

# 1. 一种基于目标检测和OCR技术的地铁标识识别方法

申请号

CN202210519524

公开（公告）日

2022. 07. 29

申请（专利权）人

浙江大学

发明人

卜佳俊；陈锶皓；于智；李靖儒；李亮城；戚忠达；邵子睿；林帅浩；谷雨

摘要

– ABSTRACT：本发明公开了一种基于目标检测和OCR技术的地铁标识识别方法，通过计算机视觉技术与目标检测技术，定位地铁场景图中的地铁导引标识的坐标；利用各标识的坐标，使用目标检测与分类技术，提取地铁标识中的图像信息；使用文本检测和文本识别技术，提取地铁标识中的文本信息；计算两种信息的联合置信度，筛选有效信息并输出。本发明为地铁标识识别领域提供了相关技术，可用于为视障人群提供导航信息，帮助其安全、高效地完成地铁行程，同时也为设计地铁导航系统的人员提供更科学的设计建议。

权利要求

1. 一种基于目标检测和OCR技术的地铁标识识别方法，其特征在于，包括以下步骤：

S110：通过网络协议和小程序前端框架，获取地铁场景图片；

S120：利用目标检测和分类技术，得到地铁场景图片中的地铁导引牌标识坐标；

S130：利用目标检测、分类技术和导引牌标识坐标，提取标识所包含的图标信息；

S140：利用OCR技术和导引牌标识坐标，提取标识所包含的文本块坐标和对应文本信息；

S150：将图标信息和文本信息进行融合，计算联合置信度；设置联合置信度阈值，筛选有效输出；

S160：向前端框架返回，所识别的可读的图标信息和文本信息序列。

2. 如权利要求1所述基于目标检测和OCR技术的地铁标识识别方法，其特征在于，步骤S110包括：通过Websokcet协议搭建长连接，将获取的图片通过Base64转码传输至服务端，完成解码操作恢复原图片，并使用高斯滤波降噪和均值平滑方式进行预处理。

3. 如权利要求1所述基于目标检测和OCR技术的地铁标识识别方法，其特征在于，步骤S120包括：收集包含地铁导引牌标识信息的地铁场景图，制作数据集并完成训练，得到能够提取地铁导引牌标识信息的Yolov5预训练模型 $f_{direct}$ ；对每一张地铁场景图，使用模型 $f_{direct}$ 获取地铁场景中地铁标识的位置，并且切割出相应的导引牌标识子图。

4. 如权利要求3所述基于目标检测和OCR技术的地铁标识识别方法，其特征在于，步骤S120中，设置置信度阈值，对模型 $f_{direct}$ 的预测结果，筛选置信度满足要求的预测框。
5. 如权利要求1所述基于目标检测和OCR技术的地铁标识识别方法，其特征在于，步骤S130包括：收集包含地铁图标信息的地铁场景图，制作数据集并完成训练，得到能够提取图标信息的Yolov5预训练模型 $f_{icon}$ ；对每一张导引牌标识子图，使用模型 $f_{icon}$ 获取标志预测框坐标信息、图标类别信息与预测框置信度信息，并放入集合 $L_{icon}$ 中。
6. 如权利要求5所述基于目标检测和OCR技术的地铁标识识别方法，其特征在于，步骤S130中，设置置信度阈值，对模型 $f_{icon}$ 的预测结果，筛选置信度满足要求的预测框。
7. 如权利要求1所述基于目标检测和OCR技术的地铁标识识别方法，其特征在于，步骤S140包括：预先训练文本检测预训练模型，并收集包含文本信息的地铁场景图片，制作成数据集，在训练好的预训练模型基础上，完成DBnet文本检测模型训练；通过DBNet定位导引牌标识子图中文本信息的位置，利用CRNN文本识别方法，识别文字块子图的文字内容及对应的置信度；使用基于字典和模糊匹配的纠错方法，对文本内容进行纠错，并和对应的置信度一起最终放入集合 $L_{text}$ 中。
8. 如权利要求1所述基于目标检测和OCR技术的地铁标识识别方法，其特征在于，步骤S140中，所述基于字典和模糊匹配的纠错方法，包括：
  - (1) 构建关键字纠错字典，关键词包括地铁站名、设施、出口、相关固定术语；
  - (2) 基于模糊匹配的方式，对置信度低于预设阈值的弱识别文本进行纠错操作，以关键词的形式与字典中的文本进行匹配，如果匹配率满足要求，即匹配成功完成纠错。
9. 如权利要求1所述基于目标检测和OCR技术的地铁标识识别方法，其特征在于，步骤S150包括：对于每一张导引牌标识子图 $I_x \in \beta$ ，共有a条图标信息，b条文本信息，采用加权平均的方式计算联合置信度 设置初始联合置信度阈值为 $t_{last}$ ，当前遍历到的子图联合置信度小于阈值 $t_{last}$ 时，抑制其识别的图标和文本信息出现在最终构建的文本输出内容中。
10. 如权利要求1所述基于目标检测和OCR技术的地铁标识识别方法，其特征在于，步骤S160包括：将经过筛选的图标信息集合 $L_{icon}$ 与文本信息集合 $L_{text}$ ，按照其坐标进行排序，组合为 $L_{result}$ ；通过已经建立的长连接，将消息序列发送给框架，并转换成对应的提示信息展示给用户。

## 说明书

一种基于目标检测和OCR技术的地铁标识识别方法  
技术领域

本发明属于计算机视觉与人工智能算法领域，尤其涉及一种基于目标检测和OCR(Optical Character Recognition, 光学字符识别)技术的地铁标识识别方法。

背景技术

地铁标识系统是地铁站内方向导航的关键信息载体，引导乘客安全地完成地铁行程，承载着重要的公共服务职能。在紧急时，地铁标识能够清晰有效地帮助乘客快速达到目标地点。

识别地铁标识中的图标信息和文本信息，并将之转化为可读的文字序列，有利于向视障人群传达地铁标识中所包含的导航信息，从而可以根据导航信息做出方向判断，到达最终目标地点。

#### 发明内容

本发明的目的在于针对现有技术中缺乏地铁标识识别相关技术的不足，提供一种基于目标检测和OCR技术的地铁标识识别方法。本发明方法以微信小程序为载体。

本发明的目的是通过以下技术方案来实现的：一种基于目标检测和OCR技术的地铁标识识别方法，包括：

S110：通过网络协议和小程序前端框架，获取用户所拍摄的地铁场景图片。

S120：利用目标检测和分类技术，得到地铁场景图片中的地铁导引牌标识坐标。

S130：利用目标检测、分类技术和导引牌标识坐标提取标识所包含的图标信息。

S140：利用OCR技术和导引牌标识坐标提取标识所包含的文本块坐标和对应文本信息。

S150：将图标信息和文本信息进行融合，计算联合置信度。设置联合置信度阈值，筛选有效输出。

S160：向前端框架返回所识别的可读的图标信息和文本信息序列。

进一步地，步骤S110中，通过网络协议和小程序前端框架，获取用户所拍摄的地铁场景图片，具体包括：

S1101：通过Websocket协议和微信小程序的前端框架，建立用户端与服务端之间的稳定长连接。

S1102：在前端框架建立获取图片的接口，等待用户输入场景图像后，将图像转为Base64格式字符串，并发送至服务端。

S1103：在服务端对所接受字符串进行解码，还原目标图像，并对其采用高斯滤波降噪和均值平滑方式进行预处理。

进一步地，步骤S120中，利用目标检测和分类技术，得到地铁场景图片中的地铁导引牌标识坐标，具体包括：

S1201：收集大量包含地铁导引牌标识的地铁场景图，制作数据集并完成训练，得到能够提取导引牌标识分类与坐标信息的Yolov5预训练模型 $f_{\text{direct}}$ 。

S1202：使用模型 $f_{\text{direct}}$ 对待检测图像进行目标定位、分类与回归，并根据置信度筛选有效的预测框，得到场景图片中 $n$ 个导引牌标识的坐标集合 $\theta = \{(x_1, y_1, w_1, h_1) \cdots (x_n, y_n, w_n, h_n)\}$ 。

S1202：根据导引牌标识的坐标集合 $\theta$ ，切割出各导引牌对应的子图集合 $\beta = \{I_1, \cdots, I_n\}$ 。

进一步地，步骤S130中，利用目标检测、分类技术和导引牌标识坐标提取标识所包含的图标信息，具体包括：

S1301：收集大量包含地铁图标信息的地铁场景图，制作数据集并完成训练，

得到能够提取图标信息的Yolov5预训练模型 $f_{icon}$ 。

S1302: 获取导引牌标识的子图集合 $\beta$ ，使用模型 $f_{icon}$ 对每一张子图 $I_x \in \beta$ ，进行目标定位、分类cls与回归。根据置信度cfd筛选预测框，得到子图中a条图标信息，其中包括预测框坐标信息、图标类别信息与预测框置信度信息，将其重新组合后放入集合中，并将所有子图所生成的图标信息放入集合中。

进一步地，步骤S140中，利用OCR技术和导引牌标识坐标提取标识所包含的文本块坐标和对应文本信息，具体包括：

S1401: 在大规模公开icdar2013数据集上训练文本检测预训练模型，并收集大量包含文本信息的地铁场景图片制作成数据集，在预训练模型基础上完成DBnet文本检测模型训练。

S1402: 获取导引牌标识的子图集合 $\beta$ ，使用DBnet对每一张子图 $I_x \in \beta$ 进行文本检测，得到子图中b个文本块的坐标集合

S1403: 根据文本块坐标集合切出各文本块对应的子图集合

S1404: 使用CRNN模型对切割的文本块子图进行文本识别，得到b个文本块对应的文本内容 $text$ 与置信度

S1405: 使用一种基于字典和模糊匹配的纠错方法，对文本内容 $text$ 进行纠错。

S1406: 将每个子图纠错后的文本内容与置信度放入集合

进一步地，步骤S1405中，一种基于字典和模糊匹配的纠错方法，具体包括：

(1)收集常见地铁站名、设施、出口、相关固定术语等，形成一个关键字纠错字典。

(2)基于模糊匹配的方式，对置信度低于预设阈值的弱识别文本进行纠错操作，以关键词的形式与字典中的文本进行匹配，如果匹配率大于预设阈值，即匹配成功完成纠错。

进一步地，步骤S150中，将地铁标识的图标信息和文本信息进行融合，计算联合置信度，具体包括：

S1501: 获得图标信息集合 $L_{icon}$ 、文本信息集合 $L_{text}$ 与子图集合 $\beta$ ，对于每一张子图 $I_x \in \beta$ ，共有a条图标信息，b条文本信息，根据加权平均的方式计算联合置信度 $p_x$ ，即

S1502: 根据初期的一些实验结果，设置初始联合置信度阈值为 $t_{last}$ 。

S1503: 在构建子图 $I_x \in \beta$ 中的图标信息和文本信息输出时，根据阈值 $t_{last}$ 筛选信息有效性，即当前遍历到的子图联合置信度小于阈值时，抑制其识别的图标和文本信息出现在最终构建的文本输出内容中。

进一步地，步骤S160中，向前端框架返回所识别的可读的图标信息和文本信息序列，具体包括：

S1601: 将经过阈值筛选的图标信息集合 $L_{icon}$ 与文本信息集合 $L_{text}$ ，按照其坐标大小关系(从左到右，从上到下)进行排序，组合为 $L_{result}$ 。

S1602: 将结果 $L_{result}$ 通过Websocket长连接发送回前端框架，前端框架接受信息序列后，转化为语音输出并被客户群体接收。

本发明的有益效果是：本发明方法通过计算机视觉技术与目标检测技术，定

位地铁场景图中的地铁导引标识的坐标；利用各标识的坐标，使用目标检测与分类技术，提取地铁标识中的图像信息；使用文本检测和文本识别技术，提取地铁标识中的文本信息；计算两种信息的联合置信度，筛选有效信息并输出。本发明为地铁标识识别领域提供了相关技术，可用于为视障人群提供导航信息，帮助其安全、高效地完成地铁行程，同时也为设计地铁导航系统的人员提供设计建议。

#### 附图说明

图1是本发明基于目标检测和OCR技术的地铁标识识别方法的流程图。

#### 具体实施方式

下面结合附图和实施例对本发明进一步说明。这些附图均为简化的示意图，仅以示意方式说明本发明的基本结构，因此其仅显示与本发明有关的构成。

本发明一种基于目标检测和OCR技术的地铁标识识别方法，针对地铁标识的图像信息和文本信息提取问题，提出一种解决方案。

具体地，如图1所示，本发明一种实施例包括：

S110：通过网络协议和小程序前端框架，获取用户所拍摄的地铁场景图片。包括：

S1101：通过Websocket协议和微信小程序的前端框架，建立用户端与服务端之间的稳定长连接。

S1102：在前端框架建立获取图片的接口，等待用户输入场景图像后，将图像转为Base64格式字符串，并发送至服务端。

S1103：在服务端对所接受字符串进行解码，还原目标图像，并对其采用高斯滤波降噪和均值平滑方式进行预处理。

S120：利用目标检测和分类技术，得到地铁场景图片中的地铁导引牌标识坐标，方便后续操作。包括：

S1201：收集大量包含地铁导引牌标识的地铁场景图，制作数据集并完成训练，得到能够提取导引牌标识分类与坐标信息的Yolov5预训练模型 $f_{direct}$ 。

S1202：使用模型 $f_{dir}$ 对待检测图像进行目标定位、分类与回归，并根据置信度筛选有效的预测框，得到场景图片中 $n$ 个导引牌标识的坐标集合 $\theta = \{(x_1, y_1, w_1, h_1) \cdots (x_n, y_n, w_n, h_n)\}$ 。 $x_x, y_x$ 表示导引牌标识 $x$ 预测框左上角坐标， $w_x, h_x$ 表示导引牌标识 $x$ 预测框的宽、高。

S1202：根据导引牌标识的坐标集合 $\theta$ ，切割出各导引牌对应的子图集合 $\beta = \{I_1, \cdots, I_n\}$ 。 $I_x$ 表示导引牌 $x$ 的子图。

S130：利用目标检测、分类技术和导引牌标识坐标，提取标识所包含的图标信息。包括：

S1301：收集大量包含地铁图标信息的地铁场景图，制作数据集并完成训练，得到能够提取图标信息的Yolov5预训练模型 $f_{icon}$ 。

S1302：获取导引牌标识的子图集合 $\beta$ ，使用模型 $f_{icon}$ 对每一张子图 $I_x \in \beta$ ，进行目标定位、分类 $cls$ 与回归。根据预设置信度 $cfd$ 筛选预测框，得到子图中 $a$ 条图标信息；其中，每条图标信息包括预测框坐标信息、图标类别信息 $cls$ 与预测框置信度信息 $cfd$ ，将其重新组合后放入子图 $I_x$ 的图标信息集合中，并将所有子图所生成的图标信息放入集合中。预测框坐标信息包括左上角坐标 $x_i, y_i$ 和宽高 $w_i, h_i$ ， $i=1 \sim a$ 。

S140: 利用OCR技术和导引牌标识坐标, 提取标识所包含的文本块坐标和对应文本信息。包括:

S1401: 在大规模公开icdar2013数据集上, 训练文本检测预训练模型, 并收集大量包含文本信息的地铁场景图片制作成数据集, 在预训练模型基础上完成DBnet文本检测模型训练。

S1402: 获取导引牌标识的子图集合 $\beta$ , 使用DBnet对每一张子图 $I_x \in \beta$ 进行文本检测, 得到子图 $I_x$ 中b个文本块的坐标集合 文本块坐标信息包括预测框左上角坐标x, y和宽高w, h。

S1403: 根据文本块坐标集合 切出各文本块对应的子图集合 表示文本块的子图,  $j=1 \sim b$ 。

S1404: 使用CRNN模型对切割的文本块子图 进行文本识别, 得到b个文本块对应的文本内容 与置信度

S1405: 使用一种基于字典和模糊匹配的纠错方法, 对文本内容 进行纠错。包括:

(1)收集常见地铁站名、设施、出口、相关固定术语等, 形成一个关键字纠错字典。

(2)基于模糊匹配的方式, 对置信度低于0.8的弱识别文本进行纠错操作, 以关键词的形式与字典中的文本进行匹配, 如果匹配率大于90%, 即匹配成功完成纠错。

S1406: 纠错后剩余m个文本内容, 将每个子图 的文本内容与置信度放入集合

S150: 将图标信息和文本信息进行融合, 计算联合置信度。设置联合置信度阈值, 筛选有效输出。包括:

S1501: 获得图标信息集合 $L_{icon}$ 、文本信息集合 $L_{text}$ 与子图集合 $\beta$ , 对于每一张子图 $I_x \in \beta$ , 共有a条图标信息, b条文本信息, 根据加权平均的方式计算联合置信度 $p_x$ , 即

S1502: 设置初始联合置信度阈值 $t_{last}$ 为0.85。

S1503: 在构建子图 $I_x \in \beta$ 中的图标信息和文本信息输出时, 根据阈值 $t_{last}$ 筛选信息有效性, 即当前遍历到的子图联合置信度小于阈值时, 抑制其识别的图标和文本信息出现在最终构建的文本输出内容中。

S160: 向前端框架返回所识别的可读的图标信息和文本信息序列。包括:

S1601: 将经过阈值筛选的图标信息集合 $L_{icon}$ 与文本信息集合 $L_{text}$ , 按照其坐标大小关系(从左到右, 从上到下)进行排序, 组合为 $L_{result}$ 。

S1602: 将结果 $L_{result}$ 通过Websocket长连接发送回前端框架, 前端框架接收信息序列后, 转化为语音输出, 最终被客户群体以听觉信息接收(语音提示)。

综上, 本发明一种基于目标检测和OCR技术的地铁标识识别方法, 通过计算机视觉技术与目标检测技术, 定位地铁场景图中的地铁导引标识的坐标; 利用各标识的坐标, 使用目标检测与分类技术, 提取地铁标识中的图像信息; 使用文本检测和文本识别技术, 提取地铁标识中的文本信息; 计算两种信息的

联合置信度，筛选有效信息并输出。本发明为地铁标识识别领域提供了相关技术，可用于为视障人群提供导航信息，帮助其安全、高效地完成地铁行程，同时也为设计地铁导航系统的人员提供更科学的设计建议。

本发明并不限于上述实施方式，采用与本发明上述实施方式相同或近似的方式，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，均在本发明专利的保护范围之内。