Translate books for the visually impaired

Xiang Wang
Department of Artificial intelligence
Peking University
2100013146@stu.pku.edu.cn

Xin Hao
Department of Artificial intelligence
Peking University
2100013152@stu.pku.edu.cn

Abstract

The growing spiritual and cultural requirements the blind need to be met. However, barrier-free cultural products take a long time to prepare and are concentrated in more developed cities, they cannot be widespread to those in need. Here we show the pipeline to translate books for the visually impaired mainly with Optical Character Recognition(OCR) and Image Caption(IC) technique. Our work demonstrates how to solve this problem and our result can build the foundation of this task. We anticipate our pipeline to be a starting point of research in translating books for the visually impaired and we hope we can bring light to the blind.

1 Introduction

It is consists of four parts: OCR, Image Extractor, Text Typeset and Image Caption. Before we feed inputs into the pipeline, we need to preprocess PDF file, turning every page of the PDF file into an image with the same size[1]. We call these images as page-images. OCR module[2] takes in page-images, and output text boxes, which contains characters and corresponding positions(coordinates). Image Extractor also takes in page-images, finds out where the pictures are and cuts out pictures from origin pages. With text boxes and positions of pictures, we can typeset the characters into a half-completed book with picture labels in proper place. Image Caption module gets pictures from Image Extractor and output descriptions of these pictures. Finally, we replace the picture labels with picture descriptions one by one, and output the translated book as .txt file. Fig. 1 demonstrates the whole pipeline.

2 Image Extractor

This module need to frame every picture in a page with a rectangular area, and for a single page, it should output pictures separately and accurately, so that the Image Caption module can produce description for every pictures. Of course we can handle this problem with deep learning, but traditional method can work good enough.

2.1 Basic idea

2.2 Some details

mask, gray-scale, stride, rectangular

---

[1]Actually the same height, because the aspect ratio of one page varies from different books.

[2]Considering OCR technology is relatively mature, we choose to use existing OCR module. After trying different OCR modules, we choose Paddle OCR because its API is easy to use.
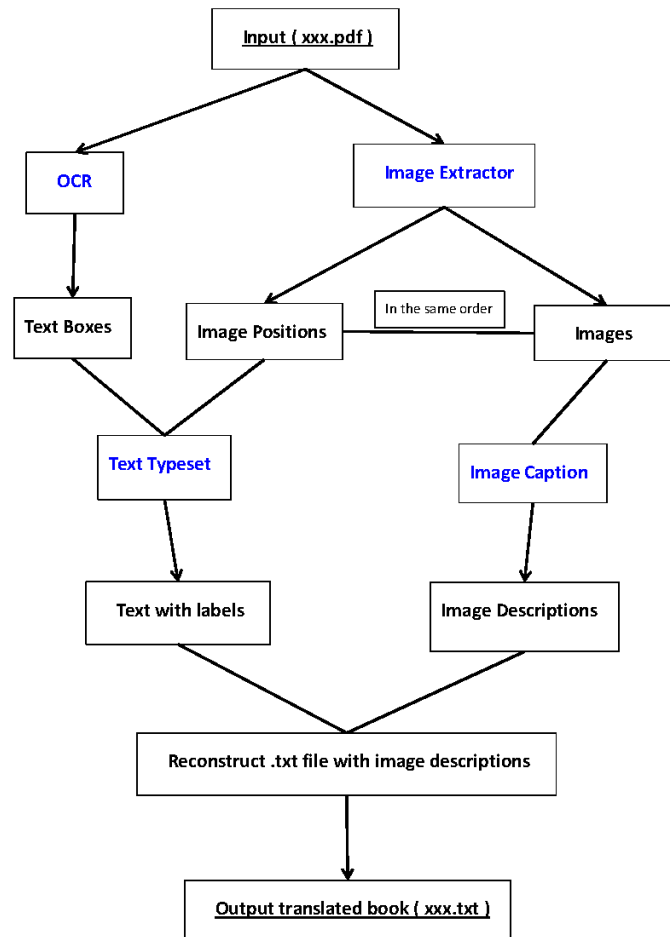
Figure 1: Pipeline

3 Text Typeset

4 Image Caption

5 Analyse

A Appendix