SEGMENT THE DOCUMENT IMAGE TO EXTRACT TEXT AND CONVERT IT TO BRAILLE

This project focuses on the recognition and conversion of Vietnamese text into Braille, a tactile writing system designed for the visually impaired. In the image preprocessing stage, methods like edge detection or segmentation using the K-means algorithm are employed to obtain an image containing only the document page, effectively eliminating all unnecessary background components. After the image preprocessing stage, a pre-trained YOLOv8 model is trained on a dataset collected from textbooks to perform tasks such as text and image segmentation in the images. Subsequently, Tesseract OCR is used to recognize characters in the images. Once the identified text is obtained, the system proceeds to convert it into Braille. To optimize convenience and accessibility, the product is deployed online through an API built with Flask.

AUTHORS

Thai Thi Kim Yen - 47.01.104.250

Hoang Thuy Quynh Huong - 47.01.104.096

LECTURER: DR. NGO QUOC VIET

Source code: https://github.com/Sherlockian1212/Document2Braille

INTRODUCTION

In today's tech-driven era, advancements are continually tackling communication and information accessibility challenges for the visually impaired.

The project's goal is a comprehensive system enabling visually impaired individuals to access and read text from images efficiently through Braille conversion, expanding information and education access. Beyond technical aspects, practical applications are integrated, and web deployment opens avenues for broader applications in education, offices, research, and entertainment.

MATERIALS & METHOD

Image Pre-processing

Input data consists of images containing text pages, along with other elements in the background, posing a challenge in text recognition.

Two methods are proposed to carry out this task, including edge detection and the K-means algorithm for image clustering. Prior to implementing these methods, a preprocessing step is conducted by applying a morphological close operation to eliminate unimportant details in the image, retaining only the essential and accurate contours.
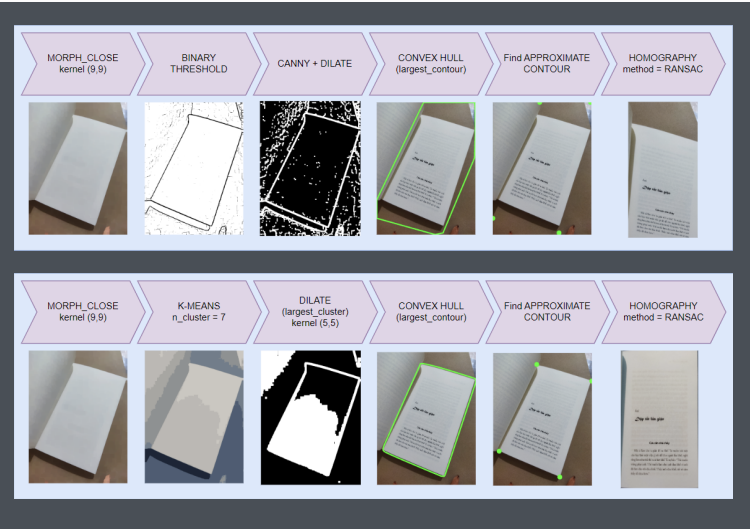

Document Image Segmentation

YOLO is a high-accuracy, fast-speed model that utilizes the pre-trained YOLOv8 model. It has been retrained on a dataset collected from 9th-grade Mathematics and Biology textbooks, which was labeled by Hà, Vy, and Yến.

This dataset comprises a total of 2149 images and has undergone a series of preprocessing steps. In this process, we applied the following steps: Preprocessing (Auto-Orient, Resize stretch to 640x640), Augmentations (Crop, Rotation, Shear, Grayscale, Cutout, Mosaic, Brightness, Blur). These steps aim to optimize the input data for the model training process. Dataset split: 88% for the training set, 8% for the validation set, and 4% for the test set.
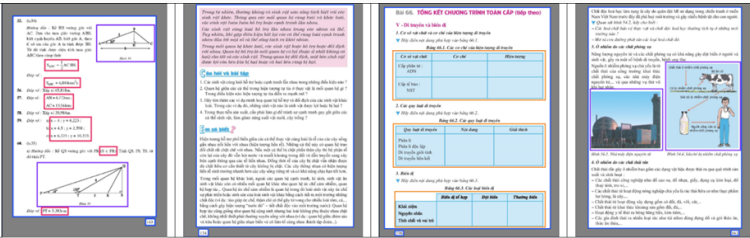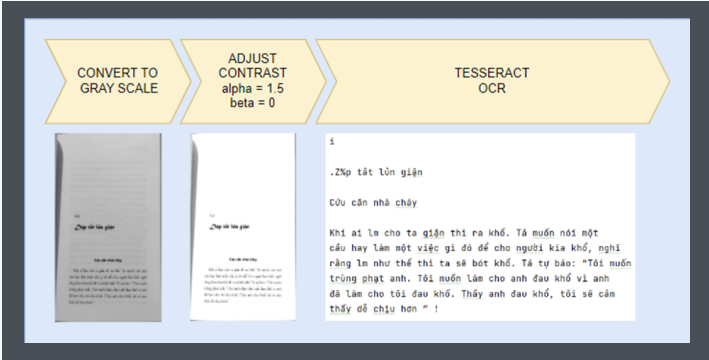

Model training and evaluation

The pre-train model is yolov8s-seg.pt, trained for 59 epochs with a batch size of 16 and an image size of 640.
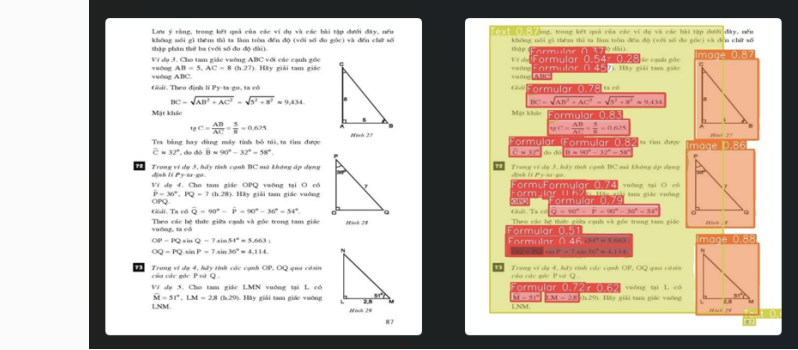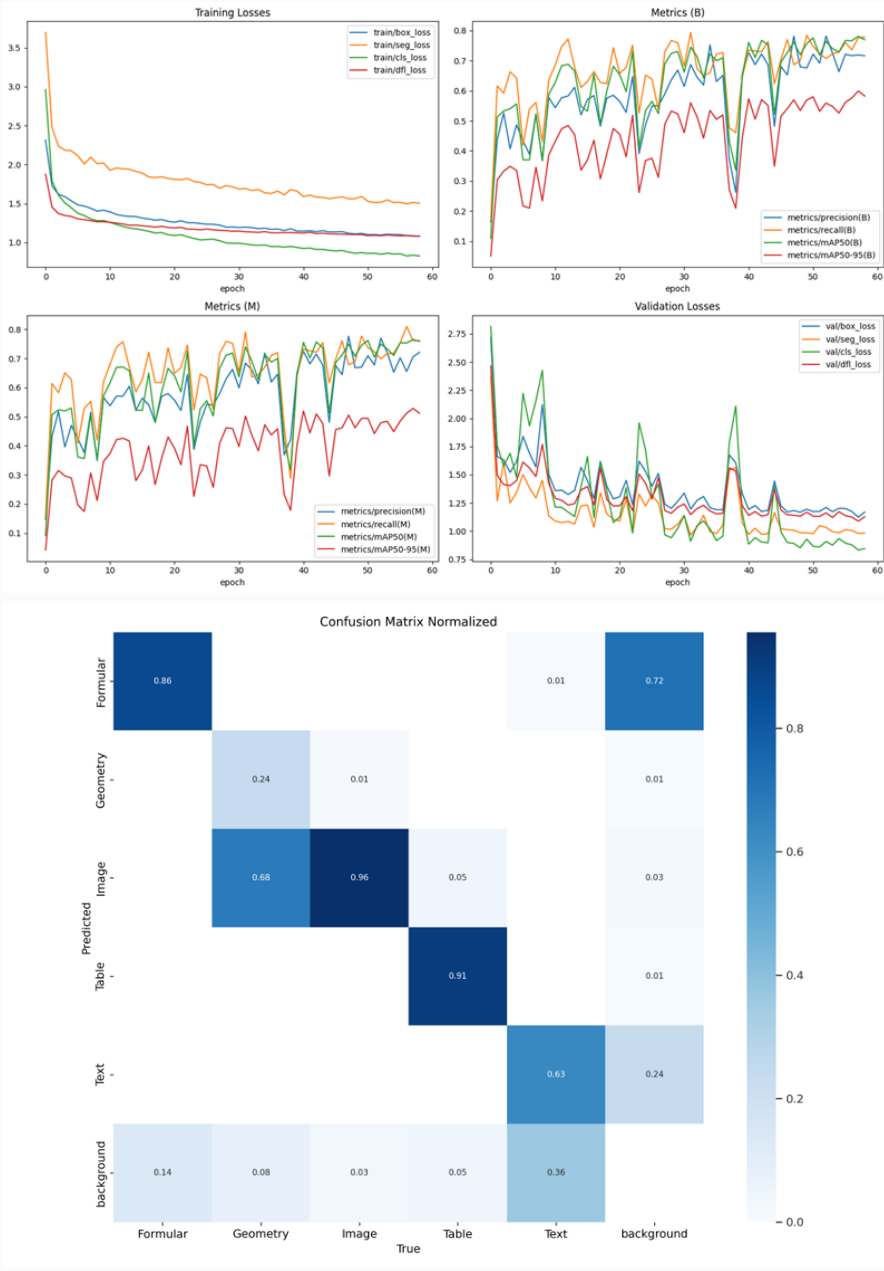

Text Extraction

Tesseract is an open-source OCR tool. Its capabilities are primarily limited to structured textual data. It performs quite poorly when dealing with unstructured text containing various noisy elements.

Conversion to Braille and web deployment

The Ministry of Education and Training has issued a Circular on "Ban hành Quy định chuẩn quốc gia về chữ nổi Braille cho người khuyết tật". It clearly outlines the principles and rules for writing Braille in the Vietnamese language. The provisions in the Circular are applied to convert the recognized and processed Vietnamese characters in the previous steps.
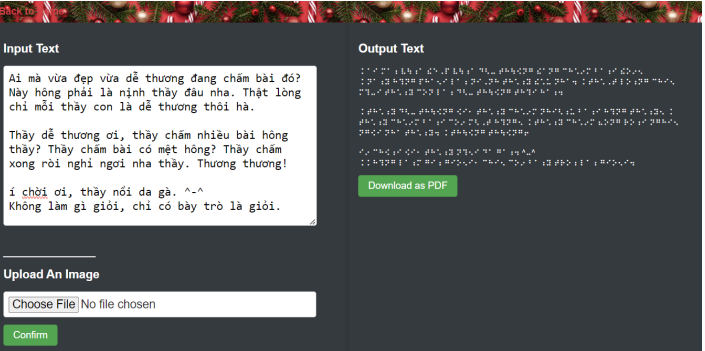
RESULTS

TThe project results include the implementation of two image preprocessing methods using edge detection and K-means. Specifically, the K-means method demonstrates better performance for images with a distinct background from the book pages, while edge detection yields better results for cases where the background have colors similar to the book pages.

Collect data, perform labeling and train the model using a pre-trained YOLOv8 model, delivering stable and actionable results. Image and Geometry labels are the two labels that are most confused with each other, in addition, other labels have quite high accuracy.

Using the Tesseract tool for comprehensive, fast, and accurate text recognition in images. Subsequently, we convert the results into Braille text and deploy an API to provide interactive capabilities with the data.

REFERENCES

Vietnamese
[1] Ngô Quốc Việt (2017). Xử lý ảnh số. NXB ĐHSP Tp.HCM.
[2] Phan Đức Chính, Tôn Thân, Vũ Hữu Bình, Trần Phương Dung, Ngô Hữu Dũng, Lê Văn Hồng, Nguyễn Hữu Thảo. Sách giáo khoa Toán lớp 9 Tập 1. Nhà xuất bản Giáo Dục Việt Nam.
[3] Phan Đức Chính, Tôn Thân, Nguyễn Huy Đoan, Phạm Gia Đức, Trương Công Thành, Nguyễn Duy Thuận. Sách giáo khoa Toán lớp 9 Tập 2. Nhà xuất bản Giáo Dục Việt Nam.
[4] Tôn Thân, Vũ Hữu Bình, Trần Phương Dung, Lê Văn Hồng, Nguyễn Hữu Thảo. Bài tập Toán 9 Tập 1. Nhà xuất bản Giáo Dục Việt Nam.
[5] Tôn Thân, Phạm Gia Đức, Trương Công Thành, Nguyễn Duy Thuận. Sách giáo khoa Toán lớp 9 Tập 2. Nhà xuất bản Giáo Dục Việt Nam.
[6] Nguyễn Quang Vinh, Vũ Đức Lưu, Nguyễn Minh Công, Mai Sỹ Tuấn. Sách giáo khoa sinh học 9. Nhà xuất bản Giáo Dục Việt Nam.
[7] Bộ Giáo dục và Đào tạo. (2019). Thông tư số 15/2019/TT-BGDĐT ban hành Quy định chuẩn quốc gia về chữ nổi Braille cho người khuyết tật.

English
[1] Ngo Quoc Viet. Digital Image Processing lecture slides.
[2] Ngo Quoc Viet. Machine Learning lecture slides.
[3] S.V. Rice, F.R. Jenkins, T.A. Nartker. (1995). The Fourth Annual Test of OCR Accuracy, Technical Report 95-03, Information Science Research Institute, University of Nevada, Las Vegas.

Chân thành cảm ơn sự nhiệt tình, tận tâm và dễ thương của thầy TS. Ngô Quốc Việt. Cảm ơn Hà, Vy và các đồng đội Vi5 Research Group đã hết lòng giúp đỡ.