DeepSolo được hướng dẫn bởi từ điển cho việc phát hiện và nhận dạng văn bản tiếng Việt trong cảnh

Đoàn Nhật Sang[1,2]

[1] Vietnam National University
Ho Chi Minh City, Vietnam

[2] University of Information Technology
Ho Chi Minh City, Vietnam

What ?

Chúng tôi áp dụng DeepSolo cho bài toán phát hiện và nhận dạng văn bản tiếng Việt, trong đó chúng tôi:

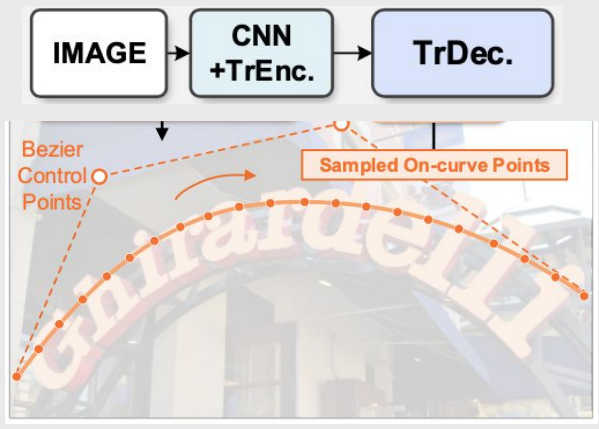
- Xây dựng bộ dữ liệu UITText cho tiếng Việt, với số lượng văn bản cong lớn.
- Cải tiến DeepSolo bằng cách tích hợp bộ từ điển vào quá trình huấn luyện và suy luận
- Áp dụng DeepSolo và phiên bản cải tiến trên dữ liệu tiếng Việt: VinText, UITText

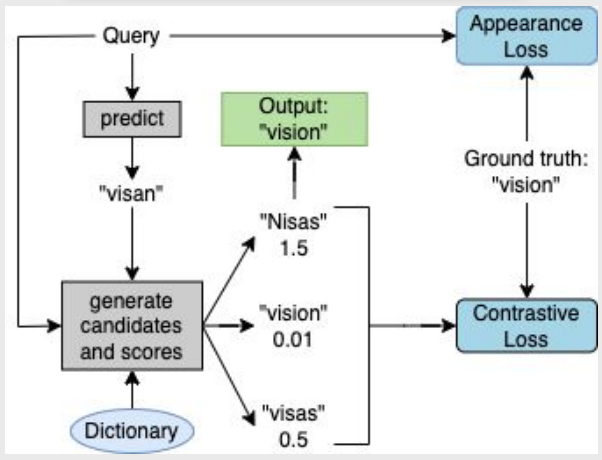
Why ?

- Phát hiện và nhận dạng văn bản trong cảnh ngày càng được quan tâm vì những ứng dụng của nó: Rút trích thông tin từ ảnh, xe tự lái, hỗ trợ người khiếm thị,...
- DeepSolo là phương pháp SOTA trên nhiều bộ dataset tiếng Anh và Trung, nhưng chưa được đánh giá trên tiếng Việt và chưa tận dụng hiệu quả thông tin ngôn ngữ biết trước.
- Các bộ dữ liệu trên tiếng Việt hiện tại chứa rất ít văn bản cong

Overview

Xây dựng dữ liệu → Tìm hiểu DeepSolo → Cải tiến DeepSolo





Description

1. Xây dựng dữ liệu

- Tìm hiểu cách xây dựng bộ dữ liệu VinText, CTW1500
- Thu thập ảnh 2000 chứa văn bản cong ở mức độ từ hoặc dòng. Các ảnh được lấy từ Internet hoặc chụp trong thực tế bởi 5 người thu thập.
- Đánh nhãn theo quy trình của CTW1500

2. Tìm hiểu DeepSolo

- Tìm hiểu các backbone: ResNet/Swin/ViTAE/..
- Tìm hiểu kiến trúc của Encoder, Decoder, các đầu dự đoán
- Tìm hiểu cách thiết kế query
- Tìm hiểu tiêu chí so khớp văn bản khi tối ưu hàm mất mát

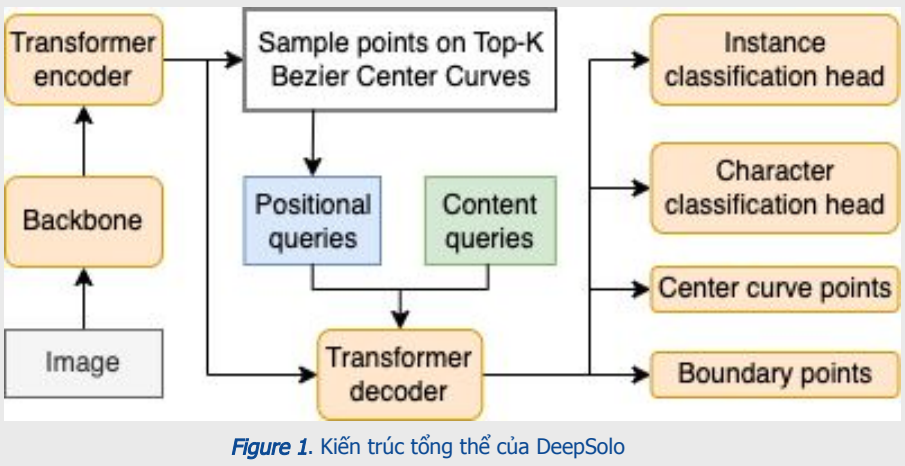


Figure 1. Kiến trúc tổng thể của DeepSolo

3. Cải tiến DeepSolo

- Tìm hiểu cách kết hợp bộ từ điển vào trong quá trình huấn luyện và suy luận
- Mỗi query đại diện cho một từ (đầu ra của DeepSolo decoder) sẽ được đưa qua character classification head để dự đoán từ
- Danh sách ứng cử là k từ trong từ điển với edit distance nhỏ nhất so với từ được dự đoán
- Thêm contrastive loss để tối đa khả năng của các ứng cử gần với ground truth, trong khi tối thiểu khả của các ứng cử khác xa với ground truth. Contrastive loss được tính dựa trên CTC Loss của các từ ứng cử và edit distance của chúng với ground truth

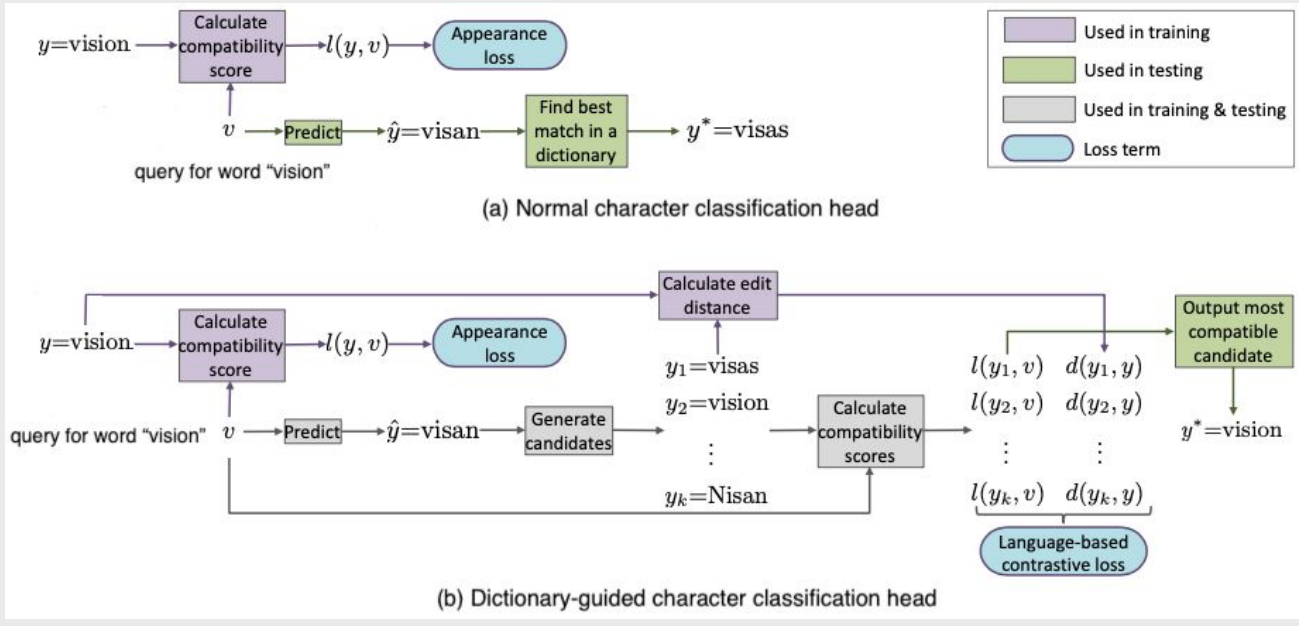


Figure 2. Character classification head của DeepSolo thông thường và sau khi được cải tiến

Đoàn Nhật Sang – Trường Đại học Công nghệ thông tin
TEL : 0899679908 Email : 21522542@gm.uit.edu.vn