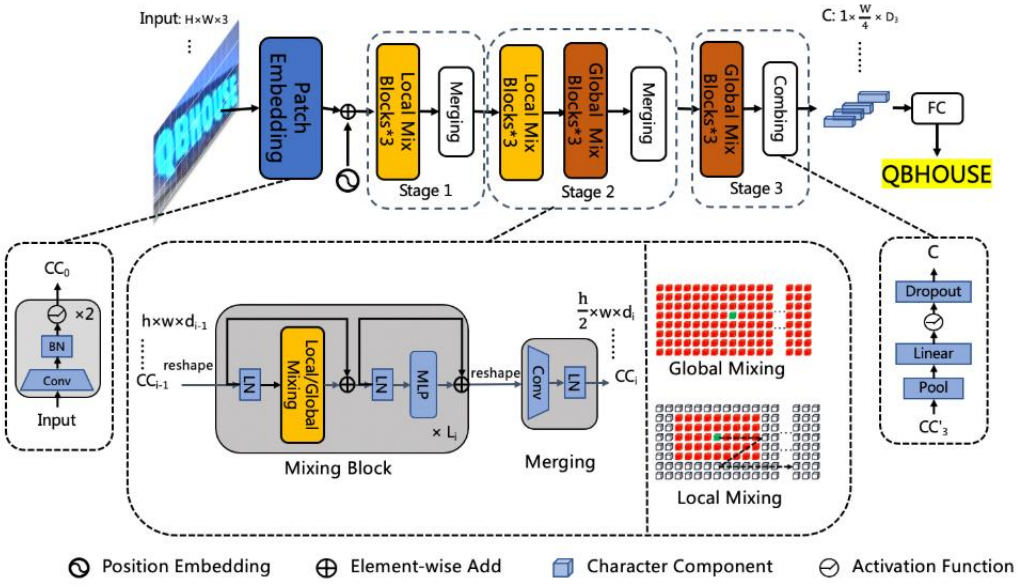

图 3-3-2-3 SVTR 模型

（3） Deep voice3 语音合成模型

语音合成 TTS 模型由 Deep voice3 提供。Deep Voice3 是由百度提出的一个全新的全卷

积 TTS 架构，模型由于采用全卷积而非 GRU 来提取文本及频谱特征，不仅可以大幅提高训练时 GPU 的利用率，还能大大加快训练速度。 Deep Voice3 能够将各种文本特征（如字符、音素、重音）转换为各种声码器参数，如梅尔谱、线性对数谱、基频、频谱包络等，而这些声码器参数可用作波形合成模型的输入。Deep Voice3 架构如图 3-3-3-1 所示，包括 3 个组件，即编码器、解码器和转换器。

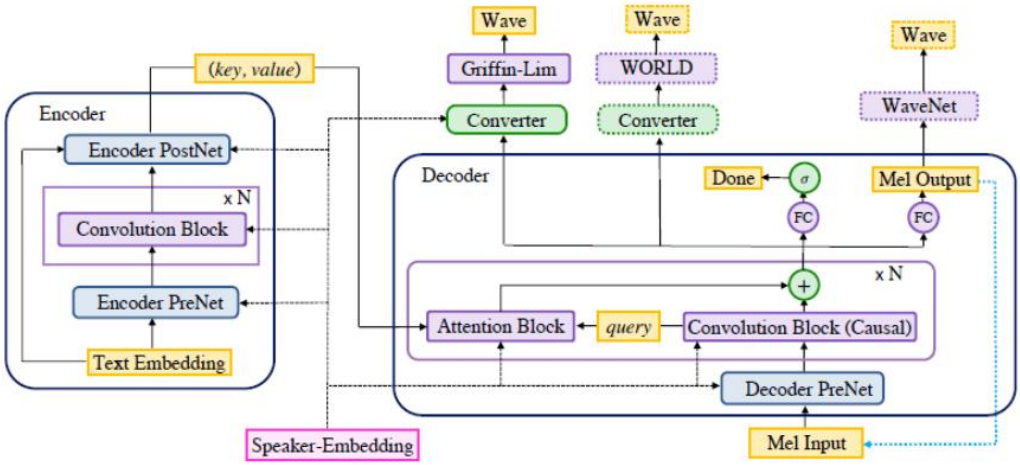

图 3-3-3-1 Deep voice 模型架构

编码器完全由卷积构成，用于提取文本特征，编码器网络首先从文本编码开始，将字符或音素转换为可训练的向量，然后将其送入全连接层以投影到目标维度。得到的输出再送入一系列卷积块，以提取时间相关的文本信息。最后，它们被投影回 Text Embedding 维度以创建注意力键向量。键向量被各个注意力块用来计算注意力权重，而最终的上下文向量被计算为值向量的加权平均。

解码器也是完全由卷积构成，利用多跳卷积注意力机制（multi-hop convolutional attention mechanism）将提取的文本特征，以一种自回归的方式解码成低维的音频特征。解码器以自回归的模式预测接下来的 r（r>1）帧梅尔谱。由于不能利用后面时刻的数据，所以解码器采用的是 causal convolution。梅尔谱数据先经过 PreNet，然后通过 casual convolution 层将其变为 query 矩阵。再与编码器的输出的 Key 和 Value 矩阵进行传统点积计算方法的 attention 运算。如此积累多层，最后经过全连接层预测接下来的 r 帧梅尔谱，并且还会预测是否该停止预测。

转换器同样是完全由卷积构成，它从解码器隐藏状态预测最终声码器的参数。与解码器不同，转换器是非因果的，因此它可以依赖未来的上下文信息。转换器网络将解码器的最后隐藏层的输出作为输入，转换器包含若干非因果卷积块，然后预测下游声码器的参数。

（4） 3D 打印技术

本系统的硬件设备采用 3D 打印技术制作。3D 打印，即快速成型技术的一种，它是一种

以数字模型文件为基础，运用粉末状金属或塑料等可粘合材料，通过逐层打印的方式来构造物体的技术。

3D打印技术的原理基本分为三部分：

1） 三维设计

3D打印的设计过程是：先通过计算机辅助设计（CAD）或计算机动画建模软件建模，再将建成的三维模型“分割”成逐层的截面，从而知道打印机逐层打印。设计软件和打印机之间协作的标准文件格式STL文件格式。一个STL文件使用三角面来大致模拟物体的表面。三角面越小其生成的表面分辨率越高。PLY是一种通过扫描来产生三维文件的扫描器，其生成的VRML或者WRL文件经常被用作全彩打印的输入文件。

2） 打印过程

打印机通过读取文件中的横截面信息，用液体状，粉状或片状的材料将这些截面逐层的打印出来，再将各层截面以各种方式粘合起来从而制造出一个实体.这种技术的特点在于其集合可以造出任何形状的物品

传统的制造技术如住塑法可以以较低的成本大量制造聚合物产品，而3D打印技术则可以以更快，更有弹性以及更低成本的办法生产数量相对较少的产品。一个桌面尺寸的3D打印机就可以满足设计者或概念开发小组制造模型的需要。

3） 完成

目前3D打印机的分辨力对大多数应用来说已经足够（在弯曲的表面可能会比较粗糙，像图像上的据齿一样），要获得更高分辨率的物品可以通过如下方法：先用当前的3D打印机打出稍大一点的物体，再稍微经过表面打磨即可得到表面光滑的“高分辨率”物品。有些技术可以同时使用多种材料进行打印。有些技术在打印的过程中还会用到支撑物，比如在打印出一些有倒挂状的物体时就需要用到一些易于除去的东西（如可溶的东西）作为支撑物。

（5） Android 软件开发

为了使视障人士愉快轻松地聆听所需要阅读的文章，我们将 OCR 模型和 TTS 模型集成到移动端，开发了一款基于深度学习的纸质书籍文字识别阅读软件“逸读”。本软件基于 Android 系统开发，运用 Android studio 的功能设计出原始的软件功能代码文件，在此基础上运用 Android studio 编译器对功能代码文件进行编译，编译成可在 Android 系统上运行的*.apk 文件，只要在系统版本为 Android 6.0 及其以上的手机或其他硬件设备上都可以运行本软件，成功地降低本软件的运行环境要求，提高可移植性。

4. 创新点

相比以往研究更多注重于理论层面，本项目坚持理论与实践结合，所研究的系统采用基于机器学习和计算机视觉的文字图像识别系统来解决视障人群的问题，创新性地将智能语音识别技术、文字识别技术与 TTS 语言合成技术相结合，并创新性地通过数据合成工具批量合成大量与目标场景类似的图像来扩大训练集以获取更高精度的模型。

5. 运营模式

“逸读”的发展规划主要分为前期、中期、后期三个时间段。

在前期，“逸读”的服务主要依托网络进行整套产品的销售，应用场景主要包含盲人图书馆，顾客家内，盲人阅读室等。系统平台初步搭建，程序设计基本完善，能够在各大应用商店下载使用，项目产品已投入市场。此时主要致力于收集用户数据，设计相关数据模型进行训练，根据盲人用户体验进行逐步改进。

在中期，“逸读”将进行第二轮融资，打通移动端，并组建专业的团队对软件及产品进行更新和升级，并且投放广告，扩大影响，吸引更多人的关注。这个时候主要考虑是如何引入下游用户，这一步骤“逸读”的战略是做 WebAPP 嵌入微信的模式，因为微信目前用户已达 8 亿，庞大的用户群体已经养成使用习惯，只需要扫码即可享受“逸读”的服务。与此同时，“逸读”的相关产品将会与一些盲人应用厂商达成合作。与盲人图书馆，盲人眼镜等达成产品合作，进行实地试点并引导使之成为示范性标杆的广告效应，进而展开更多的合作。

在后期，“逸读”致力于成为相关技术方案提供商及标准制定者的角色。“逸读”通过各种活动运营，产品销售及专利申请等创造营收，根据社会反映情况进行改进与提高。此时更应关注项目所带来的社会效益，并尽量降低成本，为更多的盲人们提供福利。

第四章 测试报告

1. 测试步骤

（1） 对单页双页书籍分别进行测试

（2） 对平滑和弯曲的书籍分别进行测试

（3） 对非书籍的物品进行测试

2. 测试结果

测试结果表明，“逸读”app 对于平滑的书页识别率高达 98%，对于弯曲的书籍也能达到 90%以上的识别率。并且“逸读”app 模型能够识别出书籍的单双页，并且能够分别对单双页书籍进行分类识别。对于非书籍类物品，“逸读”app 也能够识别出并对使用者进行语音提醒。

第五章 安装及使用

1. 安装环境要求

（1） 硬件要求

处理器主频：2GHZ 及以上；
内存：2G 及；

（2） 软件要求

系统：Android 6.0 或以上以上版本；
运行环境： Android 6.0 及以上版本。

2. 软件使用流程

连接好外部 USB 摄像头后，打开软件。
软件成功初始化并获取相应权限后，会显示界面，如图 5-2-1 所示。

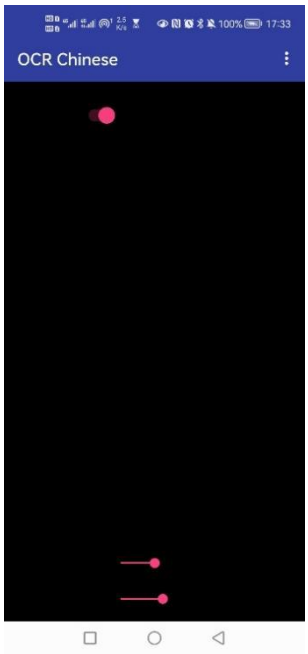

图 5-2-1 软件界面

(1) 用户点击屏幕的任意位置，即可完成拍照，软件能根据书籍是单页还是双页对书籍自动进行文字识别操作，将结果以文字输出在屏幕上，并以语音的形式通过扬声器或耳机等设备进行输出。双击屏幕可实现暂停/播放的功能。若识别为双页书籍，则第一次将对书籍左侧页面进行阅读。示例为识别双页书籍的左侧页面，如图 5-2-2 所示。

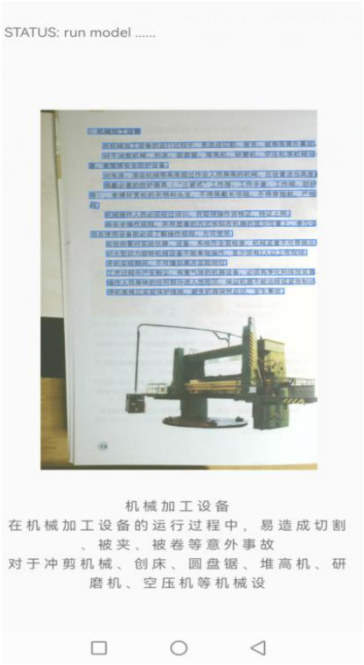

图 5-2-2 软件识别书籍左侧文字

(2) 若为单页书籍，再次单击屏幕后，软件将跳转到摄像头拍摄预览界面；若为双页书籍再次单击屏幕后，软件将识别书籍右侧页面文字。示例为识别双页书籍的右侧页面，如图 5-2-3 所示。

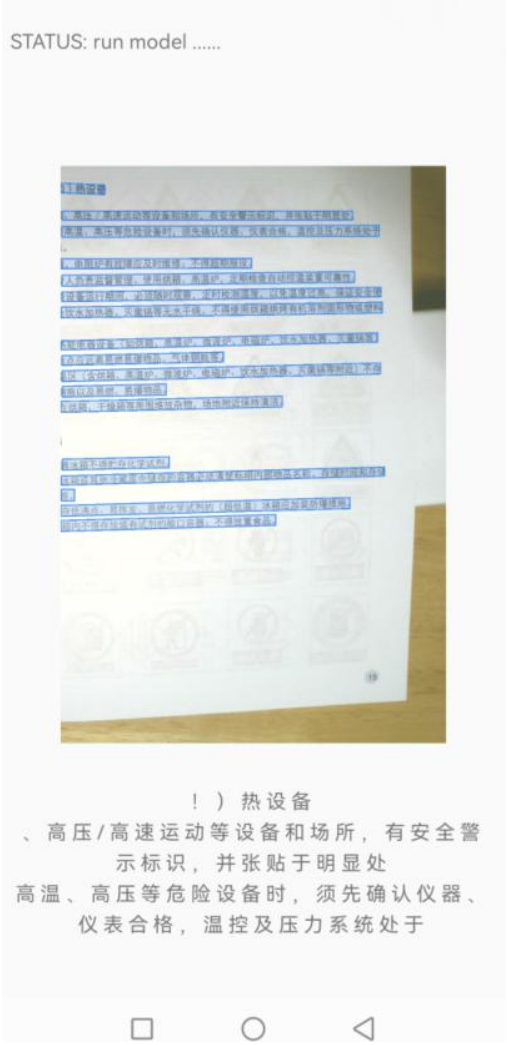

图 5-2-3 软件识别书籍右侧文字

（3） 点击退出按钮关闭软件。软件流程图如图 5-2-4 所示。

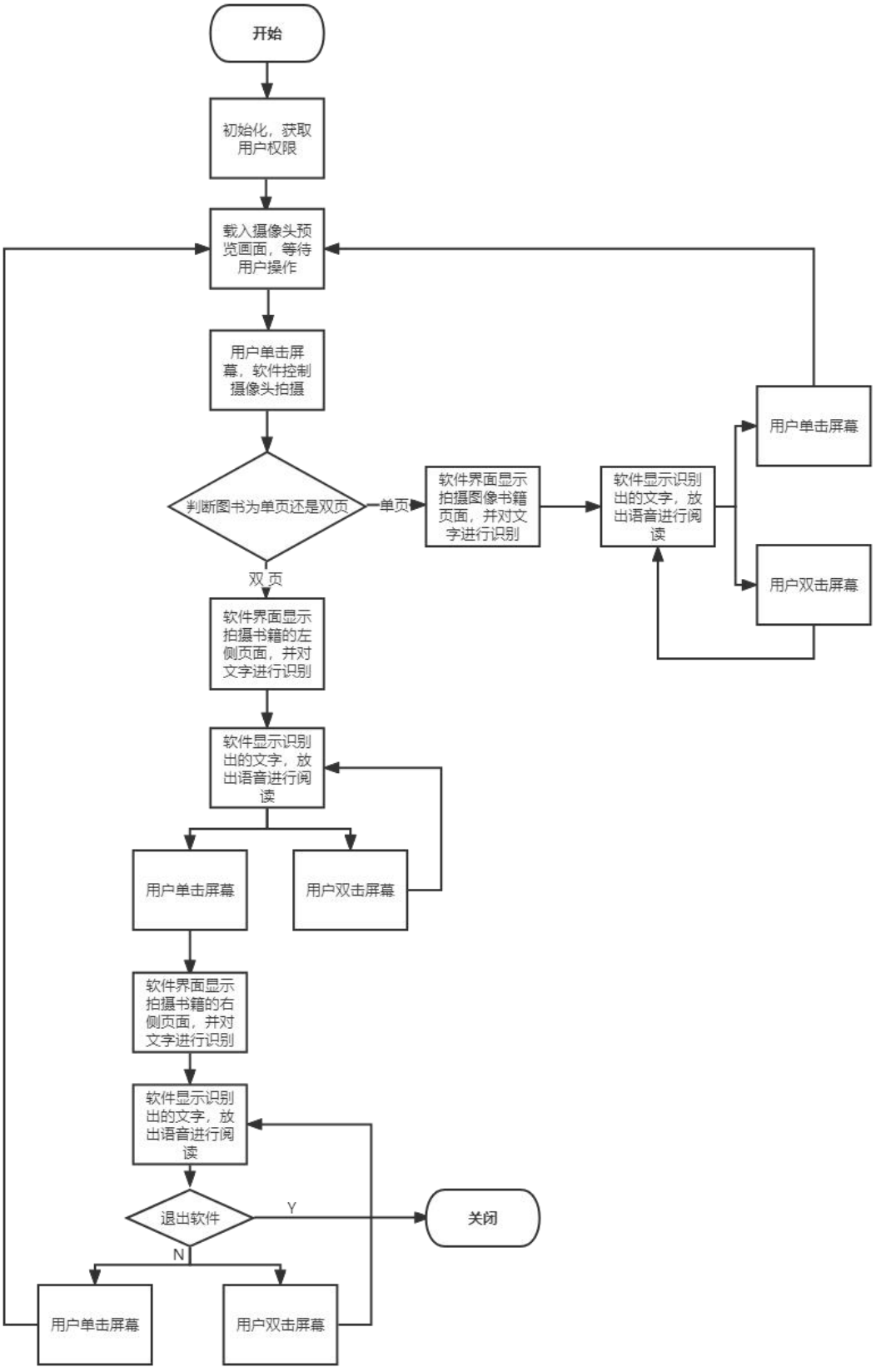
开始
初始化，获取用户权限
载入摄像头预览画面，等待用户操作
用户单击屏幕，软件控制摄像头拍摄
判断图书为单页还是双页
单页
软件界面显示拍摄图像书籍页面，并对文字进行识别
软件显示识别出的文字，放出语音进行阅读
用户单击屏幕
用户双击屏幕
双页
软件界面显示拍摄书籍的左侧页面，并对文字进行识别
软件显示识别出的文字，放出语音进行阅读
用户单击屏幕
用户双击屏幕
软件界面显示拍摄书籍的右侧页面，并对文字进行识别
软件显示识别出的文字，放出语音进行阅读
退出软件
Y
关闭
N
用户单击屏幕
用户双击屏幕

图 5-2-4 软件流程图

第六章 项目总结

1. 任务分解

程序制作：系统的硬件技术将随着客户需求不断升级。以现有的系统为基础，吸取经验教训推出更多的模块，带给盲人们更优质的阅读交互体验。结合最先进的技术和专业人才进行整个程序的制作。

系统制作与管理：架设专用服务器为客户端服务，进行系统管理，在各 APP 市场可提供下载。后期也将壮大专业的维护与更新，修复系统 bug，维护系统的运行，并提供客服方面的工作。

项目产品制作：计划利用小型摄像头搭载系统使用。利用小型摄像头拍照，将图片发送至系统实现图像识别、语音交互等功能。后期将对实际产品进行不断的升级与优化，提高盲人阅读体验，降低制作成本。

2. 克服困难

摄像头散热问题：摄像头在长时间使用时发热，画面失真，导致智能识别效果出现误差和错误，可以通过设计智能休眠摄像头的算法，减少摄像头长时间使用的情况。

模型识别问题：较厚的书籍页面文字弯曲导致图像识别不清，可以针对弯曲文字对识别模型加以训练或者采用弹性装夹的方式固定书籍。

图片分割问题：对书籍文字识别时，出现单双页文字分割不清的情况，通过建立新的图像分割模型智能识别单双页。

3. 项目特色总结

（1） 人文性：本项目以视障人群这一特定人群为切入，关注视力障碍的社会弱势群体，关注他们向往“正常”的心理需求和渴望自由学习、生活的学习需求，体现了浓浓的人文关怀。

（2） 可操作性与便捷性：本系统操作方便快捷，支持多语言文本识别，成本低，并且能装载在眼镜摄像头或手机上帮助盲人在工作与生活中解决阅读障碍，为视障人士提供了一种更加便利人性化的设计与选择。

（3） 创新性：相比以往研究更多注重于理论层面，本项目坚持理论与实践结合，所研究的系统采用基于机器学习和计算机视觉的文字图像识别系统来解决视障人群的问题，创新性地将智能语音识别技术、文字识别技术与 TTS 语言合成技术相结合，并创新性地通过数据

合成工具批量合成大量与目标场景类似的图像来扩大训练集以获取更高精度的模型

（4） 实用性：本系统的操作流程为由系统智能识别视障人士的语音输入命令，再由摄像头进行拍照通过图像识别技术识别出文字，最后由 TTS 技术放出语音辅助视力障碍人群的阅读与进行日常工作。

4. 已有成果

项目取得的主要成果如下表所示。

项目主要成果

	研究内容
模型	1. 文本检测模型 2. 文本识别模型 3. 文字转语音模型
系统	文字识别+语音合成系统
应用	移动端阅读软件和辅助设备

基于上述我们所获得的项目成果，我们在学科竞赛方面，参加武汉大学第七届“互联网+”大学生创新创业大赛获得三等奖，参加武汉大学首届青创训练营暨“百珈争创”创新创业模拟大赛获得二等奖；在科研成果方面，我们申请发明专利共 2 项，其中 2 项发明专利均已公开，有计算机软件著作权共 1 项（证明材料参见附录）。

（1） 移动端阅读软件

在初期工作阶段，我们设计出了能够让视障人士使用方便的移动端阅读软件，并且完成了文字识别模型和文字转语音模型的在移动端软件上的部署开发工作。而通过中期工作和后期工作，我们更进一步地优化了移动端阅读软件，提高了软件的识别精度及使用的流畅性。上述所提到移动端阅读软件的界面和操作方法如 2-5-1 和 2-5-2 图所示。

（2） 阅读软件辅助设备

在进行中期工作的过程中，我们发现如果视障人士仅仅使用外部摄像头进行拍照而没有两个比较固定的平面来摆放书籍和摄像头的话，非常容易出现因图像不全、图像弯曲、自动对焦失败而导致的文字识别不全、识别出错的情况。因此，在中期工作时，我们的团队设计了一款辅助设备帮助视障人士更好地使用我们的移动端阅读软件。而通过后期工作，如 6-4-2-1 和 6-4-2-2 所示，我们利用 3D 打印技术将该辅助设备实际制作出来，并配合我们开发的移动端阅读软件进行多次的使用测试，测试结果表明该辅助设备在一定程度上解决了我们之前遇到的问题。

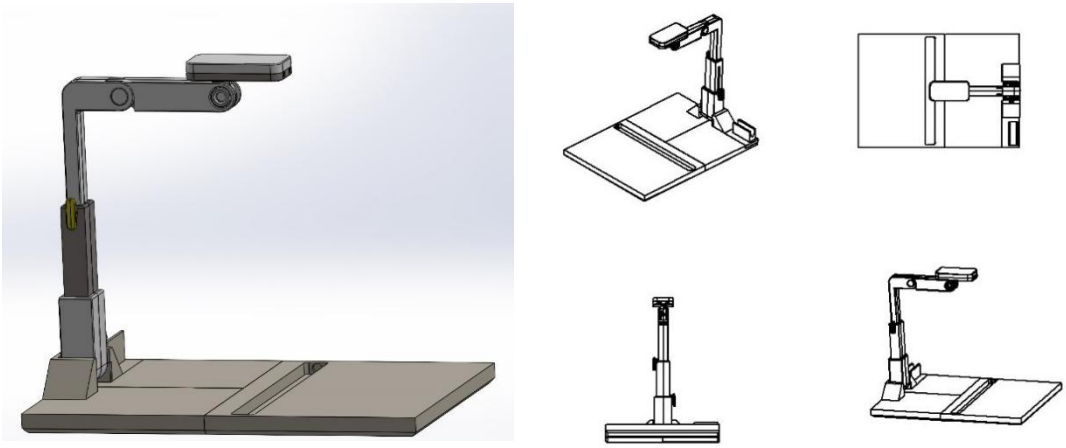

图 6-4-2-1 辅助设备设计图

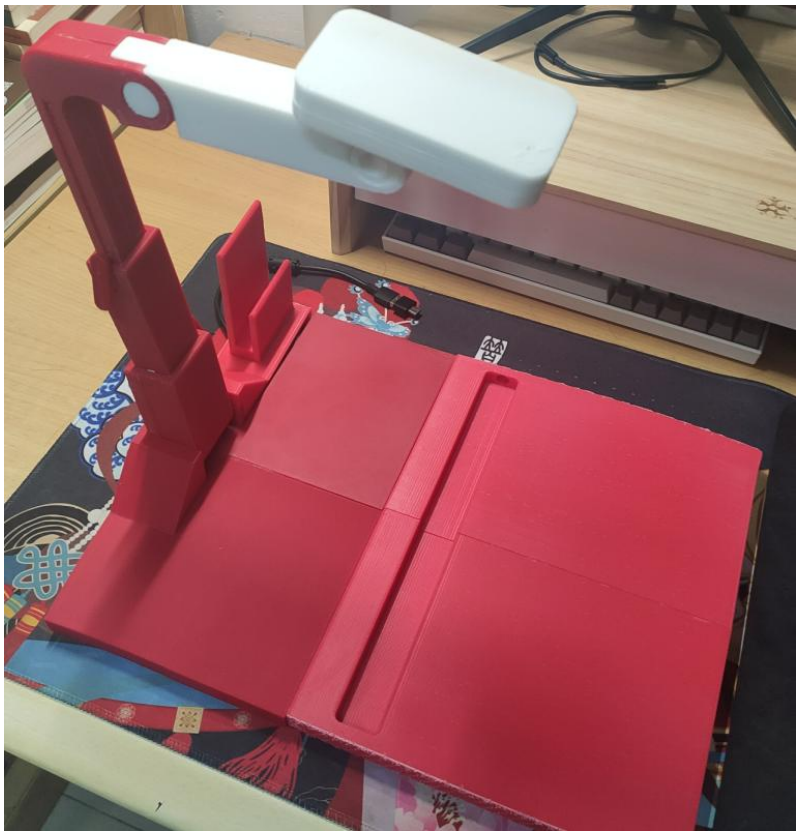

图 6-4-2-2 辅助设备成果图

如上图所示，摄像头用于对书籍成像；摄像头固定外壳和支撑架用来安装和支撑摄像头，并通过折叠结构调节摄像头的高度位置；底座上设置螺钉孔，可以直接放置在桌面上，也可以使用螺钉进行固定。

（3） 专利和软著

在科研成果方面，我们申请发明专利共 2 项，分别为《基于 OCR 识别方法的图书识别系统及设备》和《基于深度学习的中文纸质书籍阅读方法及设备方法及设备》，其中 2 项发明专利均已公开。

软件方面，在借鉴部分开源代码的基础上，自主开发了一款文字识别阅读软件，能够实现对书籍文字的自动采集、识别及阅读，满足盲人便捷阅读普通纸质书籍的需求，并取得一份计算机软件著作认证证书：《一种基于深度学习的纸质书籍文字识别阅读软件 V1.0》，登记号：2022SR0572960。

5. 商业推广

我们的产品是针对盲人自主阅读的配套设备，主要是为盲人以及他们的家人，朋友服务的。帮助盲人，快速将网络上的资料翻译成盲文，丰富教育阅读资源，改善盲人的教育条件。帮助盲人实时便利地阅读各类书籍，文章，新闻信息，增长知识。更广阔的远景是核心技术应用到盲人日常生活的各个方面，包括盲人医疗、盲人导航、盲人工作等等，全方位改善盲人的日常生活、学习、工作需求，提高视障人群的社会幸福感和获得感。

我们的产品导入方法如下：

（1） 提高产品质量、开辟新市场、改进广告内容。

（2） 改进和完善产品。突出产品特色，努力发展产品的新款式、新型号，增加产品的新用途。

（3） 树立产品形象。促销策略的重心应从建立产品知名度转移到树立产品形象，建立品牌偏好，争取新的客户。

（4） 选择适当的时机调整价格，以争取更多的消费者。

附录

国 家 知 识 产 权 局

430072

武汉市武昌区水果湖街中北路 86 号汉街武汉中央文化旅游区 K3 地块 1 栋（汉街总部国际 A 座）20 层 2 号 武汉科皓知识产权代理事务所（特殊普通合伙）
张辰(027-68776599)

发文日：

2022 年 10 月 17 日

申请号或专利号：202210755341.4　　发文序号：2022101200568250

申请人或专利权人： 武汉大学

发明创造名称： 基于 OCR 识别方法的图书识别系统及设备

发 明 专 利 申 请 公 布 及 进 入 实 质 审 查 阶 段 通 知 书

上述专利申请，经初步审查，符合专利法实施细则第 44 条的规定。根据专利法第 34 条的规定，该申请在 38 卷 4101 期 2022 年 10 月 11 日专利公报上予以公布。

根据申请人提出的实质审查请求，经审查，符合专利法第 35 条及实施细则第 96 条的规定，该专利申请进入实质审查阶段。

提示：

1. 根据专利法实施细则第 51 条第 1 款的规定，发明专利申请人自收到本通知书之日起 3 个月内，可以对发明专利申请主动提出修改。

2. 申请人可以访问国家知识产权局政府网站（www.cnipa.gov.cn），在专利检索栏目中查询公布文本。如果申请人需要纸件申请公布单行本的纸件，可向国家知识产权局请求获取。

3. 申请文件修改格式要求：

对权利要求修改的应当提交相应的权利要求替换项，涉及权利要求引用关系时，则需要将相应权项一起替换补正。如果申请人需要删除部分权项，申请人应该提交整理后连续编号的部分权利要求书。

对说明书修改的应当提交相应的说明书替换段，不得增加和删除段号，仅只能对有修改部分段进行整段替换。如果要增加内容，则只能增加在某一段中；如果需要删除一个整段内容，应该保留该段号，并在此段号后注明：“此段删除”字样。段号以国家知识产权局回传的或公布/授权公告的说明书段号为准。

对说明书附图、摘要、摘要附图修改的应当提交相应的说明书附图、摘要、摘要附图替换页。

同时，申请人应当在补正书或意见陈述书中标明修改涉及的权项、段号、页。

附录1 基于OCR识别方法的图书识别系统及设备发明专利申请进入实质审查通知书

国 家 知 识 产 权 局

430072

武汉市武昌区水果湖街中北路 86 号汉街武汉中央文化旅游区 K3 地块 1 栋（汉街总部国际 A 座）20 层 2 号 武汉科皓知识产权代理事务所（特殊普通合伙）

张辰(027-68776599)

发文日：

2022 年 10 月 17 日

申请号或专利号：202210741094.2　　发文序号：2022101200641000

申请人或专利权人： 武汉大学

发明创造名称： 基于深度学习的中文纸质书籍阅读方法及设备方法及设备

发 明 专 利 申 请 公 布 及 进 入 实 质 审 查 阶 段 通 知 书

上述专利申请，经初步审查，符合专利法实施细则第 44 条的规定。根据专利法第 34 条的规定，该申请在 38 卷 4101 期 2022 年 10 月 11 日专利公报上予以公布。

根据申请人提出的实质审查请求，经审查，符合专利法第 35 条及实施细则第 96 条的规定，该专利申请进入实质审查阶段。

提示：

1. 根据专利法实施细则第 51 条第 1 款的规定，发明专利申请人自收到本通知书之日起 3 个月内，可以对发明专利申请主动提出修改。

2. 申请人可以访问国家知识产权局政府网站（www.cnipa.gov.cn）,在专利检索栏目中查询公布文本。如果申请人需要纸件申请公布单行本的纸件，可向国家知识产权局请求获取。

3. 申请文件修改格式要求：

对权利要求修改的应当提交相应的权利要求替换项，涉及权利要求引用关系时，则需要将相应权项一起替换补正。如果申请人需要删除部分权项，申请人应该提交整理后连续编号的部分权利要求书。

对说明书修改的应当提交相应的说明书替换段，不得增加和删除段号，仅只能对有修改部分段进行整段替换。如果要增加内容，则只能增加在某一段中；如果需要删除一个整段内容，应该保留该段号，并在此段号后注明：“此段删除”字样。段号以国家知识产权局回传的或公布/授权公告的说明书段号为准。

对说明书附图、摘要、摘要附图修改的应当提交相应的说明书附图、摘要、摘要附图替换页。

同时，申请人应当在补正书或意见陈述书中标明修改涉及的权项、段号、页。

附录2 基于深度学习的中文纸质书籍阅读方法及设备方法及设备发明专利进入实质审查通知书

中华人民共和国国家版权局

计算机软件著作权登记证书

证书号：软著登字第9527159号

软件名称：一种基于深度学习的纸质书籍文字识别阅读软件V1.0

著作权人：武汉大学

开发完成日期：2020年03月01日

首次发表日期：未发表

权利取得方式：原始取得

权利范围：全部权利

登记号：2022SR0572960

根据《计算机软件保护条例》和《计算机软件著作权登记办法》的规定，经中国版权保护中心审核，对以上事项予以登记。

中华人民共和国国家版权局 计算机软件著作权登记专用章

No. 10692118

2022年05月11日

附录3 一种基于深度学习的纸质书籍文字识别阅读软件著作权登记证书

参考文献

[1] Luo C, Jin L, Sun Z. MORAN: A Multi-Object Rectified Attention Network for scene text recognition[J]. Pattern Recognition: The Journal of the Pattern Recognition Society, 2019, 90: 109-118.

[2] Ping W, Peng K, Gibiansky A, et al. Deep Voice 3: Scaling Text-to-Speech with Convolutional Sequence Learning[C], 2017.

[3] Du Y, Li C, Guo R, et al. PP-OCRv2: Bag of Tricks for Ultra Lightweight OCR System[J], 2021

[4] Lu H, King S, Watts O. Combining a Vector Space Representation of Linguistic Context with a Deep Neural Network for Text-To-Speech Synthesis[C], 2013.

[5] Kang S, Qian X, Meng H. Multi-distribution deep belief network for speech synthesis[C]. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 2013.

[6] Wang Y, Skerry-Ryan R J, Stanton D, et al. Tacotron: Towards End-to-End Speech Synthesis[C]. Interspeech 2017, 2017.

[7] Arik S O, Chrzanowski M, Coates A, et al. Deep Voice: Real-time Neural Text-to-Speech[C], 2017.

[8] Luo C, Jin L, Sun Z. MORAN: A Multi-Object Rectified Attention Network for scene text recognition[J]. Pattern Recognition: The Journal of the Pattern Recognition Society, 2019, 90: 109-118.

[9] Radwan M A, Khalil M I, Abbas H M. Neural Networks Pipeline for Offline Machine Printed Arabic OCR[J]. Neural processing letters, 2018, 48(2): 769-787.

[10] 张雪梅, 高凯. 深入浅出Android软件开发教程[M]. 深入浅出Android软件开发教程, 2015.

[11] Lu H, King S, Watts O. Combining a Vector Space Representation of Linguistic Context with a Deep Neural Network for Text-To-Speech Synthesis[C], 2013.

[12] Luo C, Jin L, Sun Z. MORAN: A Multi-Object Rectified Attention Network for scene text recognition[J]. Pattern Recognition: The Journal of the Pattern Recognition Society, 2019, 90: 109-118.

[13] Radwan M A, Khalil M I, Abbas H M. Neural Networks Pipeline for Offline Machine Printed Arabic OCR[J]. Neural processing letters, 2018, 48(2): 769-787.

[14] 纪名刚 濮. 机械设计(第8版)[M]. 机械设计(第8版), 2006.

[15] 孙桓, 陈作模, 葛文杰. 机械原理.第7版[M]. 机械原理.第7版, 2006.

[16] 陈载赋. 结构力学简明手册[M]. 结构力学简明手册, 1986.

[17] 朱瑛 张. 机械制造工艺学(第2版)[M]. 机械制造工艺学(第2版), 2009.

[18] 文小燕. 机械制造及自动化中的3D打印技术[J]. 自动化与仪器仪表, 2018(8): 4.

[19] 曾妍, 闫大鹏. 基于直角坐标系3D打印的机械结构设计分析[J]. 科技视界, 2016(12): 1.