

TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP. HỒ CHÍ MINH
KHOA ĐIỆN ĐIỆN TỬ
BỘ MÔN KỸ THUẬT MÁY TÍNH - VIỄN THÔNG



HCMUTE

ĐỒ ÁN TỐT NGHIỆP

**HỆ THỐNG NHẬP ĐIỂM TỰ ĐỘNG TỪ
ẢNH BÀI THI**

NGÀNH CÔNG NGHỆ KỸ THUẬT MÁY TÍNH

Sinh viên: **NGUYỄN TẤN ĐẠT**
MSSV: 17119117

NGUYỄN NGỌC KHÁNH
MSSV: 17119084

TP. HỒ CHÍ MINH – 07/2022

TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP. HỒ CHÍ MINH
KHOA ĐIỆN ĐIỆN TỬ
BỘ MÔN KỸ THUẬT MÁY TÍNH - VIỄN THÔNG

ĐỒ ÁN TỐT NGHIỆP

HỆ THỐNG NHẬP ĐIỂM TỰ ĐỘNG TỪ ẢNH BÀI THI

NGÀNH CÔNG NGHỆ KỸ THUẬT MÁY TÍNH

Sinh viên: **NGUYỄN TẤN ĐẠT**

MSSV: 17119117

NGUYỄN NGỌC KHÁNH

MSSV: 17119084

Hướng dẫn: **PhD PHAN VĂN CA**

TP. HỒ CHÍ MINH – 07/2022

THÔNG TIN KHÓA LUẬN TỐT NGHIỆP

1. Thông tin sinh viên

- | | |
|--|------------------------|
| - Họ và tên sinh viên: Nguyễn Tấn Đạt | MSSV: 17119117 |
| Email: tandatcr058@gmail.com | Điện thoại: 0379178909 |
| - Họ và tên sinh viên: Nguyễn Ngọc Khánh | MSSV: 17119084 |
| Email: ngoockhanhgl99@gmail.com | Điện thoại: 0396992624 |

2. Thông tin đề tài

- Tên của đề tài: Hệ thống nhập điểm tự động từ ảnh bài thi
- Đơn vị quản lý: Bộ môn Điện Tử Viễn Thông, Khoa Điện - Điện Tử, Trường Đại Học Sư Phạm Kỹ Thuật Tp. Hồ Chí Minh.
- Thời gian thực hiện: Từ ngày 7/3/2022 đến ngày 20/7/2022
- Thời gian bảo vệ trước hội đồng: Từ ngày 27/7/2022 đến ngày 30/7/2022

3. Lời cam đoan của sinh viên

Chúng tôi – Nguyễn Tấn Đạt và Nguyễn Ngọc Khánh cam đoan ĐATN là công trình nghiên cứu của bản thân chúng tôi dưới sự hướng dẫn của PGS.TS. Phan Văn Ca.

Các kết quả công bố trong ĐATN là trung thực và không sao chép từ bất kỳ công trình nào khác.

Tp.HCM, ngày 20 tháng 07 năm 2022

SV thực hiện đồ án
(Ký và ghi rõ họ tên)

Nguyễn Tấn Đạt Nguyễn Ngọc Khánh

BẢN NHẬN XÉT KHÓA LUẬN TỐT NGHIỆP

(Dành cho giảng viên hướng dẫn)

Đề tài: Hệ thống nghiệp vụ công tác ảnh bìa thư

Sinh viên: + Nguyễn Tấn Đạt

MSSV: 17119117

+ Nguyễn Ngọc Khánh

MSSV: 17119084

Hướng dẫn: PGS.TS. Phan Văn Ca

Nhận xét bao gồm các nội dung sau đây:

1. Tính hợp lý trong cách đặt vấn đề và giải quyết vấn đề; ý nghĩa khoa học và thực tiễn:

Đặt vấn đề rõ ràng, mục tiêu cụ thể; đề tài có tính mới, cấp thiết; đề tài có khả năng ứng dụng, tính sáng tạo.

Đặt vấn đề rõ ràng, mục tiêu cụ thể; đề tài có tính mới, cấp thiết; đề tài có khả năng ứng dụng, tính sáng tạo.

2. Phương pháp thực hiện/ phân tích/ thiết kế:

Phương pháp hợp lý và tin cậy dựa trên cơ sở lý thuyết; có phân tích và đánh giá phù hợp; có tính mới và tính sáng tạo.

Phương pháp thực hiện hợp lý và tin cậy dựa trên cơ sở lý thuyết; có phân tích và đánh giá phù hợp; có tính mới và tính sáng tạo.

3. Kết quả thực hiện/ phân tích và đánh giá kết quả/ kiểm định thiết kế:

Phù hợp với mục tiêu đề tài; phân tích và đánh giá/ kiểm thử thiết kế hợp lý; có tính sáng tạo/ kiểm định chặt chẽ và đảm bảo độ tin cậy.

Kết quả thực hiện phù hợp với mục tiêu đề tài; phân tích và đánh giá/ kiểm thử thiết kế hợp lý; có tính sáng tạo/ kiểm định chặt chẽ và đảm bảo độ tin cậy.

4. Kết luận và đề xuất:

Kết luận phù hợp với cách đặt vấn đề, đề xuất mang tính cải tiến và thực tiễn; kết luận có đóng góp mới mẻ, đề xuất sáng tạo và thuyết phục.

Kết luận phù hợp với cách đặt vấn đề, đề xuất mang tính cải tiến và thực tiễn; kết luận có đóng góp mới mẻ, đề xuất sáng tạo và thuyết phục.

5. Hình thức trình bày và bố cục báo cáo:

Văn phong nhất quán, bố cục hợp lý, cấu trúc rõ ràng, đúng định dạng mẫu; có tính hấp dẫn, thể hiện năng lực tốt, văn bản trau chuốt.

Văn phong nhất quán, bố cục hợp lý, cấu trúc rõ ràng, đúng định dạng mẫu; có tính hấp dẫn, thể hiện năng lực tốt, văn bản trau chuốt.

6. Kỹ năng chuyên nghiệp và tính sáng tạo:

Thể hiện các kỹ năng giao tiếp, kỹ năng làm việc nhóm, và các kỹ năng chuyên nghiệp khác trong việc thực hiện đề tài.

Thể hiện các kỹ năng giao tiếp, kỹ năng làm việc nhóm, và các kỹ năng chuyên nghiệp khác trong việc thực hiện đề tài.

7. Tài liệu trích dẫn

Tính trung thực trong việc trích dẫn tài liệu tham khảo; tính phù hợp của các tài liệu trích dẫn; trích dẫn theo đúng chỉ dẫn APA.

Tính trung thực trong việc trích dẫn tài liệu tham khảo; tính phù hợp của các tài liệu trích dẫn; trích dẫn theo đúng chỉ dẫn APA.

8. Đánh giá về sự trùng lặp của đề tài

Cần khẳng định đề tài có trùng lặp hay không? Nếu có, đề nghị ghi rõ mức độ, tên đề tài, nơi công bố, năm công bố của đề tài đã công bố.

Cần khẳng định đề tài có trùng lặp hay không? Nếu có, đề nghị ghi rõ mức độ, tên đề tài, nơi công bố, năm công bố của đề tài đã công bố.

9. Những nhược điểm và thiếu sót, những điểm cần được bổ sung và chỉnh sửa*

Không có nhược điểm và thiếu sót, những điểm cần được bổ sung và chỉnh sửa.

10. Nhận xét tinh thần, thái độ học tập, nghiên cứu của sinh viên

Không có nhận xét tinh thần, thái độ học tập, nghiên cứu của sinh viên.

Đề nghị của giảng viên hướng dẫn

Ghi rõ: "Báo cáo đạt/ không đạt yêu cầu của một khóa luận tốt nghiệp kỹ sư, và được phép/ không được phép bảo vệ khóa luận tốt nghiệp"

Đề nghị cho phép bảo vệ.

Tp. HCM, ngày ... tháng ... năm 20...

Người nhận xét

(Ký và ghi rõ họ tên)

Phan Văn Ca

PGS. TS. PHAN VĂN CA

* Giảng viên hướng dẫn có thể ghi thêm vào một câu

LỜI CẢM ƠN

Lời đầu tiên, nhóm thực hiện đề tài xin gửi lời cảm ơn chân thành đến thầy Phan Văn Ca – giảng viên khoa Điện – Điện tử, Trường ĐH Sư Phạm Kỹ Thuật Tp.HCM đã trực tiếp hướng dẫn, hỗ trợ và chia sẻ những kiến thức, kinh nghiệm quý báu trong quá trình nghiên cứu và thực hiện đề tài, chỉ ra những sai sót cũng như gợi ý những phương án thực hiện sao cho khả thi và dễ tiếp cận nhất.

Nhóm cũng xin gửi lời cảm ơn đến tất cả các quý thầy cô trong Khoa Điện - Điện tử đã tận tình giảng dạy, hỗ trợ, giúp đỡ cũng như đã tạo điều kiện về mặt vật chất lẫn tinh thần cho nhóm trong suốt những năm tháng học tập và rèn luyện ở trường.

Nhóm đề tài cũng xin cảm ơn các bạn sinh viên trong tập thể lớp 171190A, B, C đã nhiệt tình giúp đỡ, hỗ trợ tạo tập dữ liệu phục vụ cho đề tài này. Những đóng góp của các bạn luôn được nhóm tiếp cận và đánh giá cao.

Cuối cùng, nhóm thực hiện muốn gửi lời cảm ơn chân thành và sâu sắc đến gia đình, bạn bè, người thân đã luôn bên cạnh động viên và hỗ trợ nhiệt tình trong quá trình thực hiện và hoàn thành khóa luận tốt nghiệp.

Mặc dù đã cố gắng hoàn thiện đề tài nhưng do sự hạn chế về thời gian, kiến thức và chưa có nhiều kinh nghiệm thực tiễn nên đề tài chắc chắn không thể tránh khỏi những thiếu sót nhất định. Nhóm thực hiện rất mong nhận được những ý kiến đóng góp của quý thầy cô và các bạn để đề tài có thể trở nên hoàn thiện và có thể áp dụng vào thực tế để góp phần giải tỏa áp lực cho hàng triệu giáo viên trong tương lai không xa.

Nhóm thực hiện xin chân thành cảm ơn!

Tp. HCM, ngày tháng 07 năm 2022

Nhóm thực hiện đề tài
(Ký và ghi rõ họ tên)

Nguyễn Tấn Đạt Nguyễn Ngọc Khánh

TÓM TẮT

Ngày nay, cùng với sự phát triển vượt bậc của Khoa học – Kỹ thuật, tạo nên hàng loạt thay đổi quan trọng trong các lĩnh vực đời sống xã hội. Tuy nhiên với giáo dục, một lĩnh vực có ảnh hưởng vô cùng lớn đến thế hệ mai sau vẫn chưa được đầu tư một cách đúng mức. Ngoài những giờ dạy và học trên lớp học, giảng đường. Giáo viên vẫn phải bỏ ra quá nhiều thời gian không liên quan đến việc giáo dục, giảng dạy kiến thức. Cụ thể là việc nhập điểm số sau mỗi đợt kiểm tra và kì thi.

Đề tài: “Hệ thống nhập điểm tự động từ ảnh bài thi” được thực hiện dựa trên ý tưởng đó. Đây là một đề tài có ý nghĩa quan trọng trong việc giảm tải lượng công việc của giáo viên, giúp họ có thời gian, sức khỏe để tập trung vào những công việc liên quan đến giáo dục nhiều hơn.

Trong đề tài này, ngoài những kiến thức về xử lý ảnh, đề tài còn tập trung đi sâu nghiên cứu về nhận dạng chữ viết tay bằng mô hình CRNN kết hợp CTC Loss và các bước huấn luyện để tăng khả năng nhận dạng.

Hệ thống sử dụng Camera từ điện thoại để thu nhận hình ảnh của bài thi, truyền về cho máy tính để xử lý, nhận dạng thông tin của sinh viên và tự động nhập điểm số vào danh sách lớp.

MỤC LỤC

DANH MỤC HÌNH	X
DANH MỤC BẢNG BIỂU	XII
DANH MỤC CÁC TỪ VIẾT TẮT	XIII
CHƯƠNG 1 GIỚI THIỆU	1
1.1 ĐẶT VẤN ĐỀ	1
1.2 MỤC TIÊU ĐỀ TÀI.....	2
1.3 GIỚI HẠN ĐỀ TÀI.....	2
1.4 NỘI DUNG NGHIÊN CỨU	2
1.5 BỐ CỤC QUYỀN BÁO CÁO	3
CHƯƠNG 2 CƠ SỞ LÝ THUYẾT	4
2.1 TỔNG QUAN VỀ XỬ LÝ ẢNH	4
2.1.1 Tổng quan về xử lý ảnh	4
2.1.2 Quá trình xử lý ảnh	5
2.2 CƠ SỞ DỮ LIỆU	6
2.2.1 Bộ dữ liệu HANDS-VNOnDB2018	6
2.2.2 Bộ dữ liệu MNIST	6
2.3 TỔNG QUAN VỀ HỆ THỐNG NHẬN DẠNG CHỮ VIẾT TAY	8
2.3.1 Tổng quan	8
2.3.2 Kiến trúc tổng quát	9
2.3.3 Ứng dụng	10
2.4 PHÁT HIỆN VĂN BẢN (TEXT DETECTION)	10
2.4.1 Tổng quan	10
2.4.2 Kỹ thuật phân tách từ Scale Space (Scale Space Technique).....	11

2.5 NHẬN DẠNG VĂN BẢN (TEXT RECOGNITION).....	13
2.5.1 Tổng quan	13
2.5.2 Mô hình Convolutional Recurrent Neural Network (CRNN) kết hợp hàm mất mát Connectionist Temporal Classification (CTC Loss)	13
2.6 NGÔN NGỮ/PHẦN MỀM/ THƯ VIỆN	21
2.6.1 Ngôn ngữ Python	21
2.6.2 Visual Studio	21
2.6.3 Thư viện.....	21
2.6.4 Google Colab/Kaggle	22
CHƯƠNG 3 THIẾT KẾ HỆ THỐNG	23
3.1 YÊU CẦU HỆ THỐNG	23
3.2 MÔ HÌNH TỔNG QUÁT CỦA HỆ THỐNG	23
3.3 CÀI ĐẶT HỆ THỐNG CAMERA ĐIỆN THOẠI.....	24
3.4 LƯU ĐỒ GIẢI THUẬT CHO VIỆC HUẤN LUYỆN VÀ NHẬN DẠNG CHỮ VIẾT TAY	26
3.5 LƯU ĐỒ GIẢI THUẬT KHI NHẬN DẠNG VIDEO.....	28
3.6 GIAO DIỆN PHẦN MỀM DÀNH CHO NGƯỜI DÙNG	30
CHƯƠNG 4 KẾT QUẢ VÀ ĐÁNH GIÁ.....	34
4.1 KẾT QUẢ TỔNG QUAN	34
4.2 TIỀN XỬ LÝ	34
4.3 NHẬN DẠNG VỚI MÔ HÌNH CRNN KẾT HỢP CTC LOSS	38
4.4 KẾT QUẢ VÀ ĐÁNH GIÁ MÔ HÌNH	40
4.4.1 Kết quả và đánh giá trên hình ảnh	40
4.4.2 Kết quả và đánh giá trên Video	43
CHƯƠNG 5 KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN	46

5.1 KẾT LUẬN	46
5.2 HƯỚNG PHÁT TRIỂN	46
TÀI LIỆU THAM KHẢO	48

DANH MỤC HÌNH

Hình 2.1: Giới thiệu về xử lý ảnh	4
Hình 2.2 Quá trình xử lý ảnh	5
Hình 2.3: Ảnh chữ viết tay trong bộ dữ liệu HANDS – VNOnDB2018.....	6
Hình 2.4: Ảnh mẫu dữ liệu MNIST.....	7
Hình 2.5: Dữ liệu huấn luyện MSSV	7
Hình 2.6: Dữ liệu huấn luyện điểm số.....	7
Hình 2.7: Các hướng nghiên cứu chính của bài toán nhận diện chữ.....	8
Hình 2.8: Mô hình tổng quát về nhận dạng chữ viết tay	9
Hình 2.9: Dòng chữ trước khi xử lý	12
Hình 2.10: Dòng chữ xử lý thành các Blob màu trắng.....	12
Hình 2.11: Ảnh đã được phân tách từ.....	12
Hình 2.12: Mô hình kiến trúc CRNN	14
Hình 2.13: Mô hình kiến trúc CNN.....	15
Hình 2.14: Mối quan hệ giữa Feature Sequence, Feature Vector và Receptive Field	16
Hình 2.15: Mạng LSTM	17
Hình 2.16: Mạng Bi – LSTM	18
Hình 2.17: Đầu ra của Bi – LSTM theo mỗi Timestep	19
Hình 3.1: Mô hình tổng quát hệ thống nhập điểm tự động từ ảnh bài thi	24
Hình 3.2: Cách lắp đặt Camera.....	25
Hình 3.3: Hình ảnh minh họa cho ảnh từ ứng dụng IP Webcam	26
Hình 3.4: Lưu đồ giải thuật cho việc huấn luyện và nhận dạng chữ viết tay.....	26

Hình 3.5: Lưu đồ giải thuật khi nhận dạng trên Video.....	28
Hình 3.6: Giao diện người dùng	30
Hình 3.7: Hộp thoại hiển thị sau khi bấm nút Chọn tệp	31
Hình 3.8: Giao diện sau khi nhập đường dẫn	32
Hình 3.9: Giao diện hiển thị trong quá trình làm việc	32
Hình 3.10: Giao diện hiển thị trên Terminal	33
Hình 3.11: Hình ảnh hiển thị kết quả nhận dạng gồm Họ tên, MSSV và điểm số.....	33
Hình 4.1: Ảnh sau khi dùng kỹ thuật Image Alignment	35
Hình 4.2: Ảnh sau khi cắt	35
Hình 4.3: Ảnh các thông tin sau khi được cắt	36
Hình 4.4: Ảnh giấy thi sau khi qua phép Top Hat (trái) và phép Black Hat (phải)	36
Hình 4.5: Thông tin trước (trái) và sau (phải) khi tăng độ tương phản.....	37
Hình 4.6: Ảnh sau khi lấy ngưỡng OTSU Threshold.....	37
Hình 4.7: Ảnh họ tên đã được lấy ngưỡng, giảm nhiễu và xóa dòng kẻ ngang	37
Hình 4.8: Kết quả sau khi thực hiện phân tách từ bằng kỹ thuật Scale Space	38
Hình 4.9: Model nhận dạng được sử dụng trong đề tài	39
Hình 4.10: Quá trình huấn luyện Model họ tên (trái) và Model MSSV (phải)	40
Hình 4.11: Phân tách từ lỗi vì ghi chữ sát nhau	42
Hình 4.12: Ảnh nhận dạng lỗi vì không hiểu được từ đã xóa	42
Hình 4.13: Dấu “,” làm số 0 nhầm thành số 9	43
Hình 4.14: Dấu gạch dưới điểm số gây nhiễu	43
Hình 4.15: Số 2 ghi dính liền gây nhầm lẫn với số 9	43

Hình 4.16: Số 3 ghi quá tròn gây nhầm lẫn với số 8 hoặc số 0	43
Hình 4.17: Ảnh giấy thi mẫu	44
Hình 4.18: Minh họa kết quả nhận dạng trên Terminal	44

DANH MỤC BẢNG BIỂU

Bảng 4-1: Thông số cài đặt cho Model họ tên và Model số.....	38
Biểu đồ 4-1: Kết quả nhận dạng họ và tên MSSV.....	41
Biểu đồ 4-2: Kết quả nhận dạng điểm số	42

DANH MỤC CÁC TỪ VIẾT TẮT

Adaboost	Adaptive Boosting
AI	Artificial Intelligence
CNN	Convolutional Neural Network
CPU	Central Processing Unit
CRNN	Convolutional Recurrent Neural Network
CTC	Connectionist Temporal Classification
FPS	Frames Per Second
GPU	Graphics Processing Unit
LSTM	Long Short Term Memory Networks
MNIST	Modified National Institute of Standards and Technology
MSSV	Mã số sinh viên
OCR	Optical Character Recognition
OpenCV	Open Source Computer Vision Library
PNG	Portable Network Graphics
RNN	Recurrent Neural Network
SVM	Support Vector Machines
VGG16	Visual Geometry Group from Oxford
YOLO	You Only Look One

CHƯƠNG 1 GIỚI THIỆU

1.1 ĐẶT VẤN ĐỀ

Trong thời đại công nghiệp hóa hiện đại hóa hiện nay, đi cùng với sự phát triển của khoa học công nghệ, tiến lên cách mạng công nghiệp 4.0, nhu cầu tạo ra các hệ thống thông minh, tự động hóa công việc đang và sẽ rất được quan tâm và hữu ích cho cuộc sống hàng ngày. Hiện nay các ứng dụng từ AI (Artificial Intelligence) là một trong những chủ đề được nhắc đến nhiều nhất trong những năm gần đây, làm tiện lợi hơn cho những hoạt động thường ngày cũng như công việc. Ví dụ, có thể kể đến nhận dạng biển số xe bằng máy móc thay vì phải viết giấy như 5 – 10 năm về trước, vừa nhanh chóng, chính xác lại đảm bảo về mặt an ninh. Hay những hệ thống tự động báo khói, báo cháy thông qua Camera nhận diện hình ảnh giúp báo động và phản ứng nhanh hơn với những trường hợp khẩn cấp...

Tuy nhiên với giáo dục, một lĩnh vực có ảnh hưởng vô cùng lớn đến thế hệ mai sau vẫn chưa được đầu tư một cách đúng mức. Ngoài những giờ dạy và học trên lớp học, giảng đường. Giáo viên vẫn phải bỏ ra quá nhiều thời gian không liên quan đến việc giáo dục, giảng dạy kiến thức. Cụ thể là việc nhập điểm số sau mỗi đợt kiểm tra và kì thi.

Theo thống kê trung bình, mỗi giáo viên trung học phải chấm khoảng 2000 bài kiểm tra và bài thi mỗi học kỳ. Với gần 2000 bài kiểm tra, khối lượng công việc khi phải nhập điểm cho học sinh cũng đã rất lớn. Mỗi lần chấm bài xong, giáo viên phải rà soát lại bài, tính tổng điểm, dò tên học sinh để nhập điểm rồi lại rà soát kiểm tra lại rất mất thời gian, chưa kể ảnh hưởng đến sức khỏe sau khi làm việc trong thời gian dài dẫn đến có thể nhập sai điểm cho học sinh. Hiện nay đã có các thiết bị giúp chấm và nhập điểm tự động từ những bài thi trắc nghiệm, tuy nhiên về bài thi tự luận thì vẫn chưa có hệ thống nào thực hiện vấn đề đó một cách cụ thể.

Trên cơ sở đó, ở đề tài “Hệ thống nhập điểm tự động từ ảnh bài thi” nhóm thực hiện muốn tạo ra một công cụ phần nào đó giúp đỡ các giáo viên sau khi chấm điểm xong có thể nhập điểm vào File Excel một cách tự động.

1.2 MỤC TIÊU ĐỀ TÀI

Nắm được cách thức vận hành của quá trình nhận dạng chữ viết tay, các mô hình và thuật toán liên quan.

Trích xuất, nhận dạng được ba thông tin chính trên giấy thi: Họ và tên sinh viên, MSSV, điểm của bài thi.

Sau khi nhận dạng xong có thể tự động nhập điểm vào File Excel.

Hiện thị giao diện để giáo viên có thể làm việc, tương tác với hệ thống.

1.3 GIỚI HẠN ĐỀ TÀI

Trong quá trình thực hiện, nhóm xin đưa ra các giới hạn mà đề tài này gặp phải:

- Mô hình vẫn còn xử lý chậm vì chưa tối ưu hóa được giải thuật và cả những hạn chế về phần cứng của máy tính hay đường truyền Internet
- Có khá ít bộ dữ liệu để huấn luyện các mô hình nhận dạng chữ viết tay Tiếng Việt làm cho mô hình vẫn chưa thật sự đạt kết quả tốt trong việc nhận dạng
- Khi sử dụng vẫn còn cần giới hạn người dùng trong kiểu viết, cách viết, cách trình bày... để mô hình có thể hoạt động tốt nhất có thể
- Chưa có khả năng kết nối, tải dữ liệu, nhập điểm lên Internet
- Hệ thống chỉ có khả năng làm việc với các danh sách lớp ở dạng File Excel
- Giao diện còn đơn giản, chưa đẹp mắt

1.4 NỘI DUNG NGHIÊN CỨU

Đề tài “Hệ thống nhập điểm tự động từ ảnh bài thi” gồm các nội dung cụ thể như sau:

- Tìm hiểu tổng quan về hệ thống nhận dạng chữ viết tay và các thành phần chính

- Nghiên cứu, xây dựng mô hình giải thuật phát hiện chữ viết tay, nhận dạng chữ viết tay
- Nghiên cứu, xây dựng phần mềm tự động nhập điểm vào danh sách lớp File Excel
- Xây dựng giao diện cho người dùng
- Hiện thực hóa bằng mô hình trực quan

1.5 BỐ CỤC QUYỀN BÁO CÁO

Để người đọc dễ hiểu và nắm bắt thông tin rõ ràng hơn thì quyền báo cáo này được chia thành 5 chương cụ thể như sau:

Chương 1: Giới thiệu

Trong chương này sẽ trình bày về các nhu cầu ứng dụng thực tế của đề tài, mục tiêu đề tài muốn hướng đến, giới hạn và nội dung nghiên cứu của đề tài.

Chương 2: Cơ sở lý thuyết

Nội dung chương bao gồm dữ liệu cho mô hình, lý thuyết về cách thức phát hiện, nhận dạng văn bản, các thuật toán và các phần mềm, ngôn ngữ được áp dụng trong đề tài.

Chương 3: Thiết kế hệ thống

Từ những cơ sở lý thuyết có được, trình bày về mô hình và cách thức hoạt động của hệ thống.

Chương 4: Kết quả và đánh giá

Trình bày chi tiết kết quả đạt được cùng với những nhận xét, đánh giá về toàn bộ hệ thống.

Chương 5: Kết luận và hướng phát triển

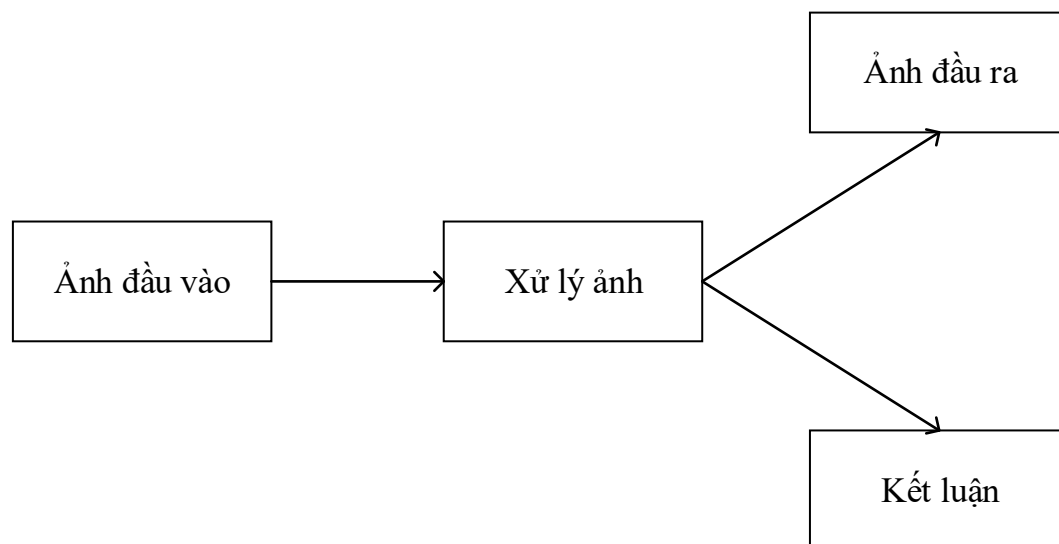
Trình bày những kết luận đã được rút ra trong suốt quá trình thực hiện đề tài cùng với hướng cải tiến, phát triển đề tài sao cho phù hợp và có thể áp dụng với thực tế hơn.

CHƯƠNG 2 CƠ SỞ LÝ THUYẾT

2.1 TỔNG QUAN VỀ XỬ LÝ ẢNH

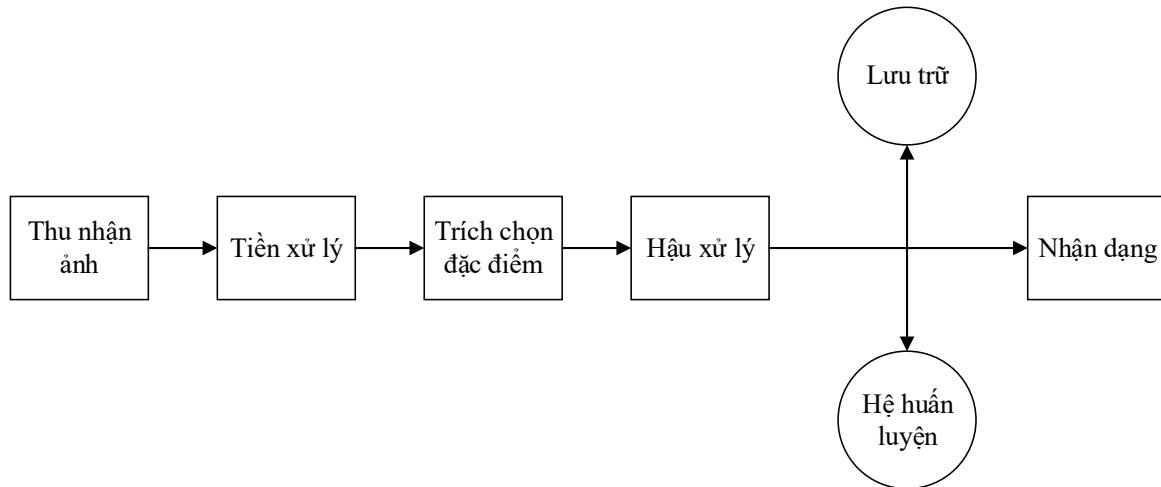
2.1.1 Tổng quan về xử lý ảnh

Xử lý ảnh là quá trình xử lý số tín hiệu của hình ảnh, là một đối tượng nghiên cứu của lĩnh vực thị giác máy. Trong đó, ảnh ban đầu sẽ được biến đổi và xử lý với các đặc tính của ý muốn người sử dụng sao cho ảnh sau khi xử lý đạt kết quả như mong muốn. Sau quá trình xử lý, kết quả đưa ra có thể là một ảnh có đặc điểm khác so với ảnh ở đầu vào hoặc có thể là một kết luận (Hình 2.1). Xử lý ảnh đã và đang phát triển mạnh mẽ trong thời gian gần đây, được áp dụng trong các lĩnh vực khác nhau như: Quân sự, quốc phòng, y tế, văn hóa, kinh tế... Hiện nay, có 4 lĩnh vực chính liên quan đến quá trình xử lý ảnh: nén ảnh, truy vấn ảnh, xử lý nâng cao chất lượng ảnh và nhận dạng ảnh. Ở phạm vi đề tài này sẽ tìm hiểu về mảng nhận dạng hình ảnh [1].



Hình 2.1: Giới thiệu về xử lý ảnh

2.1.2 Quá trình xử lý ảnh



Hình 2.2 Quá trình xử lý ảnh

- Thu nhận ảnh: Công đoạn được cho là mang tính quyết định của cả quá trình xử lý ảnh. Ảnh ban đầu sẽ được thu nhận bằng các thiết bị như sensor, camera hoặc máy quét... tiếp theo các tín hiệu thu nhận được sẽ được số hóa. Tùy thuộc vào đặc tính của đối tượng xử lý mà chọn ra các thiết bị thu nhận ảnh khác nhau. Các thông số như: độ phân giải, dung lượng bộ nhớ, chất lượng màu sắc của ảnh và tốc độ việc thu nhận hình ảnh của thiết bị có tầm quan trọng lớn [1].
- Quá trình tiền xử lý: Trong công đoạn này, hình ảnh sẽ được loại bỏ nhiễu, khử bóng, độ lệch, chuyển đổi màu sắc, tăng độ tương phản, cắt bỏ những thông tin không quan trọng trong ảnh và làm rõ các đặc trưng của mẫu chữ viết... Quá trình được thực hiện bởi các bộ lọc nhằm mục đích làm tăng chất lượng hình ảnh đầu vào để cho các công đoạn tiếp theo có thể thuận tiện xử lý [1].
- Trích chọn đặc điểm (hoặc trích chọn đặc trưng): Đặc trưng ảnh ở đây chính là đặc trưng nội dung của hình ảnh, hay nói cách khác là nội dung thực sự cần chọn trong ảnh. Nội dung cần chọn trong ảnh là chữ viết, kết cấu, hình dạng, các đặc trưng cục bộ... hoặc có thể là bất cứ thông tin nào có từ chính nội dung trong ảnh. Phạm vi đề tài bao gồm phát hiện, nhận diện chữ viết tay cần thiết trong ảnh [1].

- Hậu xử lý: Công đoạn này thực hiện nhiệm vụ xử lý các đặc điểm đã được trích chọn, các đặc điểm có thể biến đổi hoặc loại bỏ sao cho phù hợp với các kỹ thuật cụ thể được sử dụng trong hệ huấn luyện [1].
- Hệ huấn luyện và lưu trữ: Nhiệm vụ của công đoạn huấn luyện dựa trên dữ liệu đã đưa vào và lưu trong khối lưu trữ [1].
- Nhận dạng: Chuyển đổi các thông tin hình ảnh sang dạng ASCII để làm việc, nhận dạng các kí tự [1].

2.2 CƠ SỞ DỮ LIỆU

2.2.1 Bộ dữ liệu HANDS-VNOnDB2018

HAND – VNOnDB2018 (ICFHR2018 Competition on Vietnamese Online Handwritten Text Recognition Database) là bộ dữ liệu cung cấp 1.146 đoạn văn viết tay bằng tiếng Việt gồm 7.296 dòng, hơn 480.000 nét và hơn 380.000 kí tự được viết bởi 200 người Việt Nam. Vì bộ dữ liệu này là các từ ở dạng File .inkml nên sẽ được xử lý để trở thành ảnh .PNG phục vụ cho việc nhận dạng [5].

Đây là bộ dữ liệu nhóm thực hiện dùng để huấn luyện mô hình nhận dạng Họ và tên.

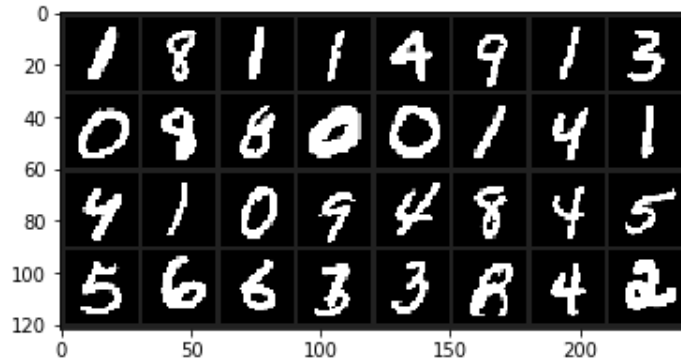


Hình 2.3: Ảnh chữ viết tay trong bộ dữ liệu HANDS – VNOnDB2018

2.2.2 Bộ dữ liệu MNIST

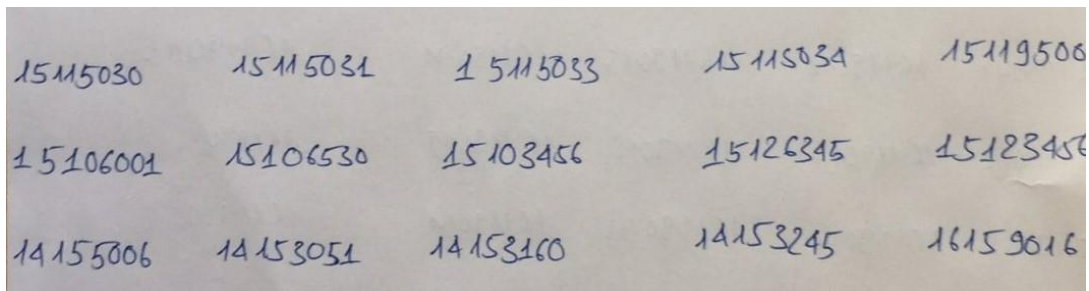
Trong bộ dữ liệu này, mỗi hình ảnh là một điểm ảnh đen trắng chứa một số được viết tay từ 0 đến 9 có kích thước là 28x28. Bộ dữ liệu vô cùng lớn với khoảng 60.000 ảnh dữ liệu huấn luyện, 10.000 ảnh dữ liệu kiểm thử và được sử dụng phổ biến trong các thuật toán dạng ảnh.

Đây là bộ dữ liệu nhóm thực hiện dùng để tự sinh chuỗi số phục vụ cho việc huấn luyện và nhận dạng MSSV và điểm số

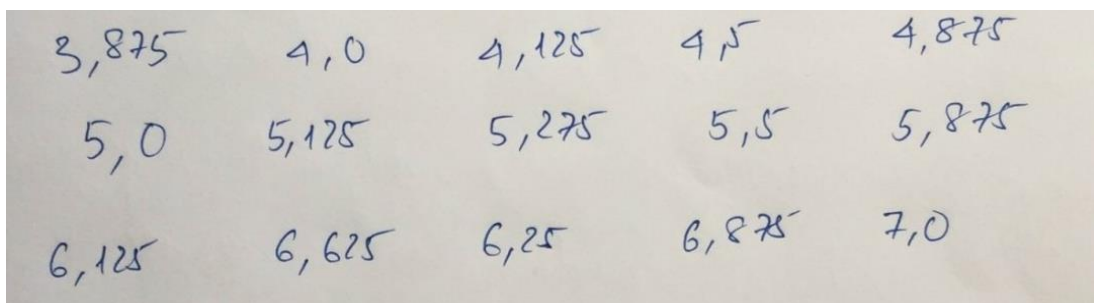


Hình 2.4: Ảnh mẫu dữ liệu MNIST

Ngoài ra, nhóm thực hiện còn tạo bộ dữ liệu thực (tự thu gom), gồm 400 hình ảnh số được viết tay bởi 4 người. Bao gồm 300 ảnh chuỗi có độ dài 8 chữ số nhằm giúp tăng cường việc nhận dạng MSSV và 100 ảnh điểm số từ 0 đến 10 với độ chia nhỏ nhất 0.125 nhằm tăng cường khả năng nhận dạng điểm số. Các hình ảnh được viết trên giấy A4, sau đó sử dụng các phương pháp tiền xử lý để trích tách hình ảnh huấn luyện.



Hình 2.5: Dữ liệu huấn luyện MSSV



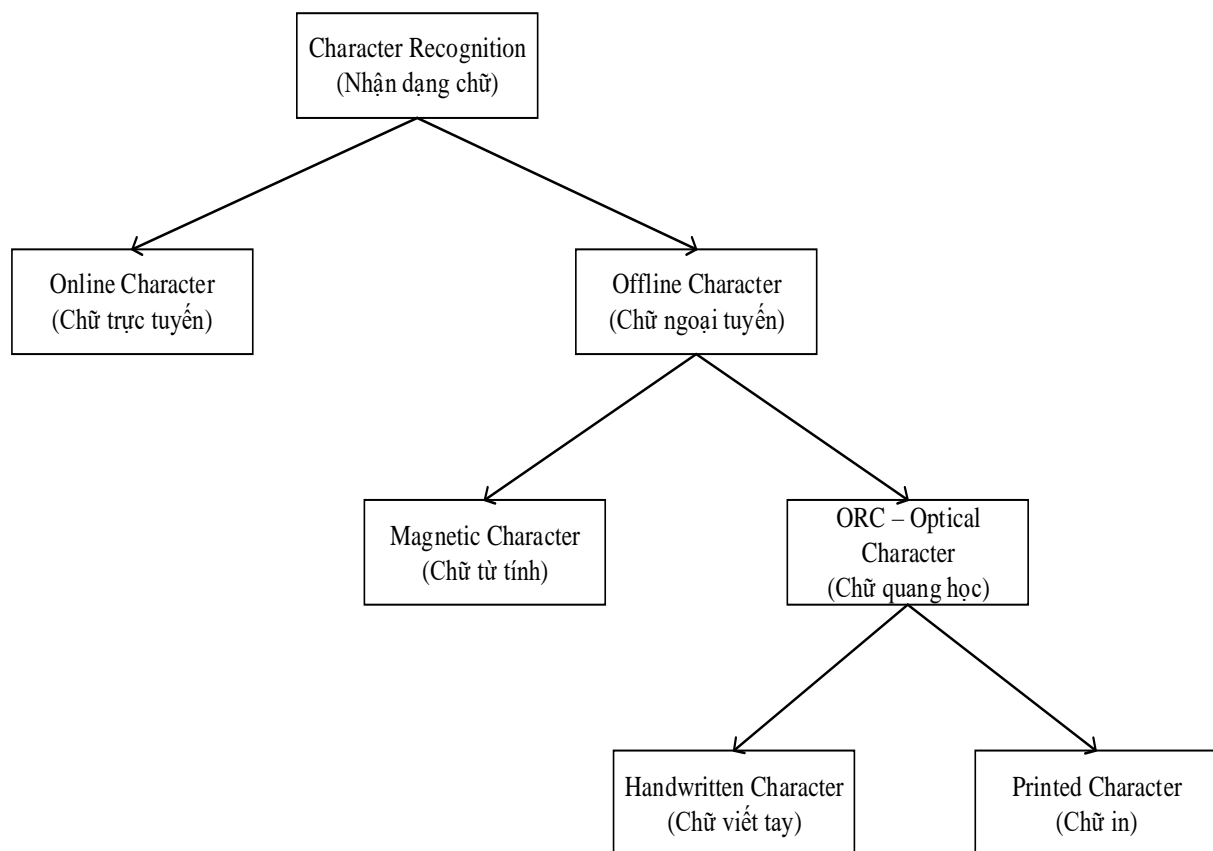
Hình 2.6: Dữ liệu huấn luyện điểm số

2.3 TỔNG QUAN VỀ HỆ THỐNG NHẬN DẠNG CHỮ VIẾT TAY

2.3.1 Tổng quan

Nhận dạng chữ viết là một lĩnh vực được rất nhiều nhà nghiên cứu quan tâm trong những năm gần đây và cho đến nay lĩnh vực này cũng đã đạt được nhiều thành tựu to lớn cả về mặt lý thuyết lẫn về phần ứng dụng vào thực tế. Ở mức độ chung nhất, lĩnh vực nhận dạng chữ viết được chia làm hai nhánh: nhận dạng chữ trực tuyến (Online Character Recognition) và nhận dạng chữ ngoại tuyến (gián tuyến) (Offline Character Recognition).

Với dữ liệu đầu vào là ảnh văn bản nên nhận dạng chữ in và nhận dạng chữ viết tay ngoại tuyến còn được gọi chung là nhận dạng chữ quang học ORC (Optical Character Recognition).



Hình 2.7: Các hướng nghiên cứu chính của bài toán nhận diện chữ

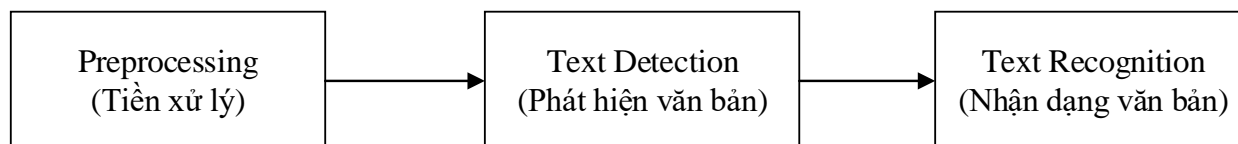
Nhận dạng chữ viết tay trực tuyến là bài toán nhận dạng song song với quá trình chữ viết được thực hiện, được thực hiện bằng cách lưu lại các thông tin về nét chữ như nét viết, thứ tự, hướng và tốc độ của nét viết trong quá trình viết chữ, cần có các thiết bị chuyên dùng để ghi lại quá trình di chuyển của nét viết như điểm bắt đầu, những điểm trên mặt phẳng mà nét bút đi qua và điểm kết thúc. Trong khi đó nhận dạng chữ viết tay ngoại tuyến là có dữ liệu đầu vào là ảnh văn bản, được quét vào máy nên việc nhận dạng được chữ viết có độ khó cao hơn nhiều so với nhận dạng chữ viết tay trực tuyến.

Gần đây, bài toán về nhận dạng chữ viết tay trực tuyến và nhận dạng chữ in đã được giải quyết gần như trọn vẹn. Tuy nhiên, bài toán nhận dạng chữ viết tay ngoại tuyến vẫn đang là thách thức to lớn đối với các nhà nghiên cứu.

Ngoài ra, đối với tiếng Việt cũng có những khó khăn riêng, tuy có sử dụng bộ chữ cái gần tương tự với bộ chữ La Tinh đang phổ biến trên các diễn đàn học thuật. Tiếng Việt còn có tập các dấu câu huyền, sắc, hỏi, ngã nặng và các chữ cái khác biệt như ă, â, ô, ơ, u, ê. Nếu khả năng nhận dạng không được tốt, máy sẽ dễ bị nhầm lẫn về mặt ngữ nghĩa. Đồng thời, hiện tại cũng có khá ít bộ dữ liệu để huấn luyện các mô hình về nhận dạng chữ viết tay tiếng Việt.

2.3.2 Kiến trúc tổng quát

Một hệ thống nhận dạng chữ viết tay thông thường sẽ gồm các phần sau:



Hình 2.8: Mô hình tổng quát về nhận dạng chữ viết tay

Giai đoạn tiền xử lý thường bao gồm các bước xử lý các điều kiện khác nhau trong ảnh, xóa nền phức tạp khỏi hình ảnh, loại bỏ nhiễu và làm rõ các đặc trưng của ảnh, chuẩn hóa ảnh để hỗ trợ cho các giai đoạn tiếp theo.

Giai đoạn phát hiện giúp xác định được vị trí có văn bản, hỗ trợ cho việc tách văn bản ra khỏi nền ảnh (những thông tin không cần thiết, dư thừa) hỗ trợ cho việc nhận dạng được

văn bản tốt hơn. Giai đoạn này thường đã bao gồm các bước phân tách hàng (Line Segmentation) và phân tách từ (Word Segmentation).

Nhận dạng văn bản giúp chuyển đổi các thông tin từ hình ảnh sang dạng ASCII để làm việc, hỗ trợ cho các phần mềm/ứng dụng theo sau.

2.3.3 Ứng dụng

Cho đến ngày nay, bài toán nhận diện chữ viết tay đã được ứng dụng nhiều trong thực tế, có thể kể đến là áp dụng vào lĩnh vực an ninh, hành chính, dịch thuật... với một số ứng dụng điển hình bao gồm:

- Chuyển đổi văn bản viết tay thành văn bản kỹ thuật số
- Dịch thuật tự động
- Xác thực chữ kí
- Trích xuất các trường thông tin cần thiết từ văn bản hành chính như công văn, quyết định...

Qua đó có thể thấy rằng việc nhận dạng chữ viết tay có ứng dụng rất quan trọng và có tiềm năng to lớn để áp dụng rộng rãi trong nhiều tác vụ trong cuộc sống.

2.4 PHÁT HIỆN VĂN BẢN (TEXT DETECTION)

2.4.1 Tổng quan

Nhiệm vụ chính của phát hiện văn bản là tạo ra các Bounding Box (hình tứ giác) bao xung quanh các văn bản có trong ảnh để tạo điều kiện cho giai đoạn nhận diện văn bản [12].

Có 3 kỹ thuật chính để phát hiện văn bản:

- Dựa vào Texture (Texture Based Method): Phương pháp này coi văn bản như một thông tin có thuộc tính riêng so với hình ảnh và nền. Phù hợp cho các hình ảnh có văn bản nằm ngang và có khả năng giải quyết tốt vấn đề tỉ lệ (Scale). Ít bị ảnh hưởng bởi các yếu tố như nhiễu, màu sắc, nền của hình ảnh. Tuy nhiên phương pháp này có thời gian tính toán khá lâu.

- Dựa vào Component (Component Based Method): Phương pháp này phân chia hình ảnh thành các cụm thông tin dựa vào biên, màu sắc để chọn ra những vùng có khả năng chứa văn bản... sau đó sẽ được lọc bớt nhờ các bộ phân loại hoặc đặt ngưỡng. Kỹ thuật gặp khá nhiều khó khăn với các văn bản có nhiều màu sắc khác nhau, nhiễu hoặc các văn bản trên hình ảnh nghiêng, vẹo. Bù lại, phương pháp này có khả năng xử lý nhanh và giải quyết vấn đề tỉ lệ và phong chữ tốt.
- Hybrid Method: Phương pháp này sử dụng kết hợp cả 2 kỹ thuật ở trên, sử dụng các thuộc tính Geometrical và Gradient để xác định khoảng, vùng có chứa văn bản trong ảnh.

Có rất nhiều phương pháp như Adaboost (Adaptive Boosting), SVM (Support Vector Machines), CNN (Convolutional Neural Network), YOLO (You Only Look Once) đã được chứng minh là có hiệu quả trong việc phát hiện văn bản với độ chính xác hơn 90%. Tuy nhiên các kiến trúc trên rất phức tạp khiến cho việc xử lý mất nhiều thời gian nên khó có thể áp dụng vào các công nghệ thời gian thực.

Ngoài ra, thuật toán EAST (Efficient and Accurate Scene Text Detection Pipeline) cũng là một trong những thuật toán phát hiện văn bản vượt trội về độ chính xác và tốc độ xử lý.

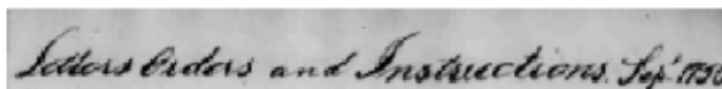
Tuy nhiên ở phạm vi đề tài, các thông tin như “Họ và tên”, “MSSV”, “điểm” ở trên giấy thi luôn nằm ở vị trí xác định, hoàn toàn có thể trích tách thủ công thông qua giai đoạn tiền xử lý và sử dụng các phương pháp phân tách từ (Word Segmentation) giúp thời gian xử lý ngắn hơn, ít tốn tài nguyên và thuận tiện cho việc nhận dạng chữ viết.

2.4.2 Kỹ thuật phân tách từ Scale Space (Scale Space Technique)

Kỹ thuật phân tách từ Scale Space được đưa ra bởi Manmatha và Srimal lần đầu tiên vào năm 1999 [6], tuy vậy kỹ thuật này vẫn đưa ra kết quả nhanh chóng, chính xác, rất phù hợp với phạm vi và mục tiêu phát hiện, phân tách chữ viết của đề tài.

Kỹ thuật sẽ đi qua 3 bước chính:

- Tiền xử lý: Chuyển ảnh về ảnh xám, lọc nhiễu và xóa đi các dòng kẻ ngang, dọc, chuẩn hóa hình ảnh.
- Phân tách dòng (Line Segmentation): Sử dụng phương pháp tách dùng lược đồ sáng (Histogram) với ý tưởng là ở phần có mực viết đi qua thì khu vực đó sẽ tối hơn.
- Phân tách từ (Word Segmentation): Trong một từ sẽ có những chữ cái được viết liền hoặc viết tách. Sau đó sử dụng bộ lọc Gauss theo cả trục ngang và dọc để làm giãn nở các chữ cái để tạo thành các Blob. Khi đó các từ sẽ nhìn như các đốm màu trắng.

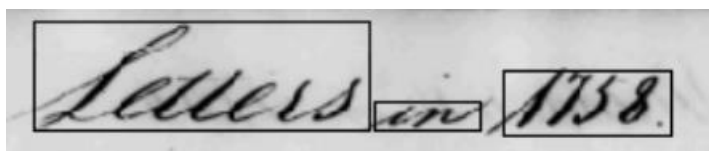


Hình 2.9: Dòng chữ trước khi xử lý



Hình 2.10: Dòng chữ xử lý thành các Blob màu trắng

Cuối cùng là vẽ đường bao quanh các Blob này và trích xuất ảnh của từng từ



Hình 2.11: Ảnh đã được phân tách từ

- **Ưu điểm:**

Phương pháp xử lý nhanh, phù hợp với các văn bản có cấu trúc. Sử dụng ít tài nguyên, phù hợp với phạm vi đề tài.

- **Khuyết điểm:**

Các thông số, tỉ lệ để tách các từ trong cùng một hàng không dính liền nhau khá phức tạp, yêu cầu thử nghiệm nhiều.

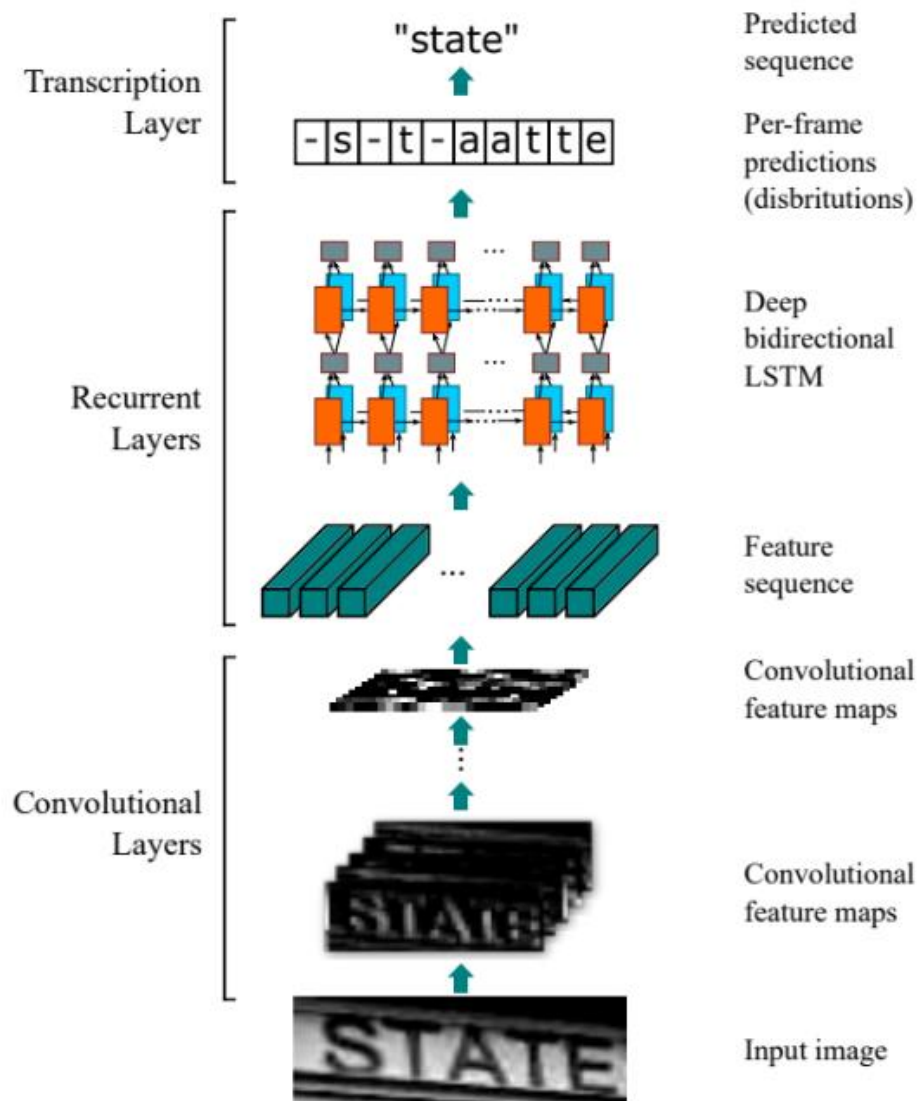
2.5 NHẬN DẠNG VĂN BẢN (TEXT RECOGNITION)

2.5.1 Tổng quan

Nhận dạng văn bản bao gồm các phương pháp như đối sánh mẫu, mô hình Markov ẩn, mạng thần kinh, máy véc tơ lựa... Tuy nhiên, những phương pháp này đã có phần lỗi thời, không còn cho ra độ chính xác cao. Những năm gần đây, người ta sử dụng các kiến trúc song song hoặc lai ghép để tạo ra các mô hình với độ chính xác cao và thời gian xử lý ngắn. Có thể kể đến các mô hình như: mô hình CRNN (Convolutional Recurrent Neural Network) kết hợp CTC Loss (Connectionist Temporal Classification), mô hình CNN kết hợp SVM, hay sử dụng Tesseract OCR [3] ...

2.5.2 Mô hình Convolutional Recurrent Neural Network (CRNN) kết hợp hàm mất mát Connectionist Temporal Classification (CTC Loss)

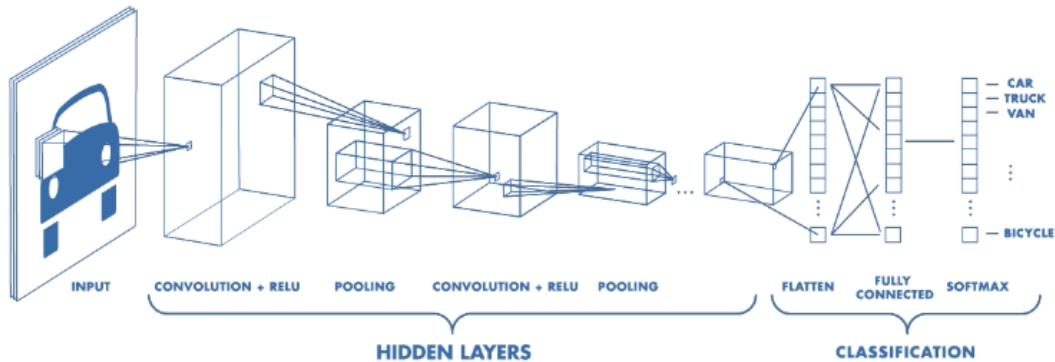
Mạng thần kinh tái phát liên tục (CRNN) là sự kết hợp của các mạng CNN, RNN và CTC Loss cho phép các tác vụ nhận dạng chuỗi với độ dài bất kì như nhận dạng văn bản, phân loại hành động, phân loại Video [3] ...



Hình 2.12: Mô hình kiến trúc CRNN

Kiến trúc mô hình CRNN gồm 3 giai đoạn chính là mạng thần kinh tích chập (Convolutional Neural Network), mạng hai chiều LSTM (Bidirectional Long Short Term Memory Network) và lớp phiên mã (Transcription Layer).

2.5.2.1 Mạng thần kinh tích chập (Convolutional Neural Network)



Hình 2.13: Mô hình kiến trúc CNN

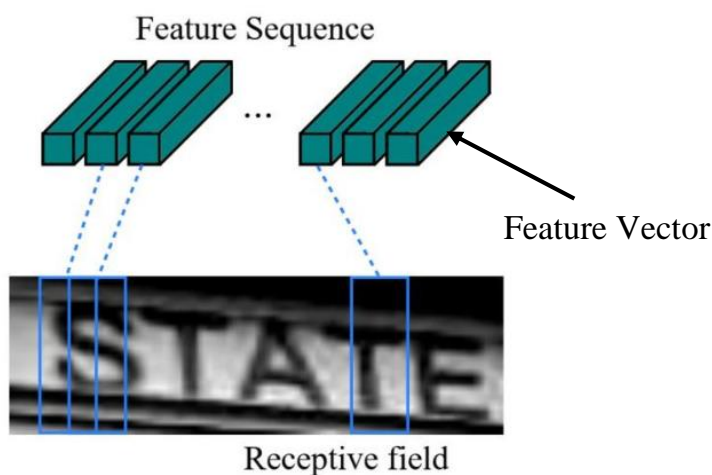
Mạng thần kinh tích chập (CNN) gồm tập hợp các lớp cơ bản [10]:

- Lớp tích chập (Convolution Layer): là lớp đầu tiên và cũng là lớp quan trọng nhất của mô hình CNN. Lớp này có chức năng chính là phát hiện ra các đặc trưng có tính không gian hiệu quả gồm màu sắc, góc, cạnh hoặc Texture của ảnh mà không bị ảnh hưởng bởi các phép xoay, tỉ lệ...
- Lớp phi tuyến (Nonlinear Layer): lớp này giúp hạn chế tình trạng Vanishing Gradient, có thể phát hiện ra các quan hệ phức tạp của dữ liệu. Có các hàm kích hoạt phi tuyến như Sigmoid, Tanh nhưng phổ biến nhất thời điểm này là hàm ReLU.
- Lớp Pooling: lớp này thường được dùng giữa các lớp tích chập, giúp giảm kích thước của dữ liệu nhưng vẫn giữ được các thuộc tính quan trọng của dữ liệu đó giúp giảm tải công việc tính toán trong mô hình.
- Lớp Fully Connected: Ở lớp này, dữ liệu đầu ra của lớp đứng trước sẽ được chuyển thành Vector và đưa vào một lớp được kết nối như một mạng thần kinh. Sau đó sử dụng hàm Sigmoid hoặc Softmax để phân loại đầu ra.

Các lớp trên được liên kết với nhau theo một trình tự nhất định.

Mạng CRNN được xây dựng bằng cách bỏ đi lớp Fully Connected Layers và giữ lại lớp tích chập (Convolutional Layer) và lớp Max Pooling từ kiến trúc CNN nhằm trích xuất các đặc trưng từ hình ảnh ở đầu vào.

Từ ảnh đầu vào, sau khi qua các lớp tích chập sẽ tạo ra được các Feature Map và từ đó một chuỗi các Feature Vector được tạo ra và được gọi là Feature Sequence để đưa vào mạng RNN (Recurrent Neural Network). Mỗi Feature Vector có thể tương ứng với một vùng Receptive Field ở ảnh gốc.



Hình 2.14: Mối quan hệ giữa Feature Sequence, Feature Vector và Receptive Field

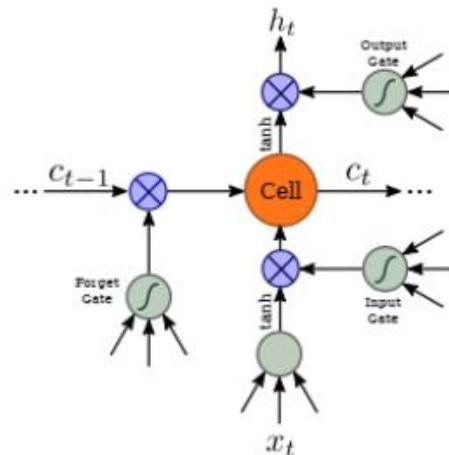
Tất cả hình ảnh phải được chuẩn hóa về cùng một chiều cao nhưng độ dài có thể khác nhau. Như ở mô hình CNN thuần túy, mọi hình ảnh được chuẩn hóa về cùng một kích cỡ cả chiều cao và chiều dài để có cùng số chiều ở đầu vào nên không phù hợp với thông tin dạng chuỗi vì độ dài của mỗi chuỗi văn bản là khác nhau.

2.5.2.2 Mạng hai chiều LSTM (Bidirectional Long Short Term Memory Network)

Lớp này giúp dự đoán các kí tự từ mỗi Feature Vector và các mối quan hệ giữa các kí tự có ở lớp tích chập trước.

Mạng LSTM (Long Short Term Memory Network) là một biến thể khác của RNN. Tuy mạng RNN có thể dùng cơ chế truyền ngược (Back Propagation) cho phép huấn luyện cả mạng RNN và CNN cùng một lúc và hoạt động trên các chuỗi có độ dài tùy ý nhưng

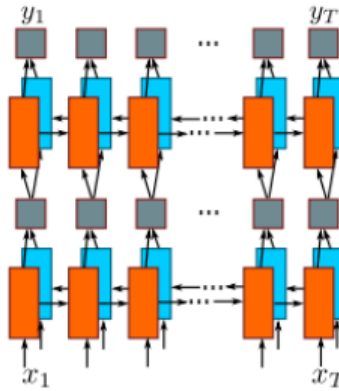
mạng RNN sẽ gặp vấn đề khi ghi nhớ các kí tự ở quá xa vì mất mát đạo hàm (Vanishing Gradient) nên cần sử dụng mạng LSTM.



Hình 2.15: Mạng LSTM

Thành phần chính của mạng LSTM là Cell State, đường nằm ngang C_{t-1} đến C_t như một dạng băng truyền, cho phép thông tin có thể đi thẳng từ đầu đến cuối mạng. Đồng thời LSTM còn có khả năng thêm vào hoặc bỏ bớt các thông tin Cell State thông qua các cấu trúc cổng.

Mạng LSTM là một mạng có định hướng (Directional Network), có nghĩa là mạng chỉ sử dụng thông tin từ quá khứ (ký tự đứng trước), nhưng với chuỗi văn bản, cả ký tự đứng trước và ký tự đứng sau đều có giá trị về mặt nội dung, vì vậy người ta đã sử dụng kết hợp hai mạng LSTM (truyền tới và truyền ngược) hay còn được gọi là mạng hai chiều LSTM (Bi – LSTM). Ngoài ra, một số mô hình còn kết hợp nhiều mạng Bi – LSTM để nâng cao độ chính xác.



Hình 2.16: Mạng Bi – LSTM

2.5.2.3 Lớp phiên mã (*Transcription Layer*)

Lớp phiên mã giúp chuyển đổi mỗi khung hình (Per – Frame Prediction) được tạo ra bởi mạng Bi – LSTM ở trước thành chuỗi kết quả cuối cùng.

Có hai chế độ phiên mã:

- Phiên mã không có từ vựng (Lexicon Free): Dự đoán chuỗi thuần túy dựa vào xác suất cao nhất có được.
- Phiên mã từ vựng (Lexicon Based): Dự đoán chuỗi dựa vào một từ điển có sẵn.

2.5.2.4 Hàm mất mát *Connectionist Temporal Classification (CTC Loss)*

Sau khi qua lớp Bi – LSTM, ta có được xác suất xuất hiện của các kí tự ứng với mỗi Timestep được biểu diễn dưới dạng một ma trận.

Từ ma trận thu được, sau đó sẽ:

- Tính toán hàm mất mát (Loss Function) để huấn luyện mạng CRNN.
- Giải mã ma trận để được chuỗi văn bản ở đầu ra.

Việc sử dụng CTC Loss sẽ giúp hoàn thành hai công việc trên

- 1) Mã hóa nhãn dữ liệu huấn luyện (Encoding Ground Truth)

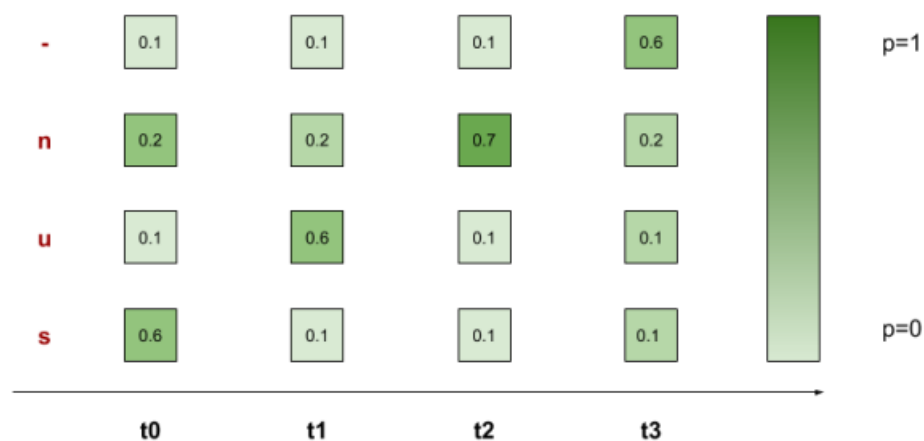
CTC Loss có thể giải quyết vấn đề này một cách rất thông minh. Cụ thể, hàm sẽ thử tất cả các Alignment của Ground Truth và tính xác suất của tổng tất cả các Alignment đó. Mạng CRNN sẽ học từ CTC Loss để tự đưa ra những Alignment này bằng cách thêm kí tự trống “-” và lặp lại bất kì kí tự nào trong Ground Truth.

Giả sử Ground Truth là từ sun và có mô hình LSTM dự toán 4 Timesteps. Thì những Alignment đúng của Ground Truth sẽ là: -sun, s-un, su-n, sun-, suun, ssun, sunn. Nhưng không thể là: s-n- (thiếu kí tự u), usn (chỉ có 3 trên 4 kí tự), us-n (không giải mã ra được từ sun).

Đối với những từ có 2 kí tự liên tục giống nhau, sẽ thêm kí tự “-” ở giữa để tạo ra một Alignment đúng. Ví dụ với từ odd, các Alignment đúng có thể là: “good” → “--ggo-oo--dd-”, nhưng không thể là “--g--odd-” “god” → “--ggoooo--dd” hoặc “-god”.

2) Tính CTC Loss

Với mỗi Ground Truth sẽ có nhiều Alignment, bất kì Alignment nào được dự đoán cũng đều là một dự đoán đúng. Do đó, hàm Loss cần tối ưu tổng của tất cả các Alignment.



Hình 2.17: Đầu ra của Bi – LSTM theo mỗi Timestep

Với từ “sun”, có 7 Alignment đúng cho ví dụ trên. Do đó theo mô hình, xác suất của từ “sun” có thể xuất hiện là:

$$\begin{aligned}
p('sun') &= p('-sun') + p('s-un') + p('su-n') + p('sun-') + p('ssun') + p('suun') + p('sunn') \\
&= (0.1 \times 0.1 \times 0.1 \times 0.2) + (0.6 \times 0.1 \times 0.1 \times 0.2) + (0.6 \times 0.6 \times 0.1 \times 0.2) + \\
&\quad (0.6 \times 0.6 \times 0.7 \times 0.6) + (0.6 \times 0.1 \times 0.1 \times 0.2) + (0.6 \times 0.6 \times 0.1 \times 0.2) + \\
&\quad (0.6 \times 0.6 \times 0.7 \times 0.2) \\
&= 0.2186
\end{aligned}$$

Từ đó, suy ra hàm Loss sẽ là: $1 - p('sun') = 1 - 0.2186 = 0.7814$

Mục tiêu của việc huấn luyện là điều chỉnh cho các thông số trong mạng CRNN, điều chỉnh lại sao cho xác suất đầu ra của từng kí tự theo Timestep cho $p('sun')$ tiệm cận và hàm Loss sẽ về gần bằng 0 để đạt hiệu quả cao nhất.

3) Giải mã văn bản (Decoding Text)

Có nhiều kỹ thuật để có thể giải mã văn bản như Beam Search, Word Beam Search, Best Path... Nhưng chúng đều có ý tưởng tương tự nhau. Sau khi có được xác suất có thể xuất hiện của các kí tự ứng với mỗi Timestep được biểu diễn dưới dạng một ma trận ở đầu ra của lớp Bi – LSTM. Tìm ra Alignment đi qua các ký tự có xác suất cao nhất trong từng Timestep. Từ đó, mô hình sẽ gộp những ký tự lặp lại liên tiếp nhau thành một ký tự và xóa hết tất cả ký tự “-” để đưa ra chuỗi dự đoán cuối cùng. Ví dụ với Alignment “--g-oo-odd-”, bỏ ký tự liên tiếp, thu được “-g-o-o-d-”, tiếp tục bỏ ký tự trắng, vậy từ thu được là “good”.

2.5.2.5 Ưu và nhược điểm

- Ưu điểm:

Mô hình CRNN kết hợp CTC Loss có thể học trực tiếp từ ảnh của các từ (Word) mà không nhất thiết phải thêm các bước tiền xử lý như nhị phân hóa ảnh (Binarization), phân đoạn ký tự (Character Segmentation), Component Localization.... Có thể áp dụng cho chuỗi văn bản có độ dài bất kì giúp nâng cao hiệu suất nhận dạng với thông tin chuỗi. Thông số tính toán ít hơn so với mạng thần kinh tích chập sâu (Deep Convolutional Neural Network).

- Nhược điểm:

Mô hình không được hoạt động tốt với các chuỗi ngẫu nhiên. Không có lịch sử ngữ cảnh cho “mô hình ngôn ngữ” trong các trường hợp đó.

2.6 NGÔN NGỮ/PHẦN MỀM/ THƯ VIỆN

2.6.1 Ngôn ngữ Python

Python là ngôn ngữ lập trình hướng đối tượng bậc cao, được sử dụng trong lập trình từ những dòng lệnh đơn giản nhất cho đến những thuật toán phức tạp nhất. Với thiết kế dễ nhìn, dễ đọc, cấu trúc rõ ràng mà lại mang lại sự chặt chẽ trong ngôn ngữ, nhanh, mạnh và có mặt ở trên mọi hệ điều hành. Với tính đơn giản, nhất quán, linh hoạt và có cộng đồng rộng lớn, việc lập trình bằng ngôn ngữ Python đã trở nên rộng rãi trên khắp thế giới.

Ngoài ra, với lĩnh vực trí tuệ nhân tạo, Python có một kho cộng nghệ phong phú bao gồm rất nhiều thư viện cho lĩnh vực này.

2.6.2 Visual Studio

Visual Studio là một trong những công cụ hỗ trợ lập trình nổi tiếng nhất hiện nay của Microsoft và chưa có một phần mềm nào có thể thay thế được.

Với các điểm vượt trội so với các phần mềm khác như:

- Visual Studio hỗ trợ lập trình trên nhiều nền tảng ngôn ngữ khác nhau từ C/C++, C#, F#, HTML, CSS, Visual Basic, JavaScript và với phiên bản Visual Studio 2015 trở đi còn có hỗ trợ lập trình trên ngôn ngữ Python.
- Visual Studio sở hữu giao diện dễ sử dụng, thân thiện với người dùng.
- Visual Studio hỗ trợ khả năng gỡ rối (Debug) dễ dàng và hiệu quả.
- Visual Studio cho phép cài đặt các thư viện dễ dàng nhờ Nuget, tích hợp được nhiều ứng dụng khác.

2.6.3 Thư viện

Ở phạm vi đề tài này, nhóm đã sử dụng các thư viện sau:

- Open CV cho xử lý ảnh
- Keras, TensorFlow, Scikit – Learn cho máy học
- NumPy cho phép phân tích dữ liệu và tính toán khoa học có hiệu năng cao
- Matplotlib cho phép làm việc với hình ảnh, đồ thị có tọa độ
- Pandas cho phép làm việc với các dạng File: Excel và csv
- Tkinter cho phép tạo giao diện tương tác với người dùng

Và một số thư viện khác phục vụ cho việc nhập xuất thư mục, tác vụ thêm...

2.6.4 Google Colab/Kaggle

Khi viết một chương trình sử dụng Framework về AI/Deep Learning như Keras, TensorFlow chúng ta có thể sử dụng bất kì Python IDE nào như PyCharm, Jupyter Notebook hoặc Visual Studio để lập trình. Nhưng khi huấn luyện mô hình, yêu cầu hệ thống phải mạnh, có tốc độ xử lý cao (thường dựa trên GPU).

Google Colab hay Kaggle cho phép chúng ta lập trình Python thông qua trình duyệt Web, phù hợp với Data Analysis, Machine Learning mà không yêu cầu về cài đặt hay cấu hình máy tính vẫn có thể sử dụng tài nguyên máy tính như CPU tốc độ cao cho đến GPUs và TPUs một cách miễn phí. Tuy vậy, việc bị giới hạn về thời gian huấn luyện (sử dụng GPUs) buộc người dùng phải thêm thắt các thao tác xử lý một cách hợp lý.

CHƯƠNG 3 THIẾT KẾ HỆ THỐNG

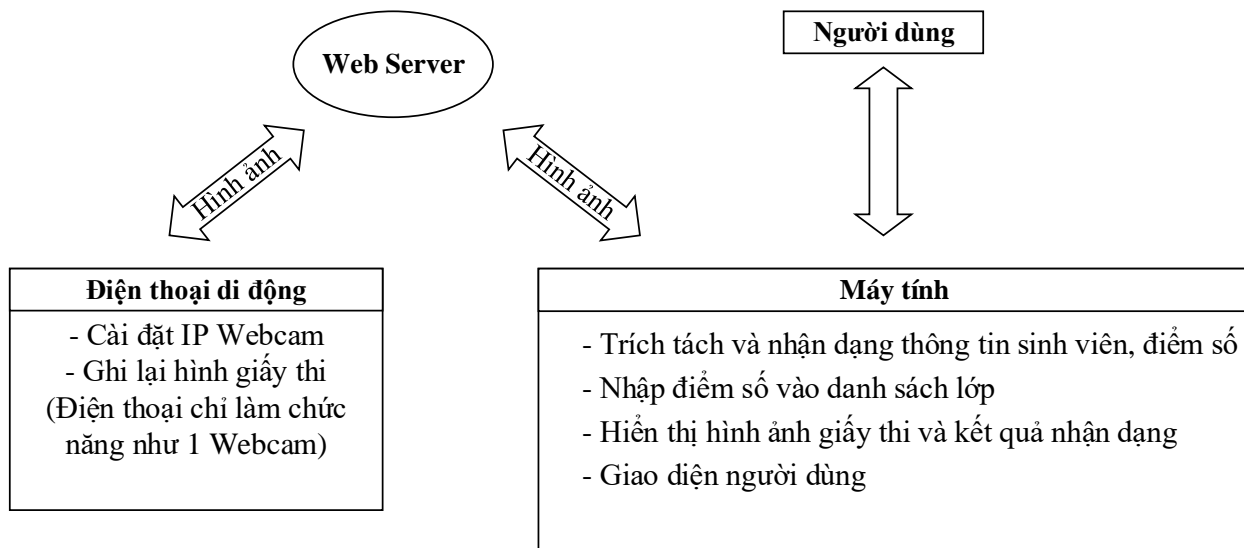
3.1 YÊU CẦU HỆ THỐNG

Hệ thống được phát triển dựa trên mục tiêu đã được đề ra và có những chức năng chính như sau:

- Chức năng chụp ảnh, truyền dữ liệu và phân tích, xử lý dữ liệu đầu vào: Sau khi người dùng thiết lập các thông số cho Camera, cố định vị trí Camera và nơi đặt giấy thi để hệ thống nhận diện. Hình ảnh giấy thi sẽ được Camera chụp lại và gửi đến máy tính một cách liên tục để xử lý thông qua Web Server.
- Chức năng nhận dạng kí tự viết tay: Máy tính sau khi nhận được dữ liệu hình ảnh từ Web Server sẽ tiến hành trích xuất, xử lý và nhận dạng các thông tin đầu vào như: “Họ và Tên”, “MSSV”, “Điểm bài thi” để phục vụ cho việc tự động nhập điểm của hệ thống.
- Chức năng quản lý của hệ thống: Thông tin sau khi được hệ thống nhận dạng sẽ được tự động ghi lại vào danh sách lớp dưới dạng File Excel để dễ dàng quản lí. Sẽ có giao diện giao tiếp giữa người dùng và hệ thống giúp người dùng dễ dàng sử dụng. Giao diện người dùng sẽ có các chức năng như : nhập địa chỉ của Web Server chứa Video quay lại hình ảnh giấy thi, nhập đường dẫn file danh sách lớp đang có trong máy tính . Ngoài ra, hệ thống sẽ hiển thị trực tiếp hình ảnh giấy thi cùng với kết quả đã nhận dạng để người dùng dễ dàng theo dõi.

3.2 MÔ HÌNH TỔNG QUÁT CỦA HỆ THỐNG

Mô hình hoạt động của hệ thống nhập điểm tự động từ ảnh bài thi có dạng như sau:



Hình 3.1: Mô hình tổng quát hệ thống nhập điểm tự động từ ảnh bài thi

Hệ thống gồm các thành phần chính là: Camera từ điện thoại và hệ thống xử lý là máy tính. Hệ thống này sử dụng Camera của điện thoại di động thông qua ứng dụng IP Webcam để ghi lại hình ảnh bài thi và truyền dữ liệu ghi được lên Web Server tại địa chỉ IP của Camera thông qua Wifi. Ở đây, điện thoại có chức năng như một Webcam thông thường, không sử dụng các tính năng khác từ điện thoại.

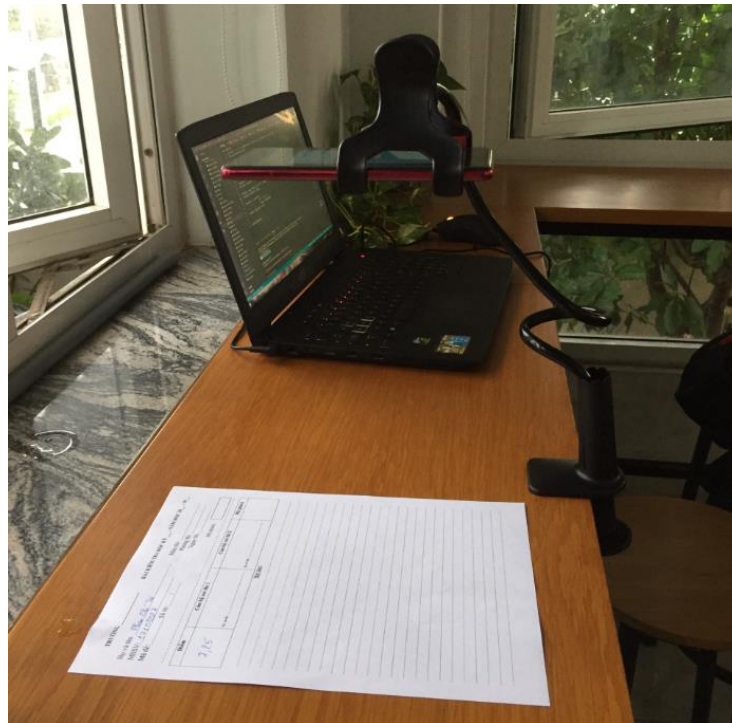
Hệ thống từ máy tính sẽ liên tục gửi yêu cầu đến Web Server để lấy ảnh đầu vào từ Camera về để trích xuất thông tin, nhận dạng và nhập điểm vào danh sách lớp. Người dùng sẽ phải cài đặt địa chỉ Web Server trên ứng dụng IP Webcam và nhập địa chỉ đó cùng đường dẫn đến danh sách sinh viên ở giao diện người dùng trên phần mềm.

Để tiết kiệm thời gian cho người sử dụng, có thể cùng lúc chấm điểm bài thi và để hệ thống nhận diện thông tin cần thiết để nhập điểm. Khi người dùng chấm xong bài thi thì sẽ đặt bài đã chấm đó dưới Camera điện thoại, trong thời gian chấm điểm thì hệ thống sẽ nhận diện liên tục đến khi khớp thông tin trong danh sách lớp thì điểm thi sẽ được nhập vào.

3.3 CÀI ĐẶT HỆ THỐNG CAMERA ĐIỆN THOẠI

Cài đặt và cố định Camera điện thoại hướng về nơi mà người dùng để giấy thi đã chấm điểm, nên để ảnh nơi có ánh sáng tốt để ảnh đầu vào được ghi chụp lại có chất

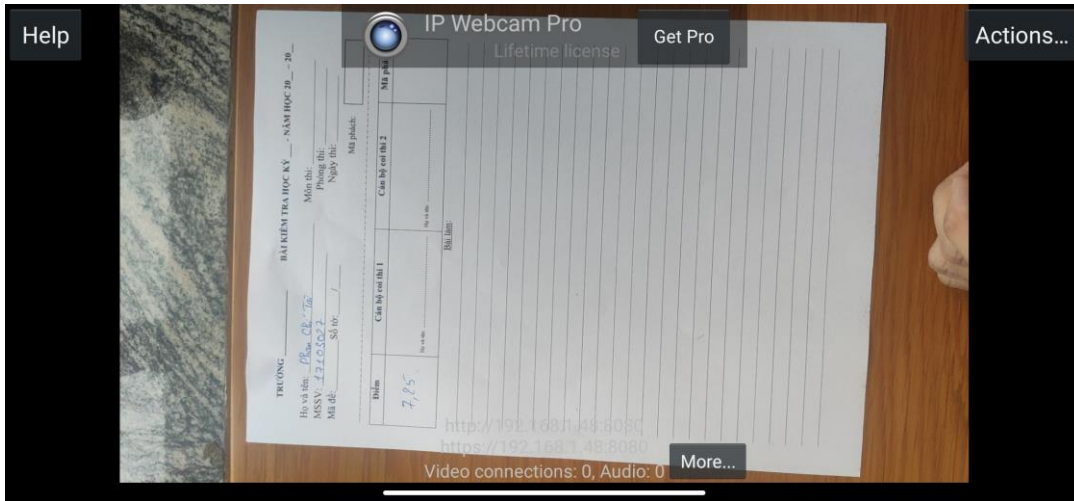
lượng cao nhằm giảm thiểu tối đa việc nhiễu thông tin khi nhận diện. Điện thoại nên được lắp chính diện với bài thi nhất có thể để Camera dễ dàng bắt trọn khung ảnh bài thi. Nên hạn chế những nơi có ánh sáng quá mạnh hoặc quá yếu có thể làm chói, nhòe, khuất bóng hay vị trí gò ghề, lồi lõm có thể gây ảnh hưởng đến quá trình nhận diện thông tin bài thi và xử lý nhận dạng.



Hình 3.2: Cách lắp đặt Camera

Để thiết lập hệ thống Camera, nhóm đã sử dụng điện thoại Xiaomi Mi 9T. Ngoài ra, điện thoại cũng đã cài đặt ứng dụng IP Webcam giúp Camera điện thoại trở thành một IP Camera cho hệ thống. Những thông số được cài đặt cho Camera của điện thoại như sau:

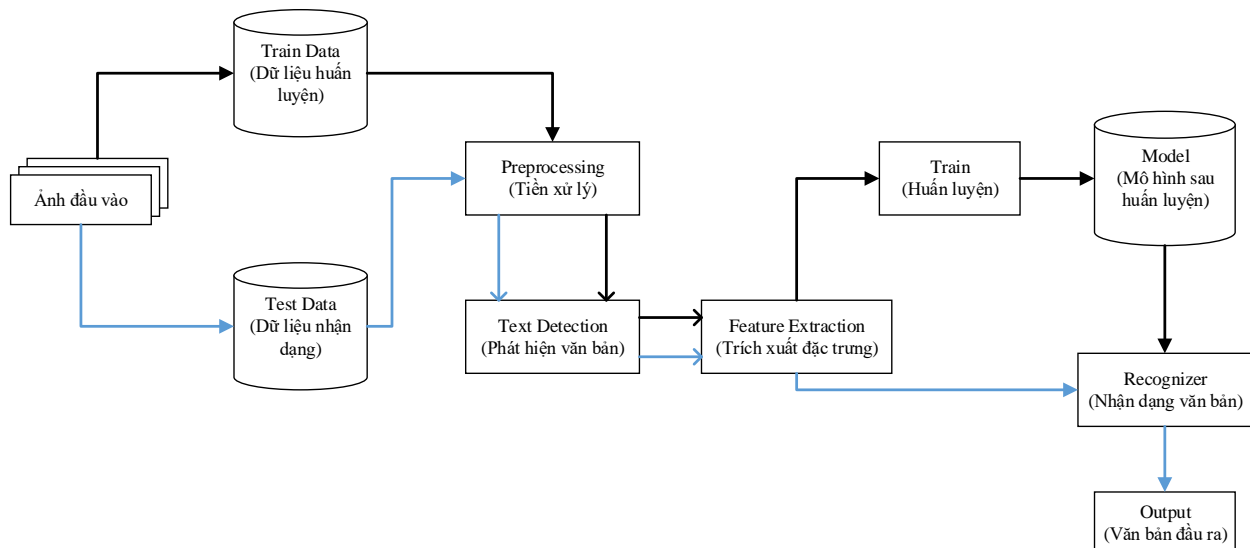
- Tốc độ khung hình: 20 fps
- Độ phân giải Video: 1920x1080 Pixels
- Khoảng cách từ Camera đến mặt bàn: 25 - 30cm
- Địa chỉ URL: <http://192.168.1.48:8080/video>
- Flash Mode: Default (sử dụng Always use Flash khi môi trường thiếu sáng)



Hình 3.3: Hình ảnh minh họa cho ảnh từ ứng dụng IP Webcam

Điện thoại và máy tính nên được sử dụng chung mạng Wifi. Sau đó hệ thống sẽ lấy hình ảnh từ địa chỉ URL trên Web Server để tiến hành xử lý nhận dạng rồi đưa ra kết quả.

3.4 LƯU ĐỒ GIẢI THUẬT CHO VIỆC HUẤN LUYỆN VÀ NHẬN DẠNG CHỮ VIẾT TAY



Hình 3.4: Lưu đồ giải thuật cho việc huấn luyện và nhận dạng chữ viết tay

Hình ảnh bài thi sau khi được ghi hình bởi Camera sẽ bao gồm cả bề mặt bài thi và các vấn đề nhiễu thông tin như nền ảnh, bụi bẩn, nhiễu hạt, đổ bóng, nếp gấp... Ngoài ra khi bài thi bị nghiêng, biến dạng, sẽ dẫn đến việc trích xuất và xử lý thông tin gặp khó khăn. Hình ảnh bài thi sẽ được hệ thống tiền xử lý với quy trình như sau:

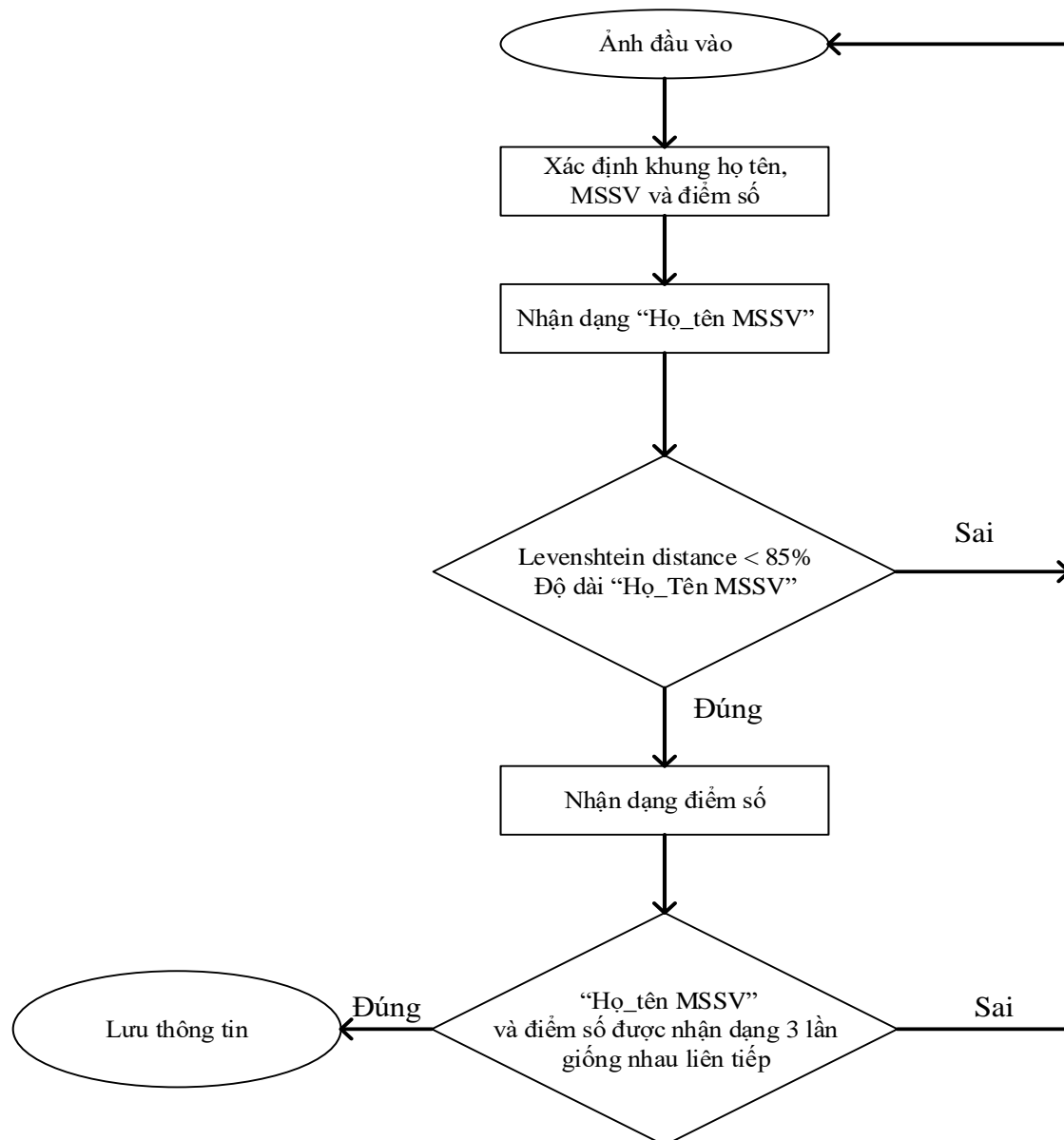
- Hình ảnh đầu vào sẽ được chuyển về dạng ảnh xám giúp màu sắc của hình ảnh, nền ảnh không ảnh hưởng đến hệ thống và làm giảm tối thiểu chi phí tính toán trong việc xử lý ảnh và nhận dạng.
- Hình ảnh bài thi sẽ được phân biệt với nền ảnh bằng phương pháp Image Alignment (trích xuất các đặc trưng riêng của bài thi và so sánh với một tờ giấy thi mẫu). Bài thi sau khi trích xuất xong đặc trưng sẽ được tách ra khỏi nền ảnh và xoay chỉnh lại cho giống với bài thi mẫu.
- Cắt tách nửa bên dưới của bài thi để giảm chi phí tính toán và tăng tốc độ xử lý cho hệ thống. Xác định và cắt thành các ảnh nhỏ của thông tin “Họ và tên”, “MSSV”, “Điểm số” một cách chính xác.
- Tối ưu độ tương phản với các phép toán xử lý ảnh kỹ thuật số là: Top Hat và Black Hat nhằm tăng khả năng nhận dạng thông tin với những ảnh bị nhòe. Lấy ngưỡng, làm mờ, giảm nhiễu và xóa hàng kẻ ngang (khi trích xuất họ tên và MSSV).

Với hình ảnh họ và tên sinh viên, nếu chỉ sử dụng quá trình tiền xử lý ở trên thôi là chưa đủ, phải cần thêm công đoạn phân tách từ (Word Segmentation) để một lần nữa giảm thiểu chi phí tính toán và tăng độ chính xác khi nhận dạng họ tên. Để xử lý vấn đề trên. Nhóm đã sử dụng kỹ thuật phân tách các từ ngay trên hình ảnh thu được bằng kỹ thuật Scale Space.

Để xử lý nhận dạng văn bản: Hình ảnh sẽ được trích xuất các đặc trưng để huấn luyện mô hình, sau khi huấn luyện sẽ được sử dụng để nhận dạng văn bản trong thực tế. Để hoàn thiện việc nhận dạng, hệ thống sử dụng hai mô hình riêng biệt để huấn luyện và nhận dạng. Một mô hình nhận dạng chữ viết tay, một mô hình khác dùng để

nhận dạng kí tự số viết tay. Hệ thống nhận dạng gồm mô hình CRNN kết hợp CTC Loss cùng lớp Attention. Trong quá trình làm việc, dữ liệu đầu ra của lớp Bi – LSTM sẽ được phiên mã Lexicon Based với nguồn dữ liệu tra cứu là danh sách lớp đầu vào để cho ra kết quả cuối cùng.

3.5 LƯU ĐỒ GIẢI THUẬT KHI NHẬN DẠNG VIDEO



Hình 3.5: Lưu đồ giải thuật khi nhận dạng trên Video

Bước đầu của giải thuật là khi Video đầu vào sẽ được xử lý chia thành từng Frame ảnh để nhận dạng riêng biệt, ở đây nhóm sẽ cho cứ mỗi 5 Frame sẽ nhận dạng một lần, việc đó sẽ giúp giảm tải xử lý và tăng độ chính xác cho quá trình nhận dạng ảnh.

Bước tiếp theo là xác định khung chứa điểm và từ đó trích xuất vị trí các thông tin của bài thi để nhận dạng, những Frame ảnh mà khung chứa điểm khó xác định, hoặc không đáp ứng được các tiêu chí về số cạnh và diện tích sẽ bị hủy bỏ xử lý và quay lại bước đầu tiên.

Trường thông tin “Họ_tên MSSV” mặc dù có thể được tìm bằng Lexicon Search cho ra kết quả với độ chính xác khá cao. Tuy nhiên, nếu kết quả cho ra cuối cùng khi tìm với thuật toán Lexicon Search có sự khác biệt quá nhiều so với kết quả Model đưa ra thì cũng sẽ được cho là không đáng tin cậy. Cụ thể, nếu Levenshtein Distance giữa kết quả sau khi tìm bằng Lexicon Search và kết quả gốc của Model có sự tương đồng trong khoảng 80% đến 85% là đủ để đảm bảo tính cân bằng giữa tốc độ nhận dạng và độ chính xác của hệ thống.

Nếu các thông tin như Họ tên và MSSV có sự hỗ trợ lẫn nhau, hay có thể dựa vào danh sách lớp để tăng độ chính xác, thì việc nhận dạng điểm số trên từng hình ảnh riêng lẻ lại gặp nhiều khó khăn bởi vì không có nguồn thông tin hỗ trợ để đảm bảo độ chính xác của kết quả nhận dạng. Nhưng đối với khi nhận dạng ở Video thì sẽ có nhiều thông tin hơn ở nhiều Frame ảnh khác nhau. Ở đây nhóm sẽ đặt ra một quy tắc là khi mô hình nhận dạng được giống nhau về cả thông tin điểm số và “Họ_tên MSSV” liên tiếp trên 3 lần, khi đó thông tin sẽ được coi là chính xác và không cần cập nhật lại điểm số mỗi khi gặp lại bài thi của sinh viên đó.

Khi văn bản chứa các thông tin của sinh viên trên bài thi đã được nhận dạng. Các thông tin ấy sẽ được lưu vào biến hệ thống chứ không phải liên tục cập nhật thông tin vào File Excel. Khi người dùng bấm nút Lưu trên giao diện, hệ thống sẽ tìm vị trí để nhập điểm trong danh sách lớp (dựa vào thông tin Họ và tên, MSSV) và nhập điểm đồng loạt cho tất cả các sinh viên vào danh sách lớp.

3.6 GIAO DIỆN PHẦN MỀM DÀNH CHO NGƯỜI DÙNG



The screenshot shows a web application window titled "HỆ THỐNG NHẬP ĐIỂM TỪ ẢNH BÀI THI". The interface is for HCMUTE (Ho Chi Minh City University of Technology and Education) and the Faculty of Electrical Engineering (KHOA ĐIỆN - ĐIỆN TỬ). It features the HCMUTE logo on the left and a circular emblem on the right. The main heading is "HỆ THỐNG NHẬP ĐIỂM TỰ ĐỘNG TỪ ẢNH BÀI THI". Below this, there is a section for "Nhập đường dẫn:" (Enter path:). It includes a "Video/URL:" label followed by a text input field, and a "Danh sách lớp:" (Class list:) label followed by a "Chọn tệp" (Select file) button and an ellipsis "...". At the bottom, there are three buttons: "Bắt đầu" (Start), "Lưu" (Save), and "Thoát" (Exit).

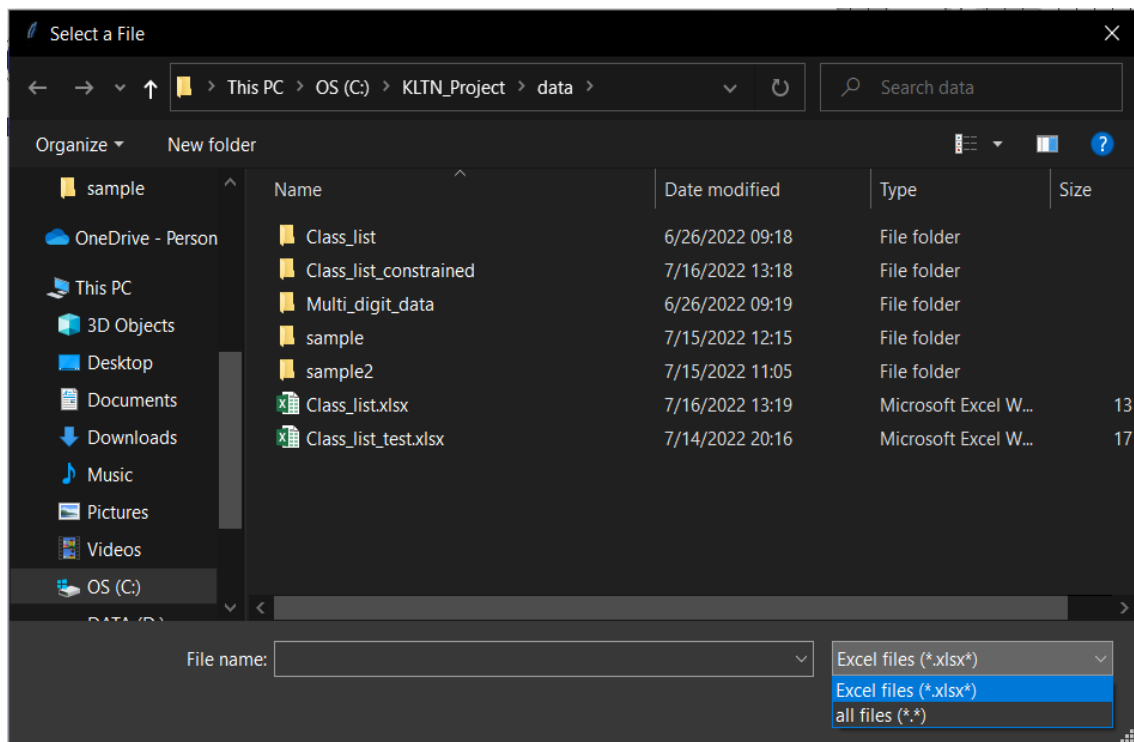
Hình 3.6: Giao diện người dùng

Giao diện phần mềm dành cho người dùng trước khi nhận dạng bao gồm những phần sau:

- Text Box để người dùng nhập địa chỉ URL của Camera trên Web Server
- Nút nhấn Chọn tệp để lấy đường dẫn đến File Excel danh sách lớp có trong máy tính
- Nút nhấn Bắt đầu để bắt đầu quá trình nhận diện

- Nút nhấn Lưu để ghi lại thông tin vào danh sách lớp
- Nút nhấn Thoát để thoát khỏi chương trình

Sau khi nhấn nút Chọn tệp sẽ hiện ra hộp thoại để người dùng có thể chọn File danh sách lớp. Người dùng có thể nhấp chuột vào Combo Box ở góc dưới bên phải của hộp thoại để chọn hiển thị tất cả các File (All Files) hoặc chỉ hiển thị các File Excel. Sau đó người dùng nhấn đúp chuột vào File Excel danh sách lớp cần làm việc.



Hình 3.7: Hộp thoại hiển thị sau khi bấm nút Chọn tệp

Khi đã chọn danh sách lớp xong, đường dẫn sẽ hiện ra bên cạnh nút Chọn tệp để người dùng tiện theo dõi:

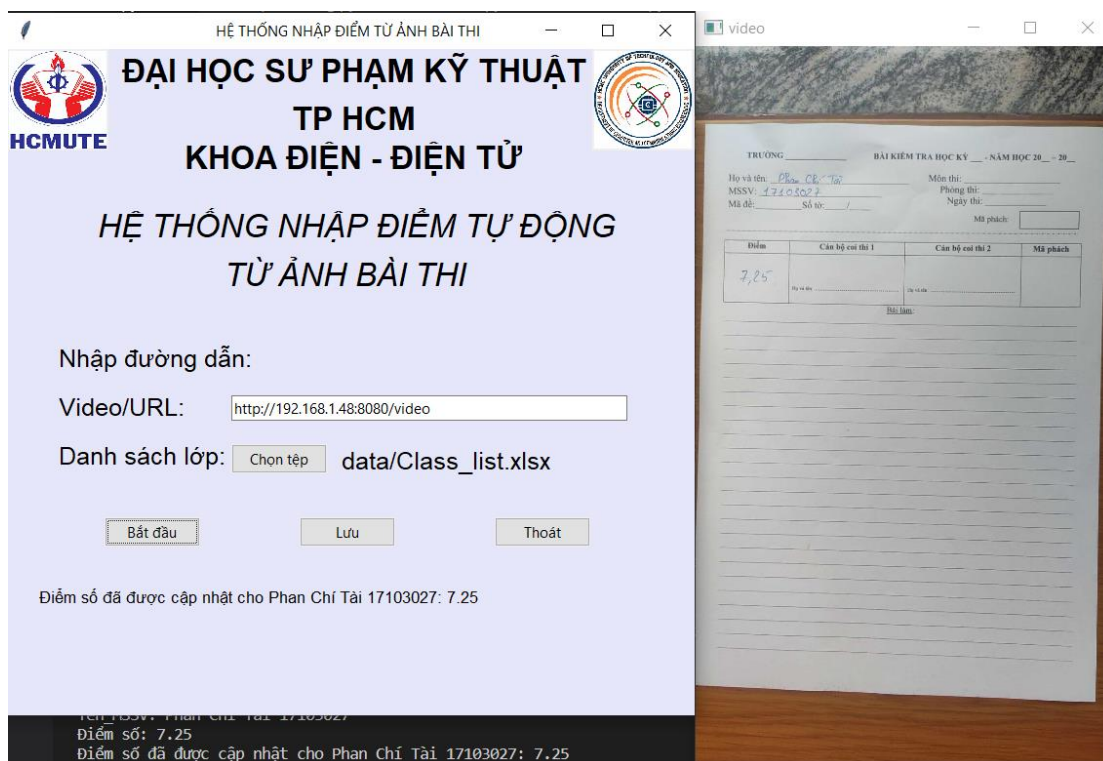
Nhập đường dẫn:

Video/URL:

Danh sách lớp: data/Class_list.xlsx

Hình 3.8: Giao diện sau khi nhập đường dẫn

Tiếp theo người dùng nhấn nút Bắt đầu để làm việc. Hình ảnh bài thi sẽ liên tục được hiển thị trên màn hình bên cạnh giao diện chính để theo dõi. Như hình dưới đang biểu diễn quá trình nhận dạng giấy thi “Phan Chí Tài 17103027”.



Hình 3.9: Giao diện hiển thị trong quá trình làm việc

Trên Terminal của hệ thống và hình ảnh Video cũng sẽ hiện thông tin kết quả nhận dạng Họ tên, MSSV và điểm số (không kể đúng sai). Sau khi kết quả nhận dạng được giống nhau 3 lần liên tục thì Terminal sẽ hiển thị dòng trạng thái “Điểm số đã được cập nhật cho Họ_ và _tên MSSV là: Điểm_số”

```
Tên_MSSV: Phan Chí Tài 17103027
Điểm số: 7.25

Tên_MSSV: Phan Chí Tài 17103027
Điểm số: 7.25

Tên_MSSV: Phan Chí Tài 17103027
Điểm số: 7.25
Điểm số đã được cập nhật cho Phan Chí Tài 17103027: 7.25
```

Hình 3.10: Giao diện hiển thị trên Terminal

Nhập đường dẫn:

Video/URL:

Danh sách lớp: data/Class_list.xlsx

Điểm số đã được cập nhật cho Phan Chí Tài 17103027: 7.25

Hình 3.11: Hình ảnh hiển thị kết quả nhận dạng gồm Họ tên, MSSV và điểm số

CHƯƠNG 4 KẾT QUẢ VÀ ĐÁNH GIÁ

4.1 KẾT QUẢ TỔNG QUAN

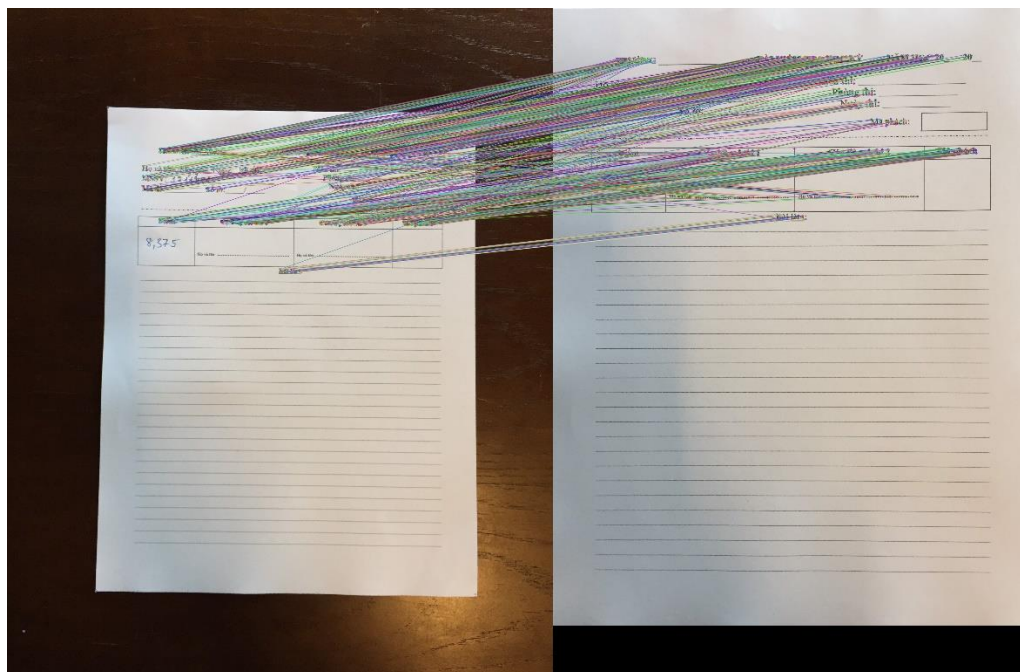
Dựa vào kiến thức, kinh nghiệm tích lũy kết hợp tình hiệu các tài liệu chuyên môn, tài liệu từ Internet cùng với sự giúp đỡ tận tình của giáo viên hướng dẫn, nhóm thực hiện đề tài “Hệ thống nhập điểm tự động từ ảnh bài thi” đã đạt được những thành công nhất định, với kết quả sau cùng của hệ thống đã đạt được những nội dung như sau:

- Trích xuất, nhận dạng được các thông tin chính trên giấy thi: Họ và tên sinh viên, MSSV, điểm số trong thời gian phù hợp với độ chính xác cao.
- Sau khi nhận dạng, có thể nhập điểm vào File Excel một cách hoàn toàn tự động.
- Hiện thị giao diện để giáo viên có thể tương tác với hệ thống.
- Hệ thống chạy ổn định trong thời gian dài, không yêu cầu nhiều về cấu hình máy tính.

4.2 TIỀN XỬ LÝ

Xử lý nhiễu và căn chỉnh hình ảnh:

Ảnh đầu vào sẽ gồm cả giấy thi và thông tin nhiễu (nền ảnh, bụi bẩn, nhiễu hạt...) đồng thời ảnh bị nghiêng, lệch so với trục diện sẽ dẫn đến việc trích xuất và nhận dạng văn bản bị sai. Nhóm đã sử dụng OpenCV để loại bỏ các thông tin thừa, nhiễu và chuyển thành ảnh xám. Sau đó sử dụng kỹ thuật Image Alignment nhờ phương pháp Feature – Based để tìm các đặc trưng của ảnh cần hiệu chỉnh sau đó sắp xếp lại hình ảnh theo một trật tự cho trước. Qua đó thu được ảnh trục diện và loại bỏ các thông tin nhiễu.



Hình 4.1: Ảnh sau khi dùng kỹ thuật Image Alignment

Với mục đích trích xuất thông tin: Họ và tên, MSSV và điểm số nên chỉ cần thông tin của nửa trên mặt giấy. Nhóm đã cắt bỏ nửa dưới của giấy thi và giữ lại nửa trên với kích thước cố định là 1683x1004 Pixels để dễ dàng cho các bước xử lý tiếp theo.

TRƯỜNG ĐH SPKT BÀI KIỂM TRA HỌC KỲ __ - NĂM HỌC 20__ - 20__

Họ và tên: Nguyễn Ngọc Khánh Môn thi: _____
MSSV: 17119084 Phòng thi: _____
Mã đề: _____ Số tờ: 1 Ngày thi: _____

Mã phách:

Điểm	Cán bộ coi thi 1	Cán bộ coi thi 2	Mã phách
8,375	Họ và tên:	Họ và tên:	

Bài làm:

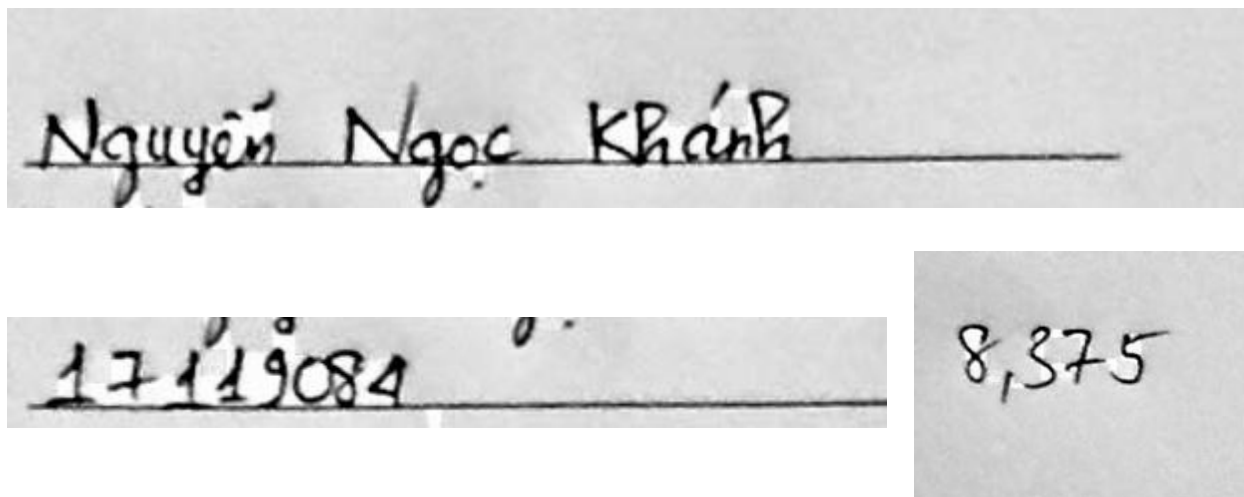
.....

.....

.....

Hình 4.2: Ảnh sau khi cắt

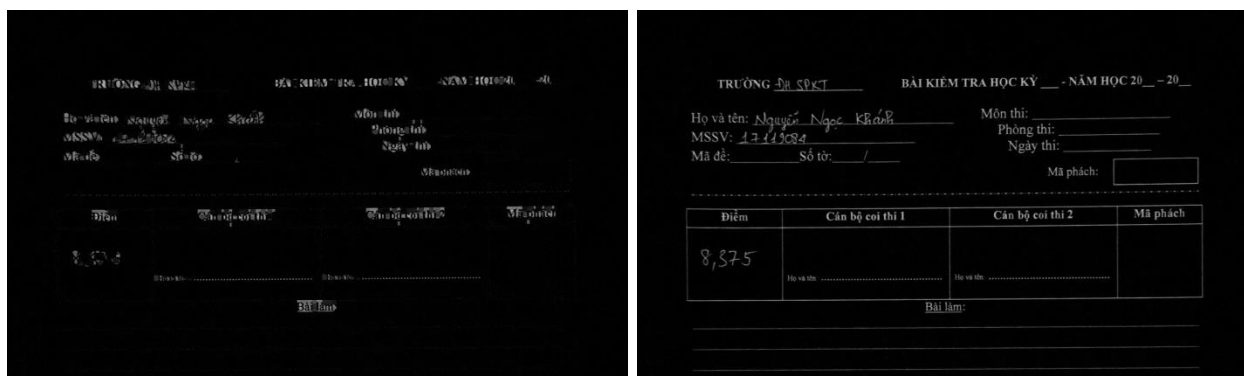
Sau đó, nhóm thực hiện xác định vị trí của Họ và tên sinh viên, MSSV và điểm số trên ảnh giấy thi rồi tiến hành cắt thành các khung thông tin cố định.



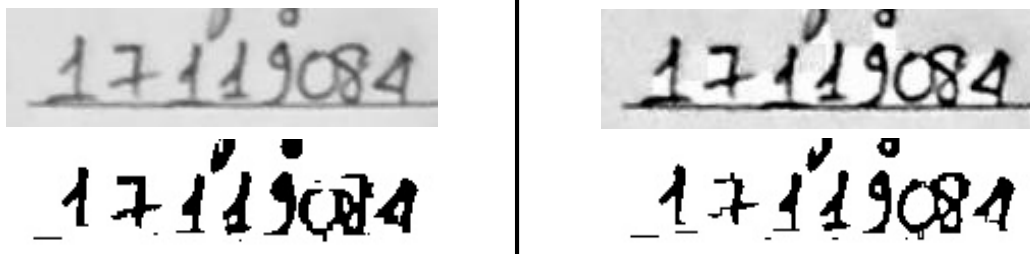
Hình 4.3: Ảnh các thông tin sau khi được cắt

Tăng độ tương phản:

Nhóm sử dụng hai phép là Top Hat (nổi bật chi tiết sáng trong nền tối) và Black Hat (nổi bật chi tiết tối trong nền sáng). Ảnh đầu ra sẽ là ảnh gốc cộng thêm ảnh qua phép Top Hat và trừ đi ảnh khi qua phép Black Hat giúp những chi tiết sáng sẽ sáng hơn nhờ phép Top Hat và chi tiết tối thì càng tối hơn nhờ phép Black Hat. Từ đó làm tăng độ tương phản cho hình ảnh. Ở bước này hỗ trợ giữ lại các đường nét của ảnh gốc bị nhòe sau khi lấy ngưỡng, giúp nâng cao khả năng nhận dạng.



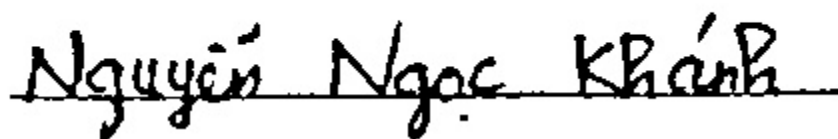
Hình 4.4: Ảnh giấy thi sau khi qua phép Top Hat (trái) và phép Black Hat (phải)



Hình 4.5: Thông tin trước (trái) và sau (phải) khi tăng độ tương phản

Ngưỡng nhị phân OTSU (OTSU Threshold):

Tuy với điều kiện ánh sáng bên ngoài, ngưỡng nhị phân OTSU không thể chống chọi tốt bằng ngưỡng nhị phân động (Adaptive Threshold), nhưng đối với ảnh cắt rất nhỏ của thông tin so với toàn bộ hình ảnh giấy thi thì sự ảnh hưởng của ánh sáng gần như không đáng kể. Ngoài ra, việc lấy ngưỡng nhị phân động cho ra ảnh có nét chữ đậm hơn nhưng kèm theo nhiều nhiễu hơn dù đã qua bước giảm nhiễu (Blur), làm mờ. Đặc biệt dòng gạch ngang ở dưới làm cho bước phân tách từ (Word Segmentation) không đạt được hiệu quả cao. Ở đây, không có phương pháp nào thực sự tốt hơn, chỉ có thể đánh đổi qua lại, điều chỉnh thông số các bước cũng như việc chuẩn bị dữ liệu để huấn luyện sẽ khác nhau ở cả hai phương pháp.



Hình 4.6: Ảnh sau khi lấy ngưỡng OTSU Threshold

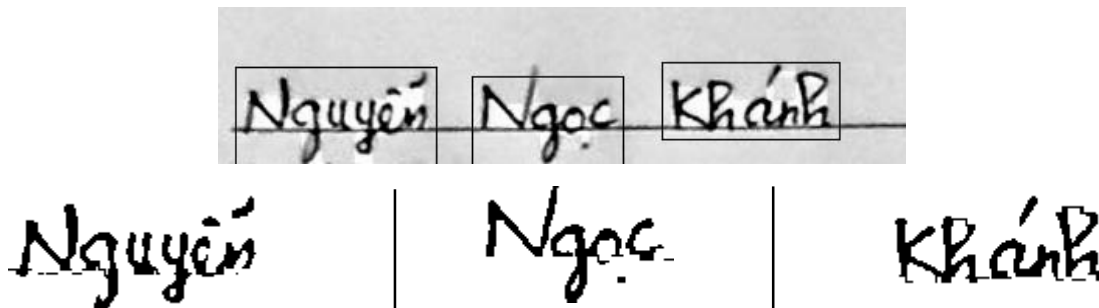
Cuối cùng là xóa hàng kẻ ngang ở dưới, làm mờ đi những đốm nhiễu, phân tách từ và điều chỉnh lại kích thước của hình ảnh để làm dữ liệu vào cho mô hình nhận dạng.



Hình 4.7: Ảnh họ tên đã được lấy ngưỡng, giảm nhiễu và xóa dòng kẻ ngang

Kỹ thuật Scale Space:

Sau khi có được hình ảnh thông tin họ và tên, nhóm sử dụng kỹ thuật Scale Space để phân tách từ hỗ trợ cho việc nhận dạng.



Hình 4.8: Kết quả sau khi thực hiện phân tách từ bằng kỹ thuật Scale Space

4.3 NHẬN DẠNG VỚI MÔ HÌNH CRNN KẾT HỢP CTC LOSS

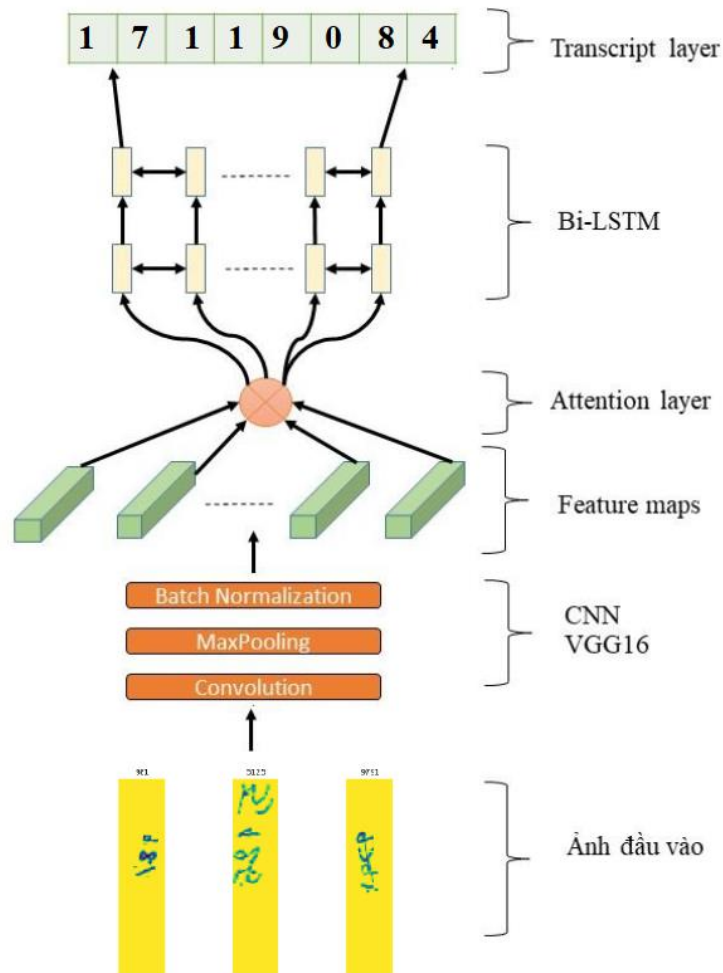
Các ảnh đầu vào được Resize và Padding để có cùng kích thước là 128x32 Pixels, sau đó các ảnh sẽ được Transpose và chuẩn hóa giá trị để đưa vào Model.

Model được sử dụng có lớp CNN từ VGG16 gồm có 7 lớp Convolution xen kẽ với 5 lớp Pooling và kết hợp với 2 lớp Batch Normalization nhằm chuẩn hóa dữ liệu cho các Model phức tạp, tránh Bias. Sử dụng với 2 lớp Bi – LSTM 256 Unit kèm Drop Out 25% để tránh việc Overfitting. Ngoài ra còn kết hợp lớp Attention, đã được chứng minh khi kết hợp với CTC Loss có thể vừa làm giảm thời gian huấn luyện và tăng khả năng nhận dạng. Kết hợp phương pháp Data Augmentation [4], giãn nở, co chữ, thêm nhiễu với các phương pháp khác như Elastics Transformation, xoay ảnh (Rotate), thay đổi tỉ lệ (Scale) làm giảm khoảng cách giữa các số, thêm đường kẻ, Random Cutout ngang và dọc...

Bảng 4-1: Thông số cài đặt cho Model họ tên và Model số

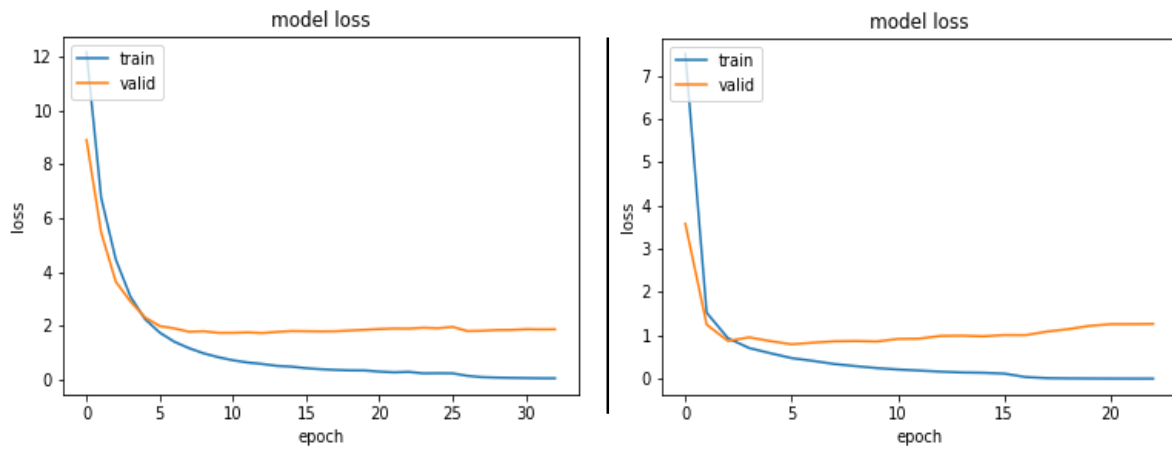
	Model họ và tên	Model số
Number of Data	110734	102255
Number of Train Data	94123	81804
Number of Valid Data	16611	20451
Split_Train_Valid	0.15	0.2

num_of_characters	148	11
max_str_len	11	10
num_of_timestamps	31	31
batch_size	128	128



Hình 4.9: Model nhận dạng được sử dụng trong đề tài

Quá trình huấn luyện Model họ tên và Model MSSV:



Hình 4.10: Quá trình huấn luyện Model họ tên (trái) và Model MSSV (phải)

4.4 KẾT QUẢ VÀ ĐÁNH GIÁ MÔ HÌNH

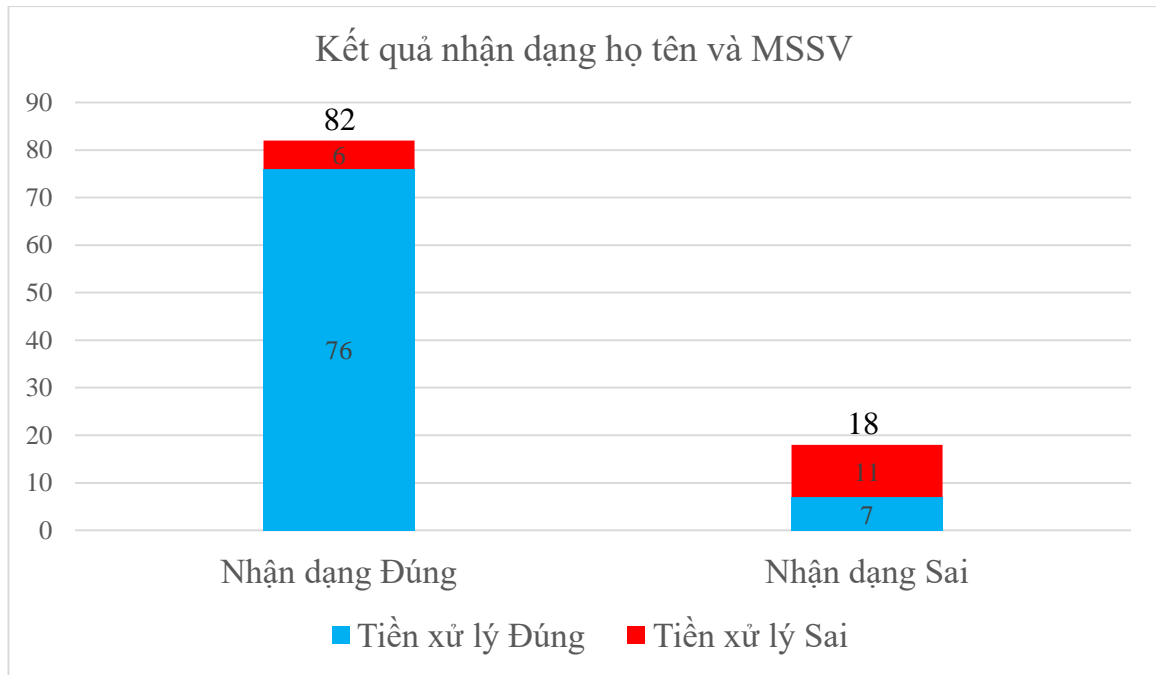
4.4.1 Kết quả và đánh giá trên hình ảnh

Kết quả về khả năng nhận diện thực tế trên 100 hình ảnh với các giấy thi khác nhau về:

- Họ và tên, MSSV, điểm số
- Kiểu chữ, cách viết
- Góc chụp ảnh
- Phong nền hình ảnh

4.4.1.1 Nhận dạng họ tên và MSSV

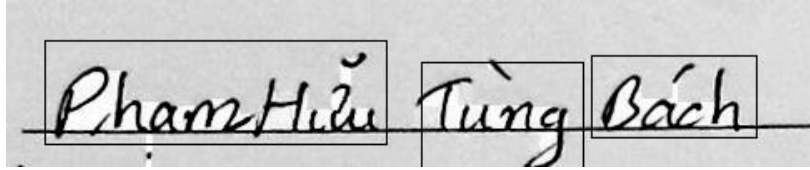
Biểu đồ 4-1: Kết quả nhận dạng họ và tên MSSV



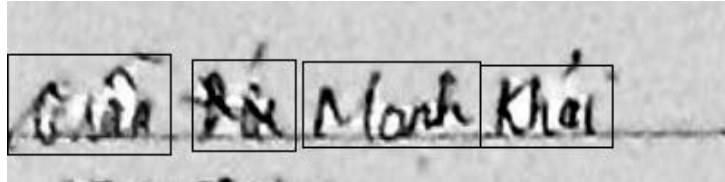
Hình ảnh tiền xử lý đúng là những hình được cắt đúng vị trí các thông tin, không có quá nhiều nhiễu, phân tách từ đúng, những hình ảnh mà đủ thông tin để con người cũng có thể nhận dạng.

Qua kết quả trên cho thấy, tiền xử lý thông tin, trích tách dữ liệu họ tên và MSSV cho độ chính xác lên đến 83%. Trong đó không có ảnh nào bị cắt sai vị trí họ tên và MSSV. Khả năng trả về đúng họ tên và MSSV là 82% và khả năng nhận dạng đúng trên tập ảnh có bước tiền xử lý đúng đạt 91,57%.

Lỗi ở đây đa phần là do bước phân tách từ sai: chữ viết nhỏ, viết gần nhau khiến việc xác định Bounding Box gặp vấn đề. Ngoài ra, ảnh có góc chụp lệch xa với góc độ trực diện khiến chữ viết khó nhận dạng, hay Model bị lẫn lộn giữ các thông tin sinh viên gần như giống nhau “Nguyễn Tuấn Việt” và “Nguyễn Tuấn Kiệt”. Cho nên độ chính xác này là chấp nhận được và có thể cải thiện thêm nếu áp dụng một số hạn chế cho người sử dụng (không tẩy xóa, chữ đẹp, rõ nét, cách chữ...)



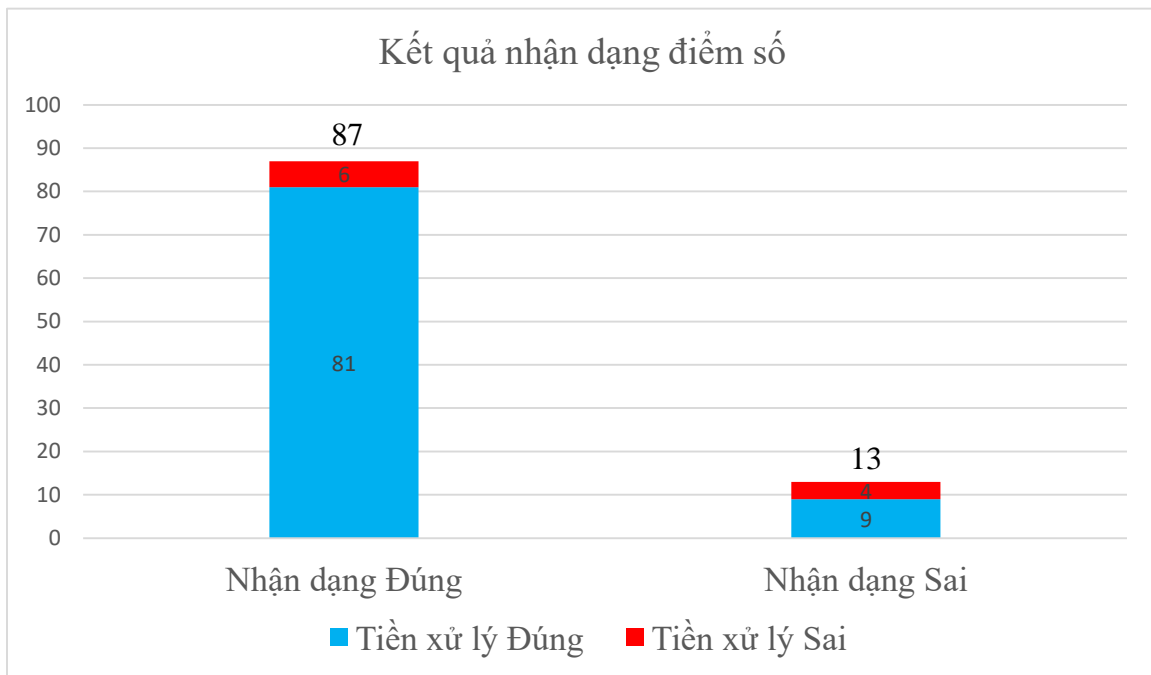
Hình 4.11: Phân tách từ lỗi vì ghi chữ sát nhau



Hình 4.12: Ảnh nhận dạng lỗi vì không hiểu được từ đã xóa

4.4.1.2 Nhận dạng điểm số

Biểu đồ 4-2: Kết quả nhận dạng điểm số

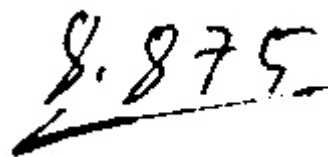


Kết quả cho thấy khả năng nhận diện điểm số khá cao, quá trình tiền xử lý đúng đạt 90%. Khả năng trả về điểm số đúng là 87%. Trong đó nhận dạng đúng điểm số trên những hình đã có tiền xử lý đúng đạt 90%.

Lỗi xảy ra khiến nhận dạng sai thường là bị nhầm lẫn giữa số 9 và 2, số 7 và 1, 7 và 2, dấu phẩy và dấu gạch dưới điểm gây nhiễu. Nhưng chủ yếu là do chữ viết xấu hoặc kiểu chữ ít khi xuất hiện trong tập Train.



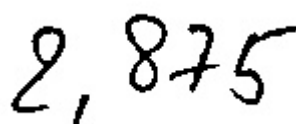
Hình 4.13: Dấu “,” làm số 0 nhầm thành số 9



Hình 4.14: Dấu gạch dưới điểm số gây nhiễu



Hình 4.16: Số 3 ghi quá tròn gây nhầm lẫn với số 8 hoặc số 0



Hình 4.15: Số 2 ghi dính liền gây nhầm lẫn với số 9

Các lỗi trên đều có thể khắc phục bằng cách thêm dữ liệu các nhiều kiểu chữ khác nhau vào quá trình huấn luyện hoặc áp đặt những hạn chế cho người sử dụng. Ví dụ số 7 phải có nét ngang, số 1 phải có nét chéo lên, số 2 không được dính liền vào thân...

4.4.2 Kết quả và đánh giá trên Video

Trên thực tế, khi thực nghiệm trên Video, các hình ảnh được đưa vào xử lý liên tục nên khả năng nhận dạng sẽ khác so với việc đưa từng hình ảnh riêng lẻ để nhận diện. Vì bây giờ thông tin sẽ được so sánh liên tục để biết thông tin đó có đáng tin cậy hay không. Đồng thời mô hình cũng đặt ra các tiêu chí nhất định để xác nhận thông tin nhận vào có đáng tin cậy hay không giúp tăng độ chính xác của kết quả.

Sau khi thử nghiệm trên 50 tờ giấy thi bất kỳ, tuy có một vài lần nhận diện sai do độ trễ của Video truyền từ Camera lên Web Server và từ Web Server vào hệ thống cũng như nhiều lúc đặt giấy thi dưới Camera nhưng với tiêu chí nhận diện đúng 3 lần giống nhau rồi mới ghi nhận thông tin của giấy thi, hệ thống đạt độ chính xác là 96% (48/50 giấy thi). Hai giấy thi còn lại do xóa bỏ chữ ở phần tên vẫn chưa xử lý được chính xác ảnh hưởng đến kết quả, tuy vậy, những lỗi xóa nhỏ vẫn có thể nhận diện được thông tin.

Hình dưới là ảnh của giấy thi khi xóa bỏ thông tin nhỏ nhưng vẫn không làm ảnh hưởng đến kết quả.

TRƯỜNG _____ BÀI KIỂM TRA HỌC KỲ ____ - NĂM HỌC 20__ - 20__

Họ và tên: Huỳnh Đức Dũng Môn thi: _____
MSSV: 17100352 Phòng thi: _____
Mã đề: _____ Số tờ: 1 Ngày thi: _____
Mã phách:

Điểm	Cán bộ coi thi 1	Cán bộ coi thi 2	Mã phách
7.25	Họ và tên:	Họ và tên:	

Bãi làm: _____

Hình 4.17: Ảnh giấy thi mẫu

```
Tên_MSSV: Huỳnh Đức Dũng 17100352
Điểm số: 7.25

Tên_MSSV: Huỳnh Đức Dũng 17100352
Điểm số: 7.25

Tên_MSSV: Huỳnh Đức Dũng 17100352
Điểm số: 7.25
Điểm số đã được cập nhật cho Huỳnh Đức Dũng 17100352: 7.25
Điểm số đã được cập nhật cho Huỳnh Đức Dũng 17100352: 7.25
```

Hình 4.18: Minh họa kết quả nhận dạng trên Terminal

Thời gian nhận dạng trung bình dưới 1 phút/giấy thi. Thời gian từ lúc người sử dụng cho giấy thi vào dưới Camera đến lúc hiển thị trên màn hình là 10 – 15 giây, độ trễ khá lớn nhưng đề tài này được sử dụng khi giáo viên chấm bài song song với quá trình nhận dạng của hệ thống và thời gian giáo viên chấm bài trung bình từ 2 – 5 phút nên thời gian xử lý của hệ thống là phù hợp.

CHƯƠNG 5 KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

5.1 KẾT LUẬN

Sau thời gian nghiên cứu và thực hiện đề tài, với sự giúp đỡ tận tình của thầy Phan Văn Ca và bạn bè, mặc dù gặp không ít khó khăn trong quá trình thực hiện, cuối cùng nhóm đã hoàn thành hệ thống và đáp ứng tốt các yêu cầu đã được đặt ra với các ưu điểm:

- Trích xuất, nhận dạng và cập nhật thông tin có độ chính xác cao và trong thời gian phù hợp
- Hệ thống hoạt động ổn định trong thời gian dài, không yêu cầu nhiều về mặt cấu hình máy tính
- Giao diện đơn giản, dễ tiếp cận với người sử dụng
- Đề tài đã tạo ra một nguồn dữ liệu cho việc huấn luyện và kiểm thử mô hình sau này, tạo điều kiện phát triển mô hình

Ngoài những ưu điểm của hệ thống, đề tài vẫn còn một số hạn chế nhất định:

- Độ chính xác khi nhận dạng những hình ảnh riêng lẻ chưa được cao
- Các giải thuật trong mô hình chưa thực sự tối ưu và hạn chế của đường truyền Internet dẫn đến việc mô hình xử lý chậm
- Hệ thống chỉ có khả năng làm việc với danh sách lớp ở dạng File Excel, chưa có khả năng kết nối, tải dữ liệu, nhập điểm lên Internet
- Giao diện có tính thẩm mỹ không được cao, chưa đẹp mắt
- Để mô hình hoạt động tốt nhất vẫn còn cần giới hạn của người dùng trong kiểu viết, cách trình bày...

5.2 HƯỚNG PHÁT TRIỂN

Từ những mặt hạn chế của đề tài đã nêu ở trên, nhóm thực hiện đưa ra những hướng mới để có thể phát triển hệ thống hoàn chỉnh và hoạt động hiệu quả hơn:

- Tối ưu hóa giải thuật nhằm tăng tốc độ xử lý và giảm thiểu chi phí tính toán của mô hình
- Phát triển các tính năng mới, kết hợp bảng điểm Online
- Thiết kế lại giao diện tăng tính thẩm mỹ, ưa mắt hơn
- Huấn luyện với những hình ảnh thực tế nhiều hơn nhằm tăng độ chuẩn xác trong các trường hợp phức tạp

TÀI LIỆU THAM KHẢO

- [1] Nguyễn Quang Hoan, “Xử lý ảnh”, Học Viện Bưu Chính Viễn Thông, 2006
- [2] Quoc Pham, “Huấn luyện mô hình CRNN cho nhận dạng chữ viết tay”, 2019, [Online]. Available: <https://pbcquoc.github.io/train-crnn>
- [3] SuNT, “Text Recognition với CRNN và CTC”, 2018, [Online]. Available: https://tiensu.github.io/blog/65_orc_crnn_model_ctc_loss_function/
- [4] Arun Gandhi, “Data Augmentation | How to use Deep Learning whhen you have Limited Data”, vol.2, 2021, [Online]. Available: <https://nanonets.com/blog/data-augmentation-how-to-use-deep-learning-when-you-have-limited-data-part-2/>
- [5] H. T. Nguyen, C. T. Nguyen, P. T. Bao, and M. Nakagawa, “A Database of unconstrained Vietnamese online handwriting and recognition experiments by recurrent neural networks,” *Pattern Recognit.*, vol. 78, pp. 291–306, 2018, doi: 10.1016/j.patcog.2018.01.013.
- [6] R. Manmatha and N.Srimal, “Scale Space Technique for Word Segmentation in Handwritten Documents”, vol.1682, 2002, doi: 10.1007/3-540-48236-9_3
- [7] B. Shi, X. Bai, C. Yao, “An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition”, 2015, arXiv: 1507.05717.
- [8] A. Choudhary, R. Rishi, and S. Ahlawat, “New character segmentation approach for off-line cursive handwritten words,” *Procedia Comput. Sci.*, vol. 17, pp. 88–95, 2013, doi: 10.1016/j.procs.2013.05.013
- [9] TheAILearner, “CTC – Problem Statement”, 2021, [Online], Available: <https://theailearner.com/2021/03/10/ctc-problem-statement/>

- [10] A. F. M. Agarap, “An Architecture Combining Convolutional Neural Network (CNN) and Support Vector Machine (SVM) for Image Classification”, pp. 5 – 8, 2017, arXiv: 1712.03541
- [11] Darmatasia and M. I. Fanany, “Handwriting recognition on form document using convolutional neural network and support vector machines (CNN-SVM),” 2017 Int. Conf. Inf. Commun. Technol. ICoIC7 2017, no. April, 2017, doi: 10.1109/ICoICT.2017.8074699
- [12] K. I. Kim, K. Jung, and J. H. Kim, “Texture-Based Approach for Text Detection in Images Using Support Vector Machines and Continuously Adaptive Mean Shift Algorithm,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 12, pp. 1631 - 1639, 2003, doi: 10.1109/TPAMI.2003.1251157
- [13] A. Nikitha, J. Geetha, and D. S. Jayalakshmi, “Handwritten Text Recognition using Deep Learning,” Proc. - 5th IEEE Int. Conf. Recent Trends Electron. Inf. Commun. Technol. RTEICT 2020, pp. 388–392, 2020, doi: 10.1109/RTEICT49044.2020.9315679
- [14] Y. Wang, W. Xiao, and S. Li, “Offline Handwritten Text Recognition Using Deep Learning: A Review,” J. Phys. Conf. Ser., vol. 1848, no. 1, 2021, doi: 10.1088/1742-6596/1848/1/012015