SAMSUNG SDS

Foresee

Techtonic 2021

Partner

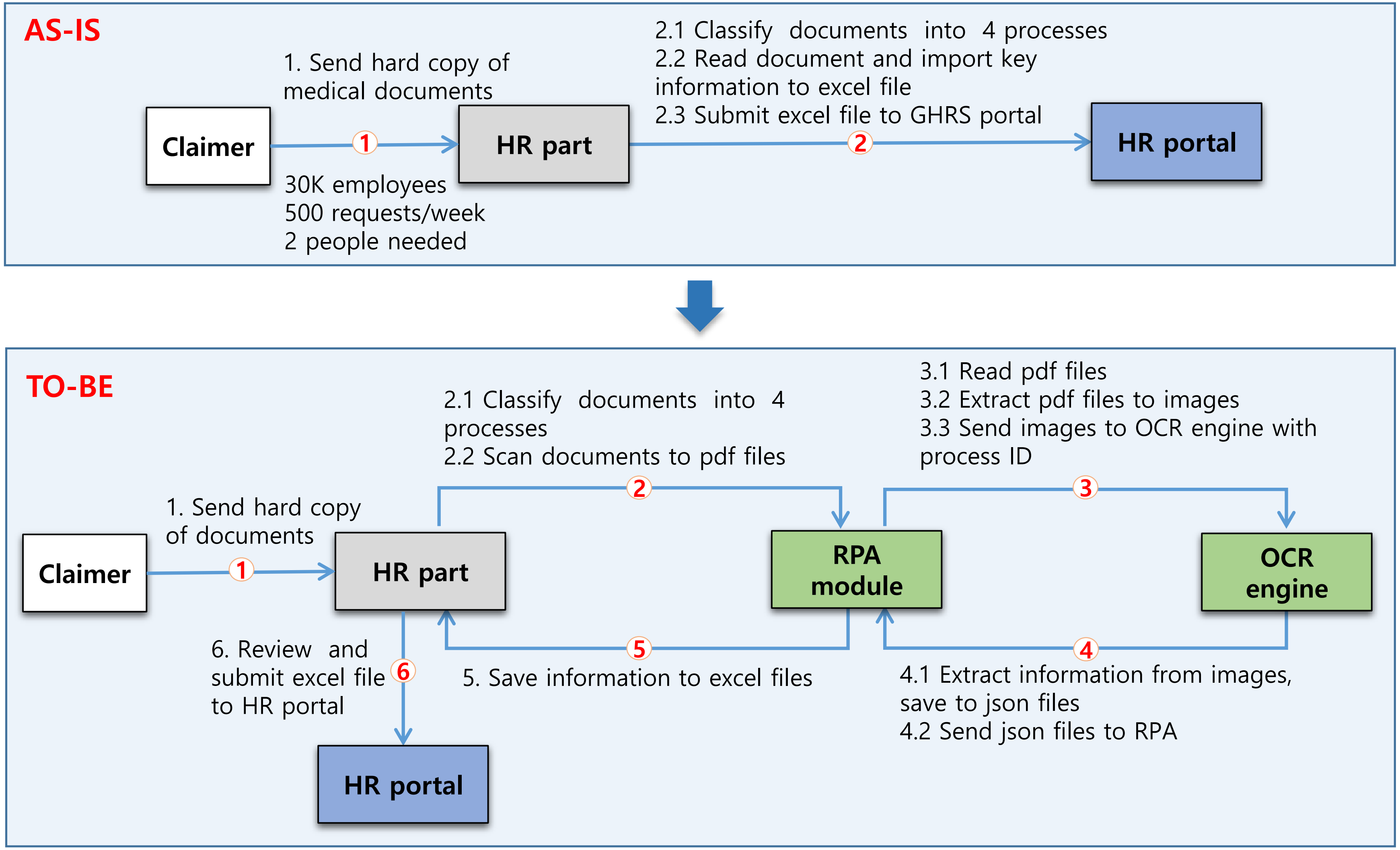
Disrupt

Samsung SDS Research Vietnam

An approach in Vietnamese handwritten OCR

Ngo T Dat

Use case: data input automation



AS-IS

1. Send hard copy of medical documents

30K employees
500 requests/week
2 people needed

2.1 Classify documents into 4 processes
2.2 Read document and import key information to excel file
2.3 Submit excel file to GHRS portal

TO-BE

1. Send hard copy of documents

2.1 Classify documents into 4 processes
2.2 Scan documents to pdf files

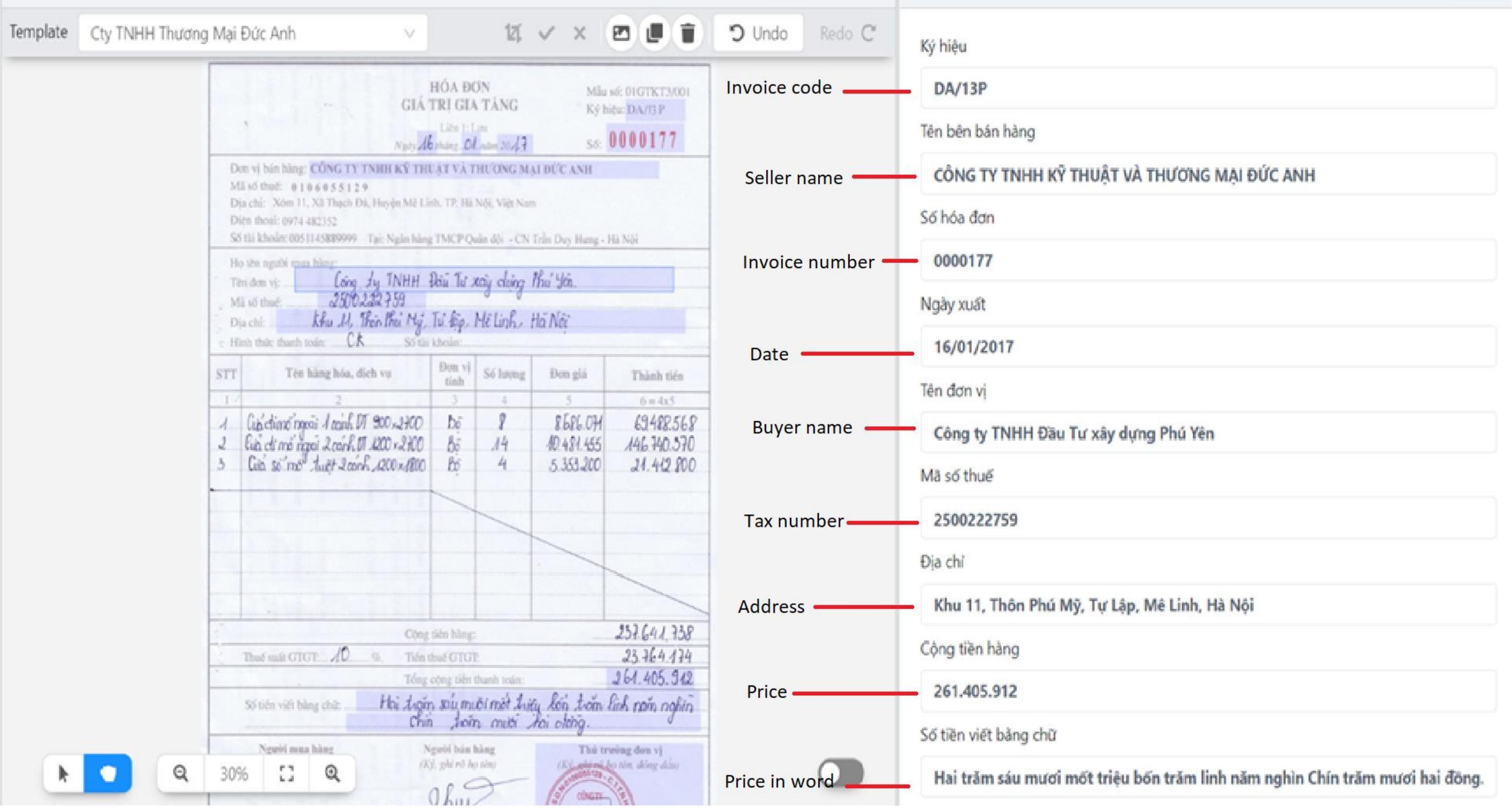
3.1 Read pdf files
3.2 Extract pdf files to images
3.3 Send images to OCR engine with process ID

4.1 Extract information from images, save to json files
4.2 Send json files to RPA

5. Save information to excel files

6. Review and submit excel file to HR portal

Financial sector: Invoice, application form



HÓA ĐƠN
GIÁ TRỊ GIA TĂNG

Liên 1: Lưu

Ngày 16 tháng 01 năm 2017

Mẫu số: 01GTKT3/001
Ký hiệu: DA/13P
Số: 0000177

Đơn vị bán hàng: CÔNG TY TNHH KỸ THUẬT VÀ THƯƠNG MẠI ĐỨC ANH
Mã số thuế: 0106055129
Địa chỉ: Xóm 11, Xã Thạch Đà, Huyện Mê Linh, TP. Hà Nội, Việt Nam
Điện thoại: 0974 482352
Số tài khoản: 0051145889999 Tại: Ngân hàng TMCP Quân đội - CN Trần Duy Hưng - Hà Nội

Họ tên người mua hàng:
Tên đơn vị: Công ty TNHH Đầu Tư xây dựng Phú Yên.
Mã số thuế: 2500222759
Địa chỉ: Khu 11, Thôn Phú Mỹ, Tự Lập, Mê Linh, Hà Nội
Hình thức thanh toán: CK Số tài khoản:

STT	Tên hàng hóa, dịch vụ	Đơn vị tính	Số lượng	Đơn giá	Thành tiền
1	2	3	4	5	6 = 4x5
1	Cửa đi mở ngoài 1 cánh DT 900×2700	Bộ	8	8.686.071	69.488.568
2	Cửa đi mở ngoài 2 cánh DT 1200×2700	Bộ	14	10.481.455	146.740.370
3	Cửa sổ mở hất 2 cánh 1200×1800	Bộ	4	5.353.200	21.412.800

Cộng tiền hàng: 237.641.738
Thuế suất GTGT: 10 %. Tiền thuế GTGT: 23.764.174
Tổng cộng tiền thanh toán: 261.405.912
Số tiền viết bằng chữ: Hai trăm sáu mươi mốt triệu bốn trăm linh năm nghìn chín trăm mười hai đồng.

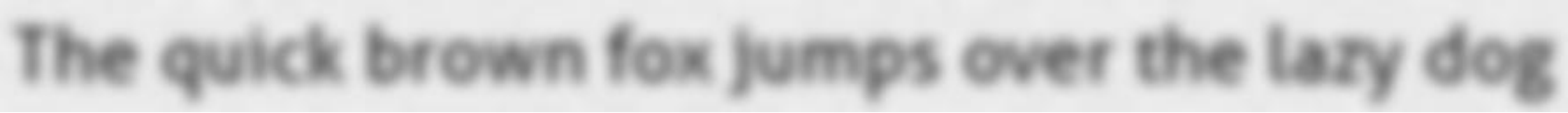
Người mua hàng
Người bán hàng (Ký, ghi rõ họ tên)
Thủ trưởng đơn vị (Ký, ghi rõ họ tên, đóng dấu)

CÔNG TY

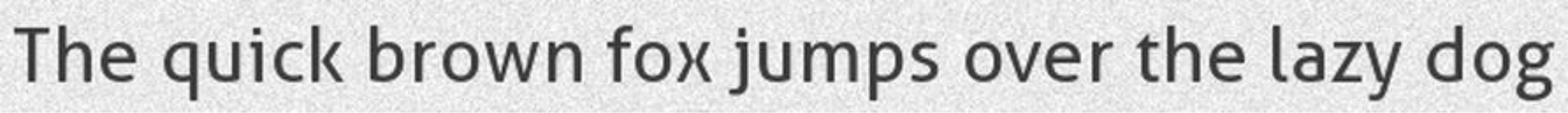
Invoice code	Ký hiệu	DA/13P
Seller name	Tên bên bán hàng	CÔNG TY TNHH KỸ THUẬT VÀ THƯƠNG MẠI ĐỨC ANH
Invoice number	Số hóa đơn	0000177
Date	Ngày xuất	16/01/2017
Buyer name	Tên đơn vị	Công ty TNHH Đầu Tư xây dựng Phú Yên
Tax number	Mã số thuế	2500222759
Address	Địa chỉ	Khu 11, Thôn Phú Mỹ, Tự Lập, Mê Linh, Hà Nội
Price	Cộng tiền hàng	261.405.912
Price in word	Số tiền viết bằng chữ	Hai trăm sáu mươi mốt triệu bốn trăm linh năm nghìn Chín trăm mươi hai đồng.

Limitations of the traditional OCR approach

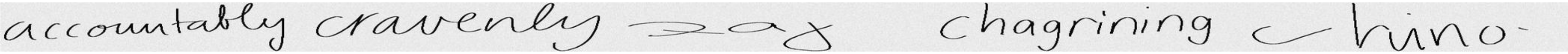
- Bad image samples



The quick brown fox jumps over the lazy dog



The quick brown fox jumps over the lazy dog



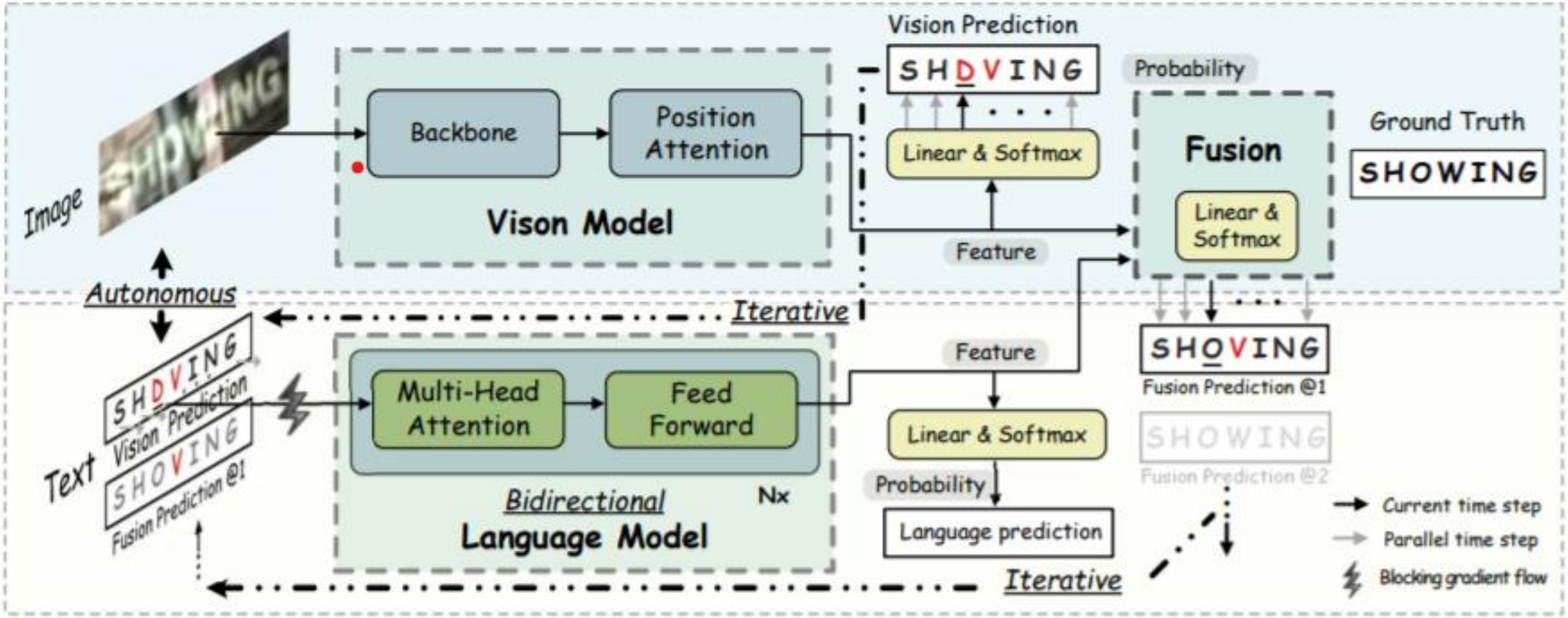
accountably cravenly zag chagrining rhino

- No knowledge about grammar rules or frequently used words.
- Do not understand semantic meaning of words.

⇒ Using Language Model to aid OCR Text Recognition Model can be a good solution.

Overall architecture

The ABINet architecture has demonstrated a feasible way to ensemble Vision Model (VM) and Language Models (LM).

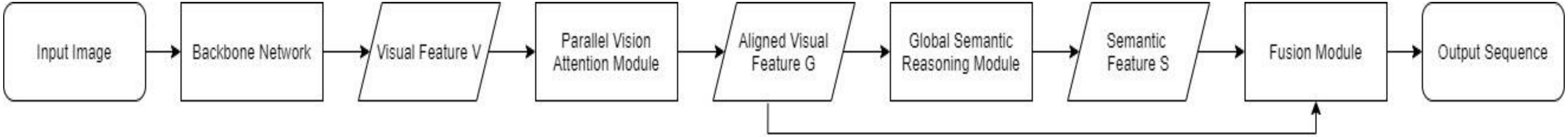


ABINet Architecture

Fang et. al., 2021. "Read Like Humans: Autonomous, Bidirectional and Iterative Language Modeling for Scene Text Recognition". In 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)

Phase 1 - Vision Model & Language Model

Vision Model - SRNet

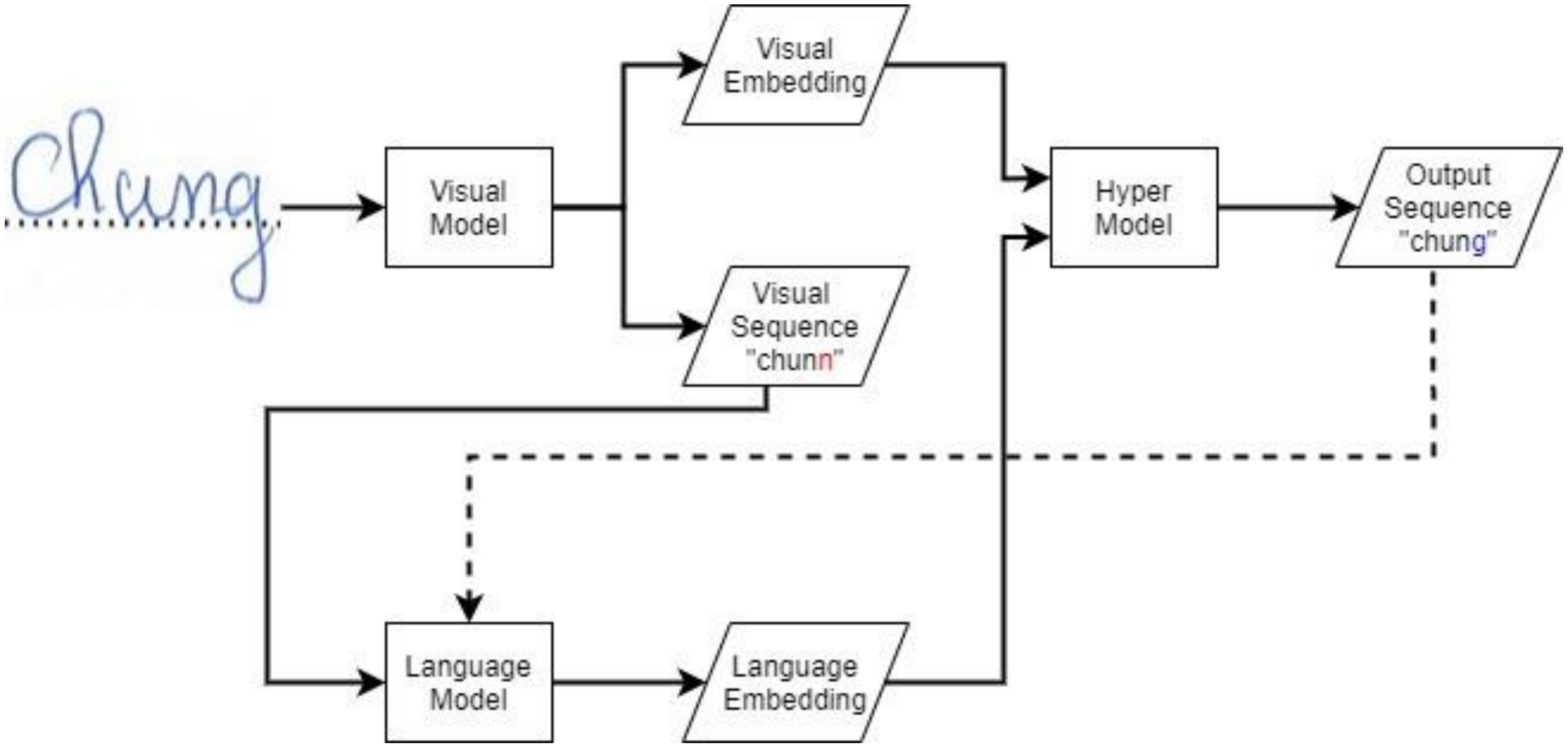


SRNet
Yu et. al., 2020. "Towards Accurate Scene Text Recognition with Semantic Reasoning Networks". In 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)

Language Model - Transformers

- Introduced in 2017
- A ground breaking achievement in NLP field and now expanding to computer vision.

Phase 2 - Ensemble models



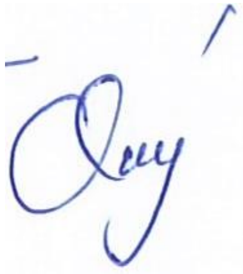
- Training procedure includes two phases.
- In phase 2, two models are jointly trained and a hyper-model (HM) was introduced to stack two phase 1 models' embedding vectors.
- Inference procedure can be single forwarded or iteratively forwarded.

Single word benchmark

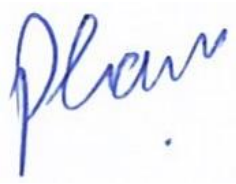
- Training data:
  - Two datasets: Vietnamese corpus (50K words, 26M samples), Vietnamese text images (20K images from our employees and 180K generated).
- Testing data:
  - 1300 handwritten words image dataset. One image contains a single cropped word (sized 32x128)
- The HM drastically improved recognition accuracy while not pushing to much computing burden.

Model	Accuracy (percent)	Inference time	
		(ms/word)	(ms/document)
VM (Vision Model)	87.8	2.846	1343
HM (Hyper Model)	95.5	3.076	1452

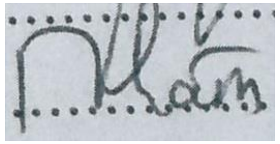
Sample demonstration



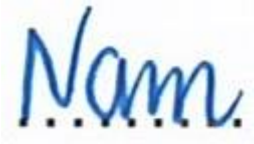
VM prediction: “cúy”
HM prediction: “qúy”
Ground truth: “quý”



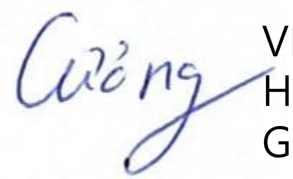
VM prediction: “phạn”
HM prediction: “phạm”
Ground truth: “phạm”



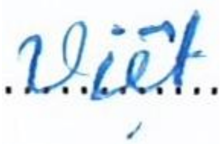
VM prediction: “thâđ”
HM prediction: “chai”
Ground truth: “nhâm”



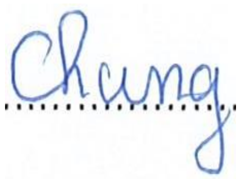
VM prediction: “namc”
HM prediction: “nam”
Ground truth: “nam”



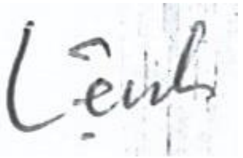
VM prediction: “cuông”
HM prediction: “cương”
Ground truth: “cương”



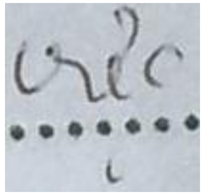
VM prediction: “liitt”
HM prediction: “liiệt”
Ground truth: “việt”



VM prediction: “chunn”
HM prediction: “chung”
Ground truth: “chung”



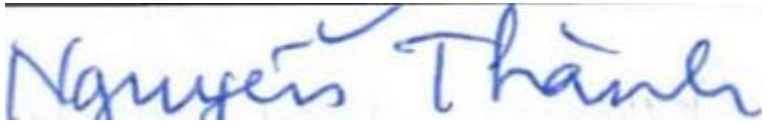
VM prediction: “linh”
HM prediction: “lệnh”
Ground truth: “lệnh”



VM prediction: “qiệc”
HM prediction: “đucc”
Ground truth: “vực”

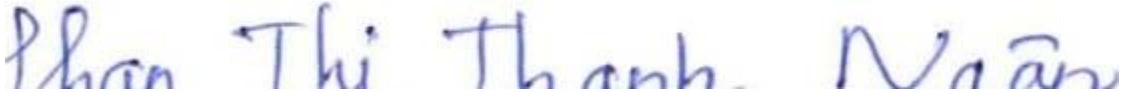
Full name benchmark

- Tested on 400 Vietnamese full names (2 to 4 words):
  - HM: 87% accuracy
  - VM: 77% accuracy



Nguyễn Thành

VM Prediction: "Nouyễn Thành"
HM Prediction: "Nguyễn Thành"
Ground truth: "Nguyễn Thành"

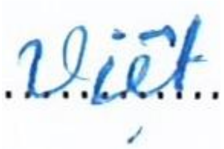


Phan Thị Thanh Ngân

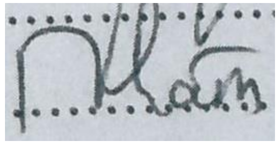
VM Prediction: "Phan Thi Thanh Noân"
HM Prediction: "Phan Thị Thanh Ngân"
Ground truth: "Phan Thị Thanh Ngân"

Future works

- Improve the LM so it can performs well in cases where visual output has more than two errors.



VM prediction: “liitt”
HM prediction: “liiệt”
Ground truth: “việt”



VM prediction: “thâđ”
HM prediction: “chai”
Ground truth: “nhâm”

- Enhance the language model context knowledge even more by switching from character tokenizing mechanism to word tokenizing mechanism,.

Thank you

SAMSUNG SDS