SCENE TEXT VIETNAMESE

"Nhận diện chữ tiếng Việt trong ảnh ngoại cảnh và sinh hoạt hằng ngày"

Nhóm trình bày: Nhóm 17
Sinh viên:
1. Nguyễn Lê Thanh - 19522238
2. Phạm Vĩ - 19521101

CÁC CHỦ ĐỀ CHÍNH

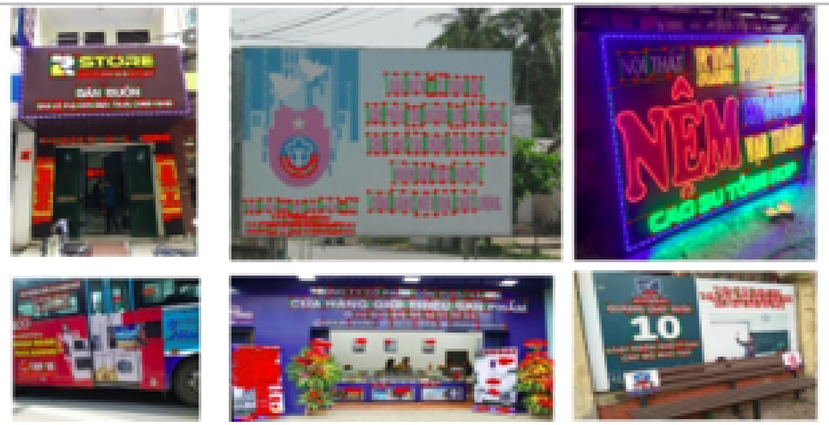
1. Tóm tắt lại bài toán Scene Text Vietnamses.
2. Các phương pháp đã thử nghiệm.
3. Các kết quả đạt được.
4. Tổng kết và hướng phát triển.

1. Tóm tắt bài toán Scene text Vietnamese.

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.

2. CÁC PHƯƠNG PHÁP ĐÃ THỬ NGHIỆM

THỬ NGHIỆM 2 HƯỚNG TIẾP CẬN TRÊN BỘ DỮ LIỆU VINTEXT.

Name	#imgs	#text instances	Examples
VinText	2000	About 56000	

- Hướng tiếp cận 1:
  - Detect: model SAST
  - Recognize: model SRN
- Hướng tiếp cận 2:
  - model Dict-guide

Vintext Dataset

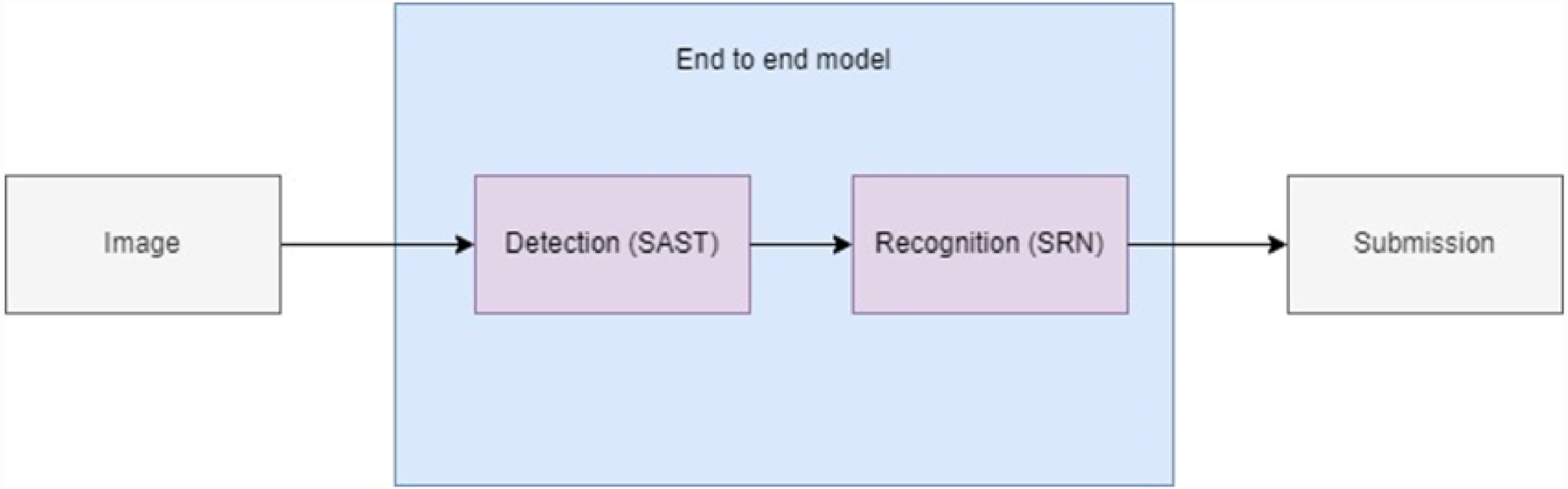
- Folder Labels - chứa các file annotation của từng image.
- Folder train_images – chứa 1200 ảnh từ im0001 đến im1200.
- Folder test_image – chứa 300 ảnh từ im1201 đến 1500.
- Folder unseen_test_images – chứa 500 ảnh từ im1501 đến im2000
- File general_dict.txt
- File vn_dictionary.txt

Ảnh trong folder train_image

```
690,402,712,402,716,420,692,420,###
768,445,800,442,799,448,769,450,###
782,506,799,506,799,514,784,515,###
771,461,799,458,799,484,776,486,PHÒN
356,113,363,114,363,130,351,129,Ở
367,114,367,114,378,115,378,130,Y
382,112,400,111,401,132,382,131,TẾ
405,116,456,120,457,134,404,132,QUANG
460,119,498,121,499,135,460,135,NINH
```

file .txt trong folder labels

HƯỚNG TIẾP CẬN 1

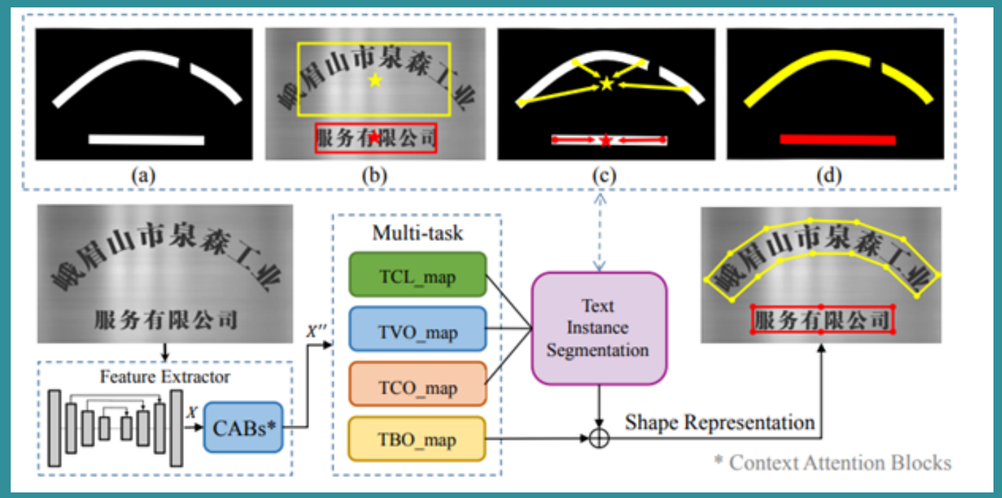


Quy trình xử lý hướng tiếp cận 1

Model Detect SAST

XỬ LÝ DỮ LIỆU

```
65,35,82,35,82,39,65,39,###
94,10,117,10,117,41,93,41,CHẤT
118,15,147,15,148,46,118,46,LƯỢNG
149,9,165,9,165,43,150,43,TỐT
167,9,180,9,179,43,167,42,ĐỂ
181,12,193,12,193,43,181,43,CÓ
195,13,215,14,215,46,196,46,VIỆC
217,13,237,14,239,47,217,46,LÀM,
240,14,265,14,266,45,240,44,NĂNG
267,11,291,11,291,46,267,45,SUẤT
293,15,310,16,311,46,292,45,CAO
312,11,327,10,324,44,312,44,ĐỂ
```



Ý tưởng:

- Dùng một vòng lặp for đọc từng file annotation của mỗi ảnh.
- Chạy một vòng lặp for đọc các dòng trong file annotation để lấy các text box.
- Dùng split(',', 8) để tách từng point và text ra sau đó lưu vào dictionary.

```
im1432.jpg  [{"transcription": "cafe", "points": [["277", "32"], ["367", "22"], ["364", "60"], ["278", "70"]]}, {"transcription": "Nhỏ", "points":
im1297.jpg  [{"transcription": "LÊ", "points": [["252", "39"], ["294", "60"], ["275", "155"], ["239", "127"]]}, {"transcription": "NGUYÊN", "point
im1327.jpg  [{"transcription": "ĐỊA", "points": [["487", "107"], ["512", "105"], ["518", "120"], ["488", "122"]]}, {"transcription": "PHẬN", "poin
im1370.jpg  [{"transcription": "BỆNH", "points": [["89", "82"], ["228", "85"], ["235", "193"], ["101", "191"]]}, {"transcription": "XÁ", "points":
im1469.jpg  [{"transcription": "###", "points": [["441", "128"], ["476", "141"], ["470", "150"], ["437", "140"]]}, {"transcription": "###", "point
im1461.jpg  [{"transcription": "BAN", "points": [["96", "88"], ["114", "88"], ["114", "97"], ["96", "97"]]}, {"transcription": "CHỈ", "points": [[
im1315.jpg  [{"transcription": "Z", "points": [["135", "27"], ["155", "26"], ["148", "51"], ["131", "50"]]}, {"transcription": "X", "points": [["2
im1408.jpg  [{"transcription": "Welcome", "points": [["265", "116"], ["358", "117"], ["358", "142"], ["263", "144"]]}, {"transcription": "to", "po
im1377.jpg  [{"transcription": "CHÙA", "points": [["215", "338"], ["313", "322"], ["311", "381"], ["212", "399"]]}, {"transcription": "LONG", "poi
```

CHUẨN BỊ FILE CONFIG

- backbone ResNet50_vd

Precision	Recall	Hmean
0.8978295394388566	0.715008431703204	0.7960572635531565

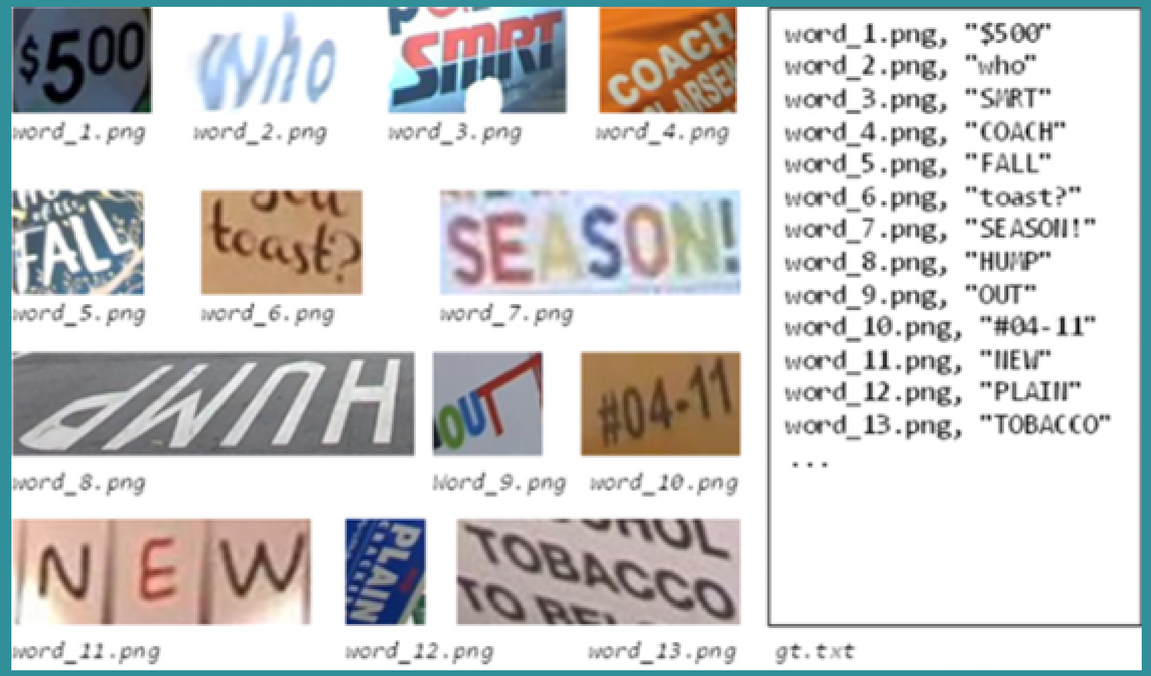
Ảnh sau khi được predict

```
Global:
    debug: false
    use_gpu: true
    epoch_num: 300
    log_smooth_window: 20
    print_batch_step: 2
    save_model_dir: ./output/SAST/
    save_epoch_step: 1
    eval_batch_step:
    - 40000
    - 50000
    cal_metric_during_train: false
    pretrained_model: ./pretrain_models/SAST/latest
    checkpoints:
    save_inference_dir: ./inference/SAST
    use_visualdl: True
    infer_img: null
    save_res_path: ./output/sast.txt
…
Train:
    dataset:
      name: SimpleDataSet
      data_dir: ./data #path_img_dir
      label_file_list: [ train_label.txt ]
      ratio_list: [1]
…
Eval:
    dataset:
      name: SimpleDataSet
      data_dir: #path_img_dir
      label_file_list:
      - unseen_label.txt
```

Model Recognize SRN

XỬ LÝ DỮ LIỆU

Data train model recognition sẽ là các ảnh nhỏ và một file txt lưu đường dẫn ảnh cùng với chữ trong ảnh đó.



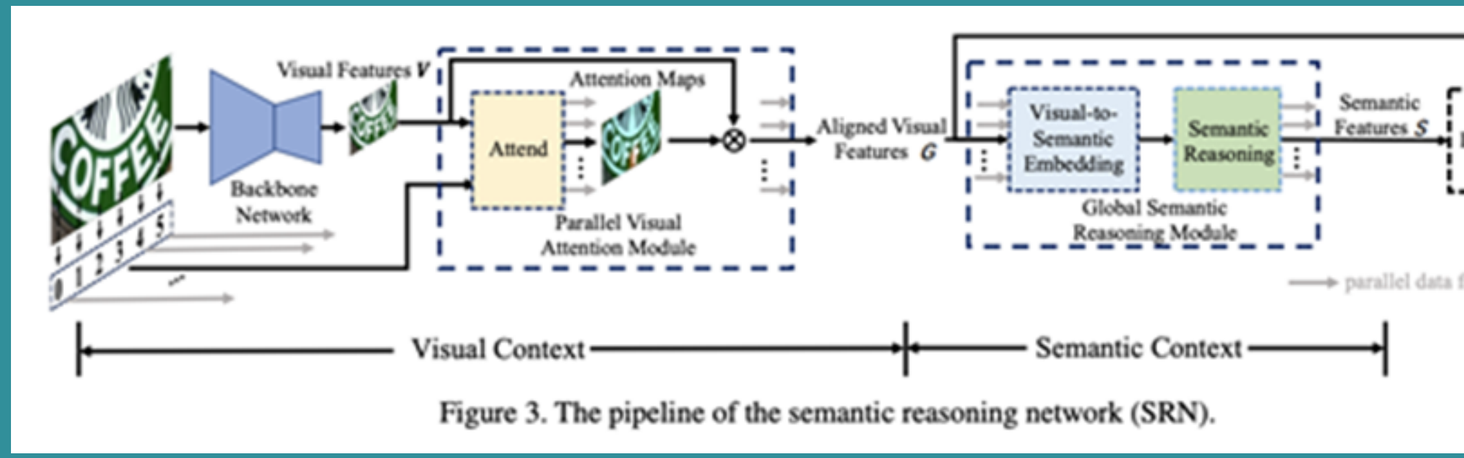


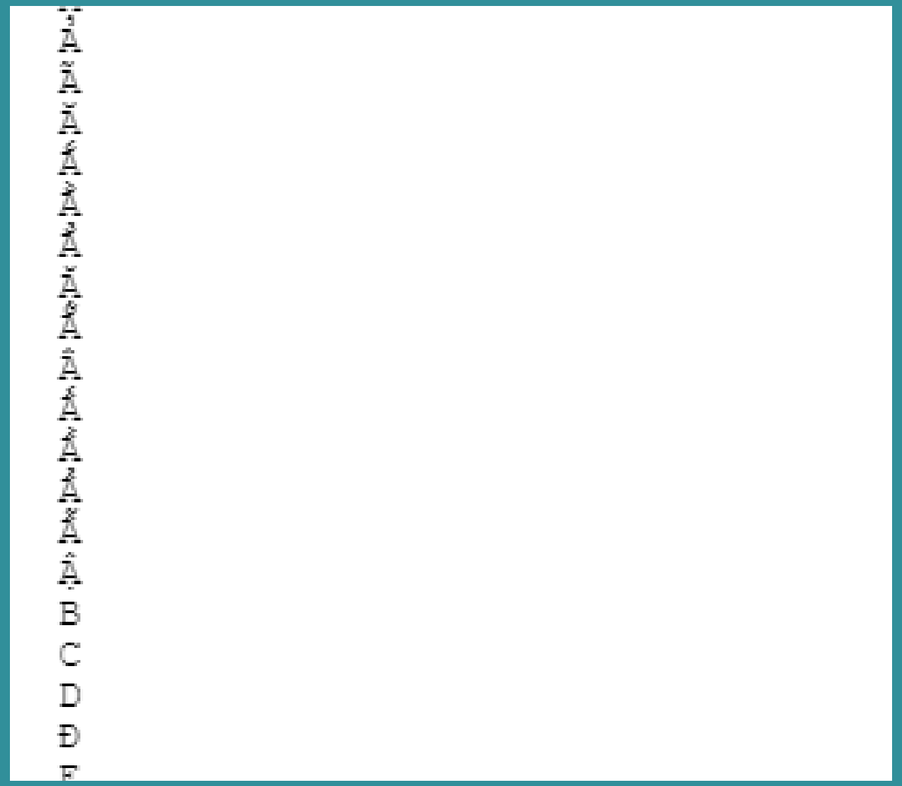
Figure 3. The pipeline of the semantic reasoning network (SRN).

Ý tưởng:

- Dựa vào các point trong labels để cắt ảnh thành những ảnh nhỏ.

CHUẨN BỊ DICTIONARY

Để dự đoán được tiếng việt cần một file dictionary dành cho tiếng Việt chứa tất cả các kí tự.



CHUẨN BỊ FILE CONFIG

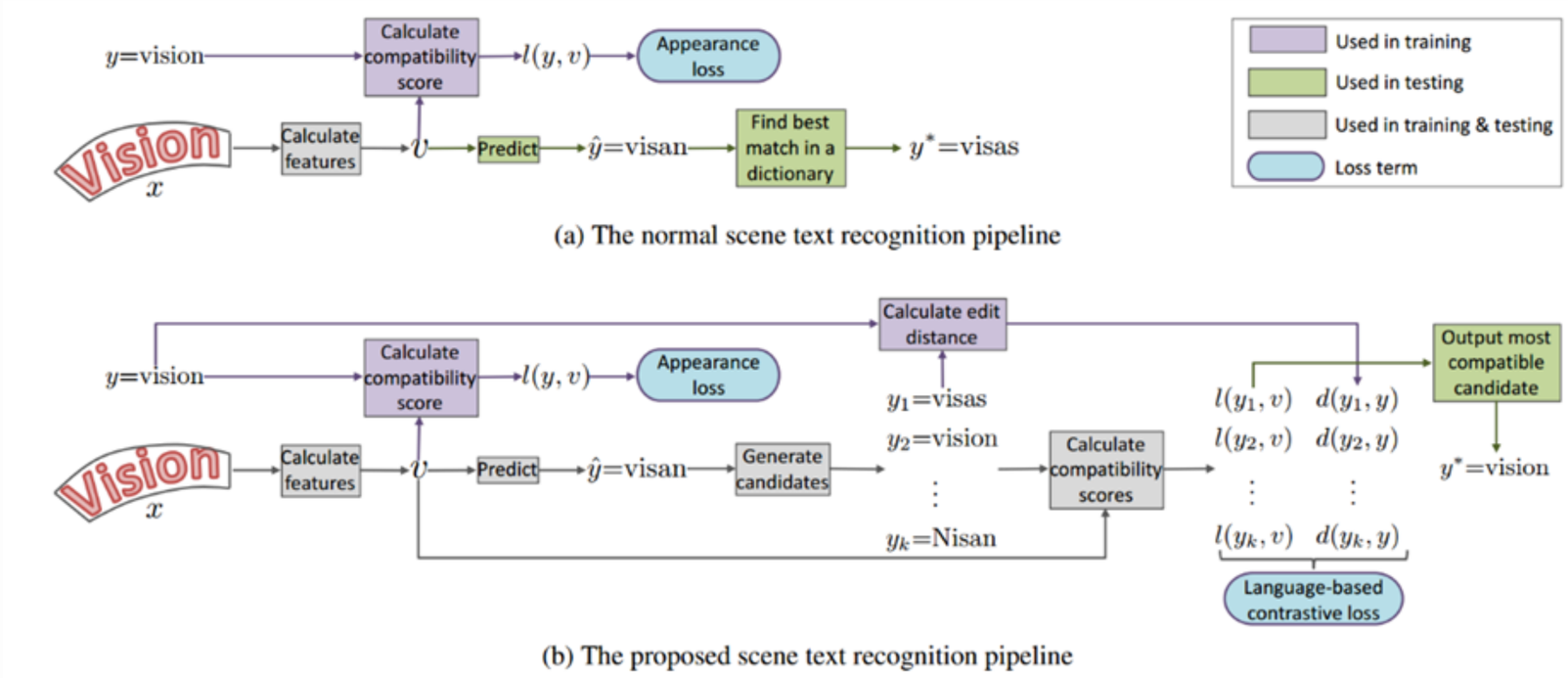
backbone là ResNetFPN.

```
Global:
debug: false
use_gpu: true
epoch_num: 200
log_smooth_window: 20
print_batch_step: 5
save_model_dir: ./output/SRN
save_epoch_step: 1
eval_batch_step:
- 30000
- 40000
cal_metric_during_train: false
pretrained_model: ./pretrain_models/SRN/latest
checkpoints:
save_inference_dir: ./inference/SRN
use_visualdl: false
infer_img: doc/imgs_words/ch/word_1.jpg
character_dict_path: /content/drive/MyDrive/PaddleOCR/PaddleOCR/ppocr/utils/dict/
character_type: ch
max_text_length: 25
num_heads: 8
infer_mode: false
use_space_char: true
...
```

KẾT HỢP 2 MODEL

Hình ảnh predict sau khi kết hợp 2 model

HƯỚNG TIẾP CẬN 2



Quy trình hướng tiếp cận 2

3. KẾT QUẢ ĐẠT ĐƯỢC

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

4. Tổng kết và hướng phát triển

- Đã thử nghiệm được 2 hướng tiếp cận cho bài toán
- Hướng tiếp cận 1 (SAST + SRN) có kết quả đánh giá khá thấp, model SRN cho kết quả predict sai khá nhiều.
- Cải thiện 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.

DEMO