**ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

bdôca



**SCENE TEXT VIETNAMESE**

**(“Nhận diện chữ tiếng Việt trong ảnh ngoại cảnh và sinh hoạt hằng ngày”)**

|  |  |  |
| --- | --- | --- |
| Sinh viên thực hiện: | | |
| STT | Họ tên | MSSV |
| 1 | Nguyễn Lê Thanh (NT) | 19522238 |
| 2 | Phạm Vĩ | 19521101 |

**TP. HỒ CHÍ MINH – 05/2022**

# GIỚI THIỆU VỀ ĐỒ ÁN

Đây là bài toán trong cuộc thi AI Challenge 2021 **(“Nhận diện chữ tiếng Việt trong ảnh ngoại cảnh và sinh hoạt hằng ngày”**). Mục tiêu là phát hiện (**detect**) và nhận diện (**recognize**) chữ trong ảnh, cụ thể ở đây sẽ tập trung vào chữ trong khung cảnh (**scene text**) được thu lại từ nhiều nguồn camera khác nhau ở Việt Nam.

Trong báo cáo này, nhóm chúng em đã thử nghiệm 2 hướng tiếp cận khác nhau trên bộ dữ liệu **Vintext dataset** để đánh giá và so sánh.

* ***Hướng tiếp cận 1:*** Sử dụng framework PaddleOCR, trong đó sử dụng model SAST để predict những bounding box chứa text sau đó đưa qua model SRN để predict text ở trong bouding box.
* ***Hướng tiếp cận 2:*** Sử dụng dict-guide được build dựa trên ABCNet.

# NỘI DUNG

## Vintext Dataset

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **#imgs** | **#text instances** | **Examples** |
| VinText | 2000 | About 56000 |  |

Trong tập Vintext Dataset có dạng:

**Vintext**

*Folder Labels* - chứa các file annotation của từng image.

*Folder train\_images* – chứa 1200 ảnh từ im0001 đến im1200.

*Folder test\_image* – chứa 300 ảnh từ im1201 đến 1500.

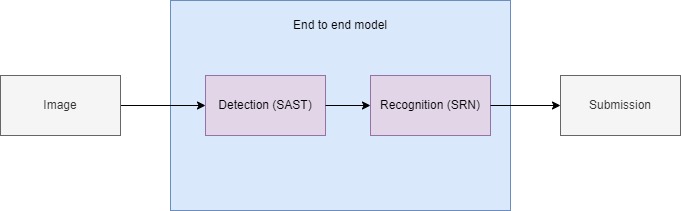
*Folder unseen\_test\_images* – chứa 500 ảnh từ im1501 đến im2000

*File general\_dict.txt*

*File vn\_dictionary.txt*

## Hướng tiếp cận 1:

Chia bài toán nhận diện chữ tiếng Việt thành 2 task nhỏ hơn là phát hiện chữ tiếng Việt và nhận diện chữ tiếng Việt. Để giải quyết 2 task trên, nhóm em sử dụng phương pháp sau:

* Phát hiện chữ tiếng Việt (detect): SAST
* Nhận diện chữ tiếng Việt (recognize): SRN

*Hình 1: Quy trình xử lý hướng tiếp cận 1*

### Model Detection SAST:

*Hình 2: Pipeline SAST*

1. ***Tiền xử lý dữ liệu, convert về format của PaddleOCR để train:***

Để train model detection PaddleOCR, file annotation cần có format như sau:

*“Image\_file\_name”\_\_ “Image annotation information encoded by json.dumps”*

**Ví dụ**: img\_001.jpg [{“transcription”: “text”, “points”: [[310, 104], [416, 141], [418, 216], [312, 179]]}, {…}]

* *Trong đó:*
  + - points sẽ là các cặp (x, y) 4 góc của text box theo chiều ngược kim đồng hồ, bắt đầu từ góc dưới bên trái.
    - transcription là text ở trong text box hiện tại. Khi chứa “###” thì có nghĩa là text box này invalid và sẽ skip đi khi train model.

Ý tưởng để convert về format PaddleOCR:

* Dùng một vòng lặp for đọc từng file annotation của mỗi ảnh.
* Chạy một vòng lặp for đọc các dòng trong file annotation cho đến hết để lấy các text box.
* Dùng split(‘,’, 8) để tách từng point và text ra sau đó lưu vào dictionary (trong lệnh split thêm tham số 8 để tránh trường hợp text chứa dấu phẩy sẽ bị thiếu dấu phẩy trong text).

1. ***Chuẩn bị file config:***

Sử dụng model SAST với backbone ResNet50\_vd được train sẵn dataset tiếng anh đạt kết quả cao với ICDAR2015 dataset.

*Ảnh có chứa văn bản

Mô tả được tạo tự độngHình 3: File config/SAST.yml*

1. ***Traning***

Khi train model, sau mỗi save\_epoch\_step, checkpoint sẽ được lưu ở đường dẫn được đặt tại dòng Global.save\_model\_dir trong file config gồm 3 file:

* iter\_epoch\_1.pdopt
* iter\_epoch\_1.pdparams
* iter\_epoch\_1.states

1. Ảnh có chứa văn bản

   Mô tả được tạo tự động***Đánh giá mô hình:***

|  |  |  |
| --- | --- | --- |
| **Precision** | **Recall** | **Hmean** |
| 0.8978295394388566 | 0.715008431703204 | 0.7960572635531565 |

*Hình 4: Ảnh sau khi được predict*

### Model Regconize SRN:

*Hình 5: Pipeline SRN*

1. ***Tiền xử lý dữ liệu:***

Ảnh có chứa văn bản, tờ báo, ảnh chụp màn hình

Mô tả được tạo tự độngData train model recognition thì nó sẽ khác so với model detection đó là data sẽ là các ảnh nhỏ và một file txt lưu đường dẫn ảnh cùng với chữ trong ảnh đó.

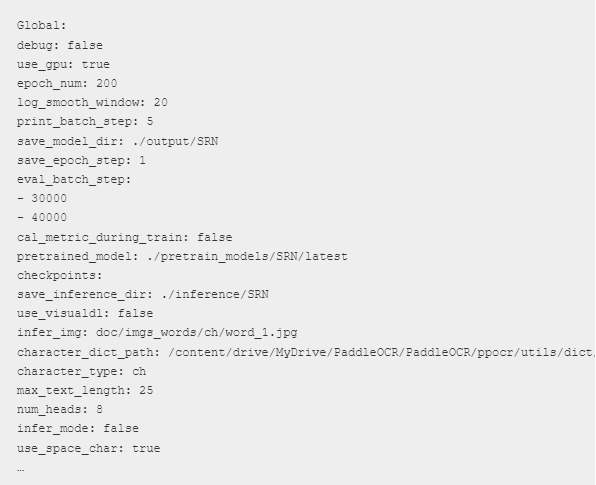
*Hình 6: Ví dụ ảnh đầu vào và label đầu vào của model SRN*

1. **Chuẩn bị dictionary:**

Để dự đoán được tiếng việt thì ta cũng cần một file dictionary dành cho tiếng Việt chứa tất cả các kí tự. Vì PaddleOCR chưa hỗ trợ tiếng Việt nên cần tạo 1 file dictionary để train.

1. ***Chuẩn bị file config:***

Sử dụng model SRN với backbone là ResNetFPN.

*Hình 7: File config/SRN.yml*

### Inference với model đã train xong

Sau khi detect và recognize, nhóm em đã inference cả 2 model. Kết quả cho ra có 3 tệp trong mục lưu mô hình:

* *inference.pdiparams*: Tệp tham số của mô hình suy luận.
* *inference.pdiparams.info*: Thông tin tham số của mô hình suy luận (có thể bị bỏ qua).
* *inference.pdmodel*: Tệp chương trình của mô hình.

### Kết hợp cả 2 mô hình dể predict và đánh giá:

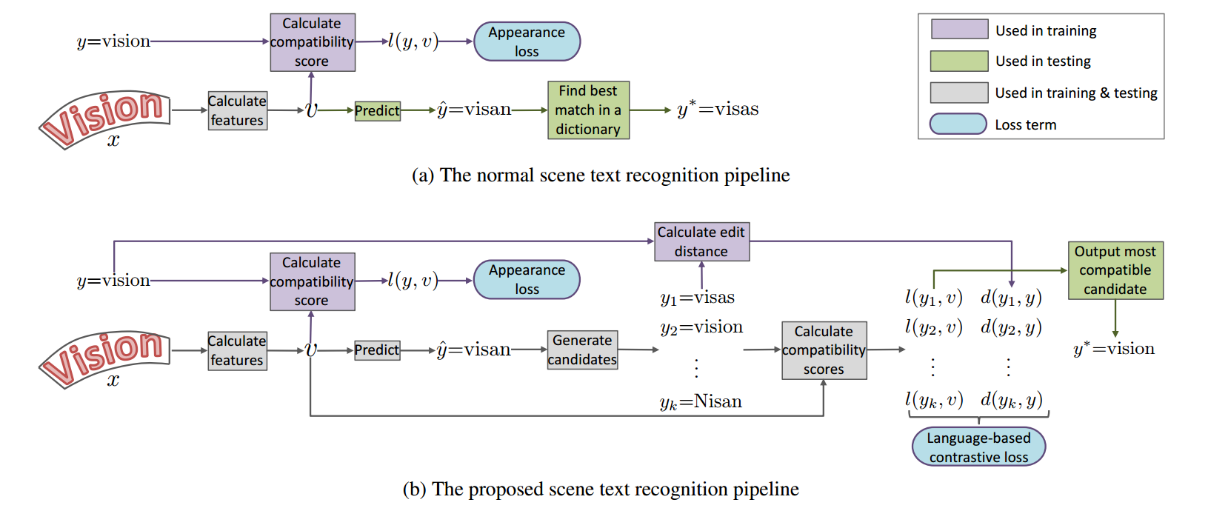
Cuối cùng, nhóm em kết hợp cả 2 model SAST và SRN để xử lý bài toán nhận diện chữ tiếng Việt. Kết quả đánh giá sau khi kết hợp:

|  |  |  |
| --- | --- | --- |
| **Precision** | **Recall** | **Hmean** |
| 0.51 | 0.30 | 0.38 |



*Hình 6: Hình ảnh sau khi predict*

## Hướng tiếp cận 2:

 Trong hướng tiếp cận thứ 2 này, nhóm chúng em đã sử dụng Dict-guide của VinAI để giải quyết bài toán.

Hình 6: pipeline dict-guide

Kết quả đánh giá sau khi train model:

|  |  |  |
| --- | --- | --- |
| **Precision** | **Recall** | **Hmean** |
| 0.9004 | 0.8014 | 0.8480 |

*Hình 7: Ảnh sau khi được predict*

# KẾT LUẬN

Bảng so sánh kết quả đánh giá 2 phương pháp tiếp cận:

|  |  |  |
| --- | --- | --- |
|  | **SAST+ SRN** | **Dict-Guide (VinAI)** |
| **Precision** | 0.51 | 0.9004 |
| **Recall** | 0.30 | 0.8014 |
| **Hmean** | 0.38 | 0.8480 |

Nhóm đã thử nghiệm được 2 hướng tiếp cận cho bài toán “Nhận diện chữ tiếng Việt trong ảnh ngoại cảnh và sinh hoạt hằng ngày”. Tuy nhiên, hướng tiếp cận đầu tiên có kết quả đánh giá khá thấp. Phần detect của model SAST cho kết quả khá khả quan, nhưng phần recognize của model SRN thì lại cho kết quả sai khá nhiều (những chữ cái có dấu không predict được).

*Hình 8: Ví dụ recognize sai*

Để cải thiện và phát triển hướng thứ nhất bằng cách tăng cường dữ liệu với phương pháp cắt dán các bounding box có chứa chữ vào ảnh nền (không chứa văn bản) và thay đổi màu sắc, tính chất của bức ảnh.

**TÀI LIỆU THAM KHẢO**

PaddleOCR. URL: <https://github.com/PaddlePaddle/PaddleOCR>

A Single-Shot Arbitrarily-Shaped Text Detector based on Context Attended Multi-Task Learning. URL: <https://arxiv.org/abs/1908.05498>

Towards Accurate Scene Text Recognition with Semantic Reasoning Networks. URL: <https://arxiv.org/abs/2003.12294>

Nguyen, N., Nguyen, T., Tran, V., Tran, T. M., Ngo, T. D., Nguyen, H. T., & Hoai, M. (2021, June). CVPR 2021 Open Access Repository. CVPR 2021 Open Access Repository: <https://openaccess.thecvf.com/content/CVPR2021/html/Nguyen_Dictionary-Guided_Scene_Text_Recognition_CVPR_2021_paper.html>

1. Dict-guide. URL: <https://github.com/VinAIResearch/dict-guided>

**PHỤ LỤC**

Phụ lục hình 3:

* **Global:**
  + use\_gpu: sử dụng gpu hoặc không
  + epoch\_num: số epoch tối đa khi train .
  + save\_model\_dir: là đường dẫn thư mục lưu checkpoint cho mỗi epoch khi train.
  + pretrained\_model: là đường dẫn thư mục pretrained model.
  + checkpoints: là đường dẫn thư mục load checkpoint khi train.
* **Train**:
* data\_dir: là đường dẫn thư mục chứa ảnh train
* label\_file\_list: list các file annotation
* ratio\_list: tỉ lệ train data set

Phụ lục hình 7:

* character\_dict\_path – đường dẫn file dictionary
* use\_space\_char – dự đoán khoảng trắng hay không
* max\_text\_length – độ dài tối đa kí tự trong một box.

**PHỤ LỤC PHÂN CÔNG NHIỆM VỤ**

|  |  |  |
| --- | --- | --- |
| **STT** | **Thành viên** | **Nhiệm vụ** |
| 1 | Nguyễn Lê Thanh | Tìm hiểu về model detect SAST, viết báo cáo |
| 2 | Phạm Vĩ | Tìm hiểu về model recognize SRN, demo |
| 3 |  |  |

***(Chú ý: Ghi rõ từng nhiệm vụ chi tiết của mỗi thành viên).***

Khi làm bài thu hoạch SV phải thực hiện đúng các quy định sau:

1. Sử đụng đúng Template này. Không đúng thì **– 2 điểm**.
2. Trình bày đề tài tối thiểu là 05 trang và tối đa là 10 trang A4
   * Không tính Trang Bìa, và Phụ lục phân công công việc trong template.
   * Phụ lục code
3. Copy y chan trên Internet thì gọi là đạo văn và **nhận 0 điểm** cho bài thu hoạch.
4. Nộp đúng hạn theo thông báo của GV.
5. Không trình bày code trong nội dung bài thu hoạch. Nếu muốn trình bày code thì trình bày tại phần phụ lục.

Sản phẩm đồ án môn học:

1. File word (thêm pdf) báo cáo, ko in ra, chỉ nộp file mềm.
2. Demo kết quả.
3. Slide thuyết trình (Nếu nhóm được chọn).