**BỘ GIÁO DỤC VÀ ĐÀO TẠO**

**TRƯỜNG ĐẠI HỌC QUY NHƠN**

170 An Dương Vương, TP. Quy Nhơn, Bình Định

Điện thoại: 02563 846 156 Fax: 02563 846 089 Web: www.qnu.edu.vn

*Trách nhiệm - Chuyên nghiệp - Chất lượng - Sáng tạo - Nhân văn*

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

****

**XỬ LÝ ẢNH**

**ĐỀ TÀI**

**PHÁT TRIỂN ỨNG DỤNG NHẬP ĐIỂM THI TỰ ĐỘNG TỪ ẢNH BẢNG ĐIỂM**

**GIẢNG VIÊN HƯỚNG DẪN: LÊ THỊ KIM NGA**

**THÀNH VIÊN**

**Ung Minh Hoài**

**Nguyễn Cảnh Đệ**

**Soukpasong Kedsouvannasarn**

**Neuathong Fongkhamdeng**

**Đoàn Văn Giàu**

**Bình Định, ngày 28, tháng 06, năm 2022**

**MỤC LỤC**

[**I. Tổng quan** 3](#_Toc106922929)

[**1. Đặt vấn đề** 3](#_Toc106922930)

[**2. Mục tiêu** 3](#_Toc106922931)

[**II. Cơ sở lý thuyết** 4](#_Toc106922932)

[**1. Ảnh số và điểm ảnh** 4](#_Toc106922933)

[**2. Xử lý ảnh là gì?** 4](#_Toc106922934)

[**III. Nội dung nghiên cứu** 6](#_Toc106922935)

[**IV. Phương pháp** 8](#_Toc106922936)

[**V. Cài đặt** 13](#_Toc106922937)

[**VI. Kết quả và đánh giá** 14](#_Toc106922938)

[**VII. Tài liệu tham khảo** 14](#_Toc106922939)

1. **Tổng quan**

**1. Đặt vấn đề**

Trường Đại học Quy Nhơn hiện nay có khoảng hơn 15000 sinh viên. Mỗi sinh viên một học kì trung bình 7 môn học. Kết quả mỗi môn học được đánh giá dựa trên 3 cột điểm: chuyên cần, giữa kì và cuối kì. Danh sách điểm đánh giá bởi giảng viên, có xác nhận của sinh viên và gửi về phòng Đào tạo. Với số lượng sinh viên hiện tại, việc nhập điểm từng môn với 3 cột điểm tốn rất nhiều công sức và dễ xảy ra nhầm lẫn. Có một số phương pháp đã được thực hiện như:

* Nhập thủ công trực tiếp: quá trình chấm bài thi giảng viên được cấp một phiếu điểm dùng để điền điểm thi của sinh viên, sau khi quá trình chấm bài kết thúc giảng viên ký xác nhận và gửi lên phòng Đào tạo để tiến hành việc nhập điểm. Phòng Đào tạo nhập điểm trực tiếp vào hệ thống quản lý điểm sinh viên. Việc nhập điểm đảm bảo kiểm tra nhiều lần tránh sai sót, nhầm lẫn. Công việc này tiêu tốn nhiều công sức mà hiệu quả mang lại không cao.
* Nhập điểm từ file: giảng viên nhập điểm vào bảng Excel theo mẫu của phòng Đào tạo cung cấp. Sau đó bảng điểm sẽ được nhập liệu vào hệ thống quá trình này được diễn ra một cách nhanh chóng. Tuy nhiên, cách này vẫn có những sai sót trong quá trình nhập điểm vào file, tốn nhiều thời gian nhập và độ chính xác cũng không cao.
* Nhập điểm trực tuyến: giảng viên được cấp quyền truy cập vào hệ thống quản lý điểm sinh viên. Cách này giảng viên nhập điểm trực tiếp vào hệ thống mà không thông qua phòng Đào tạo. Tuy nhiên, việc thực hiện theo cách này dẫn đến sự không đồng bộ của hệ thống trong trường hợp có nhiều giảng viên cùng truy cập.

Từ những vấn đề trên cho thấy sự cần thiết của một hệ thống mới cho phép đọc bảng điểm viết tay bằng các kỹ thuật xử lý ảnh nhận dạng ảnh bảng điểm. Bảng điểm được hệ thống xử lý và tự động nhập vào hệ thống. Tiết kiệt được thời gian, công sức, hiệu quả mang lại cao hơn những phương pháp trước đây.

**2. Mục tiêu**

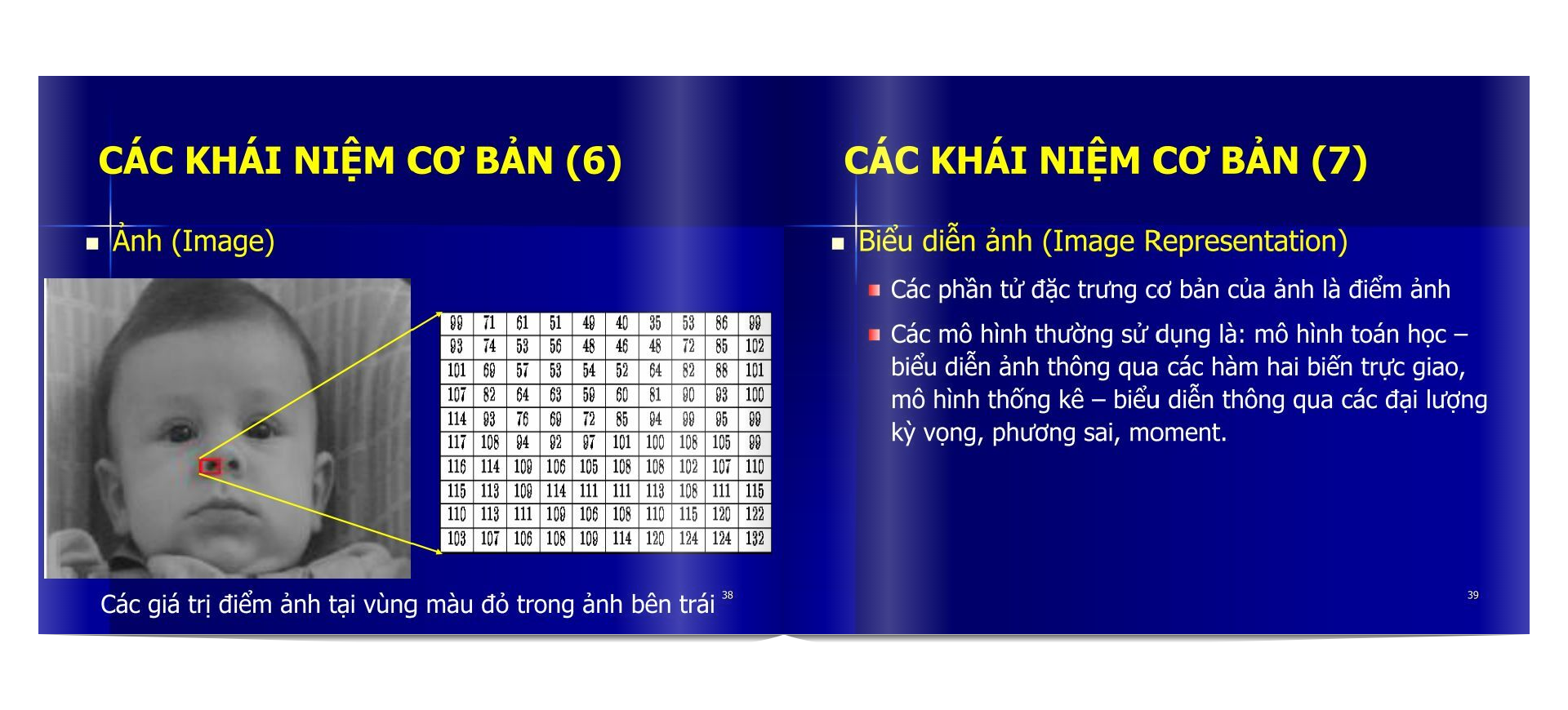
Hệ thống nhập điểm thi tự động được xây dựng để đáp ứng được mục tiêu tăng tính hiệu quả trong quá trình nhập điểm.

* **Input:** ảnh scan bảng điểm của một môn học (\*.pdf).
* **Output:** nhận dạng chính xác điểm số viết tay và xuất bảng điểm được cập nhật cột điểm viết tay ra file excel

1. **Cơ sở lý thuyết**
2. **Ảnh số và điểm ảnh**

Trước khi đi vào khái niệm về sử lý ảnh, chúng ta sẽ tìm hiểu về khái niệm của ảnh số và điểm ảnh.

* Ảnh số: một hình ảnh có thể được định nghĩa là hàm hai chiều, f (x, y), trong đó x và y là tọa độ không gian (mặt phẳng) và biên độ của f tại bất kỳ cặp tọa độ (x, y) nào được gọi là cường độ hoặc mức độ màu xám của hình ảnh tại điểm đó. Khi x, y và các giá trị cường độ của f đều là các đại lượng hữu hạn, rời rạc, chúng ta gọi hình ảnh là hình ảnh kỹ thuật số. Hay có thể hiếu 1 cách đơn giản rằng "Ảnh số là số hóa làm cho một hình ảnh kỹ thuật số trở thành một xấp xỉ của một cảnh thực”.
* Điểm ảnh: hình ảnh kỹ thuật số chứa một số lượng hữu hạn các hàng và cột của các phần tử. Mỗi phần tử được gọi là pixel.
* Với ảnh trắng đen ta có ví dụ sau:

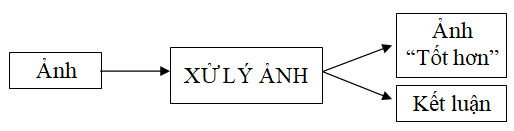
****

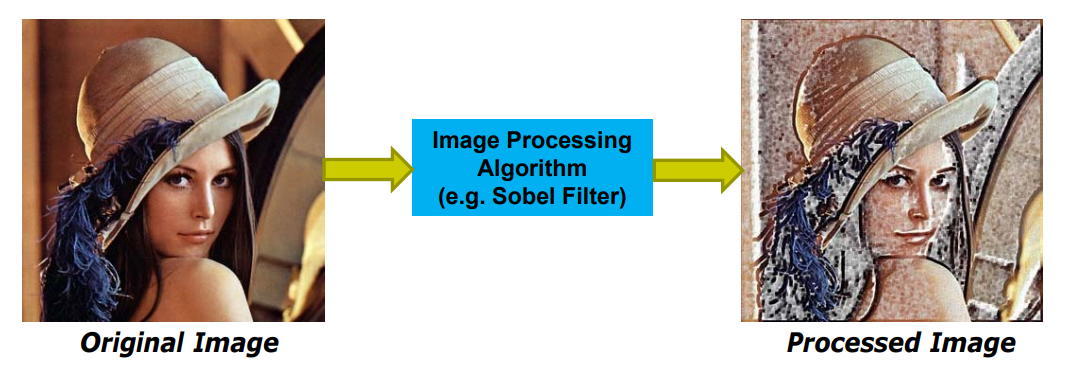
**Hình 1. Minh họa các điểm ảnh**

* Các giá trị điểm ảnh tại vùng màu đỏ trong ảnh bên trái.

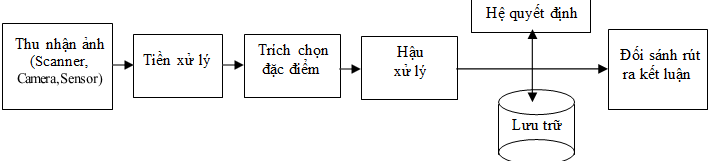
1. **Xử lý ảnh là gì?**
   1. **Khái niệm**

Xử lý ảnh là một phần trong lĩnh vực xử lý tín hiệu số. Trong đó sử dụng các thuật toán thay đổi hình ảnh đầu vào để tạo hình ảnh mới nhằm mục đích tăng cường chất lượng thông tin hình ảnh đối với quá trình tri giác của con người và biểu diễn trên máy tính. Nói một cách khác, xử lý ảnh là việc hạ xuống những thứ mình không quan tâm và làm nổi lên những thứ mình quan tâm trong ảnh số.





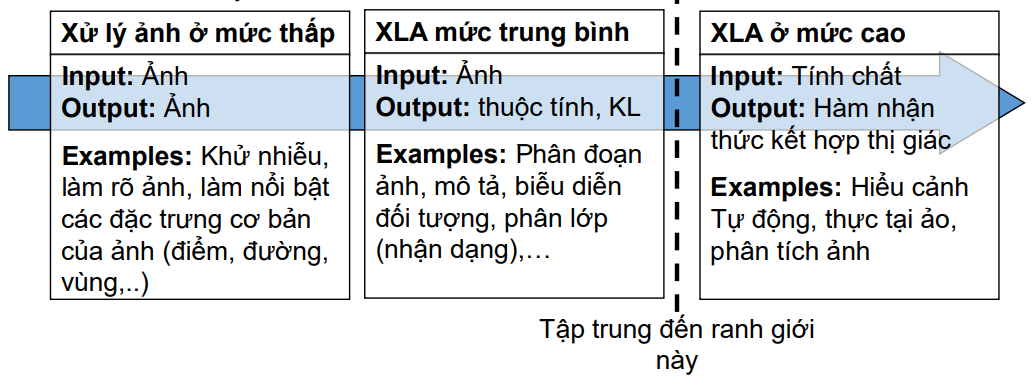
**Hình 2.** Ảnh sau khi được xử lý từ ảnh gốc



**Hình 3.** Các bước trong một hệ thống xử lý ảnh

1. **Nhiệm vụ chính của xử lý ảnh**

* Xử lý ảnh tập trung vào 2 nhiệm vụ chính:
* Nâng cao chất lượng ảnh
* Xử lý dữ liệu ảnh cho: lưu trữ, truyền thông và biểu diễn trong các hệ thống thị giác máy.
* Xử lý ảnh chia làm 3 mức:



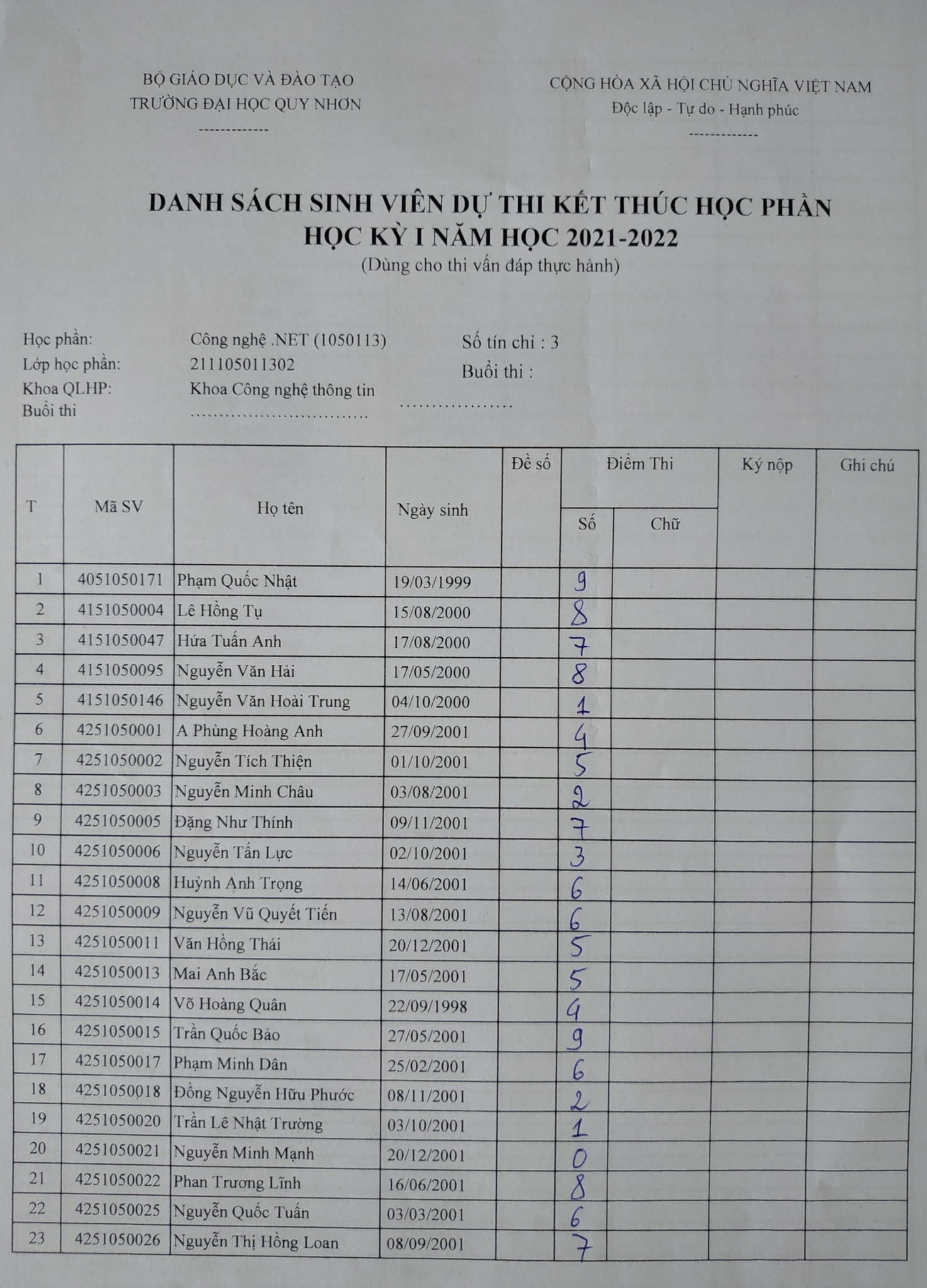
**Hình 4.** Các mức độ trong xử lý ảnh

1. **Các giai đoạn chính của xử lý ảnh**

* Thu nhận ảnh: ảnh được thu nhận từ thế giới thực qua máy chụp hình, từ tranh ảnh qua máy quét hoặc từ vệ tinh thông qua bộ cảm biến số hoặc tương tự.
* Số hóa ảnh: Số hóa các ảnh thu nhận được để lưu trữ vào máy tính.
* Phân tích ảnh: Gồm nhiều theo tác: Tăng cường ảnh, khử nhiễu, chỉnh méo, trich chọn đặc trưng,…
* Đối sánh nhận dạng: Phân lớp phục vụ cho các mục đích khác nhau.

1. **Nội dung nghiên cứu**
2. **Giới thiệu bài toán**

Sau khi sinh viên thi, bài thi được các giảng viên nhận về chấm và lên điểm bằng tay vào bảng điểm.



**Hình 5.** Minh họa ảnh bảng điểm đầu vào

Nghiên cứu cần xây dựng chương trình nhận dạng điểm viết tay trên cột số thuộc cột điểm thi và xuất kết quả sang file bất kỳ (\*.csv, \*.xls, …) theo cấu trúc mỗi dòng chứa <điểm>

Với bài toán trên, nghiên cứu đề ra hai mục tiêu chính:

* Thứ nhất: là định vị chính xác các ô điểm số viết tay trên cột số.
* Thứ hai: là nhận dạng chính xác điểm số viết tay trong các ô đã được định vị với độ chính xác 70% trở lên trên mỗi bảng điểm.

1. **Đối tượng và phạm vi nghiên cứu**

Nghiên cứu được thực hiện trên đối tượng là các ảnh bảng điểm được thu nhận qua máy ảnh, ảnh đầu vào có kích thước 2560, 1715.

Phạm vi các điểm nghiên cứu nhận dạng gồm có điểm 1, 2, 3, 4, 5, 6, 7, 8, 9 là các số tròn, không chứa phần thập phân. Ngoài ra, trường hợp sinh viên vắng thi thì ô điểm sẽ nhận giá trị mặc định “1”.

1. **Dữ liệu**

Nghiên cứu thu thập một số ảnh bảng điểm từ các thầy cô trong trường Đại học Quy Nhơn. Dự án sử dụng mô hình Tensorflow dùng để train và test dữ liệu.

1. **Phương pháp**
2. **Định vị và lọc vùng bảng**

Để nhận dạng điểm viết tay trên cột số, nghiên cứu cần phải định vị các ô ở cột số. Để định vị các ô, nghiên cứu cần tiếp cận vùng bảng. Vùng bảng là vùng rộng nhất trên ảnh. Trước khi định vị vùng bảng, ảnh bảng điểm được tiền xử lý qua các bước:

Nhị phân ảnh: từ ảnh đầu vào là ảnh xám ta đưa về ảnh nhị phân bằng kỹ thuật phân ngưỡng. Nếu giá trị tại điểm ảnh (i,j) > giá trị ngưỡng thì (i,j) có màu trắng, ngược lại có màu đen. Với giá trị ngưỡng là giá trị trung bình của vùng 3x3 với tâm là điểm (i,j) đang xét.

Âm bản: từ ảnh nhị phân dùng phép toán âm bản nhằm làm nổi các đối tượng như các đường thẳng, văn bản, điểm số,…

1. **Định vị ô**

Ảnh sau khi tiền xử lý bước đầu làm rõ đối tượng. Tiền hành phát hiện các đường thẳng dọc và ngang của ảnh sử dụng nhân có ma trận (3x3). Tiếp tục xói mòn ảnh để loại bỏ những pixel nhiễu xung quanh đối tượng giúp cho phần viền (cạnh) của đối tượng trở nên mịn hơn và giãn nỡ để các nét được to và rõ hơn. Sau khi nhận diện được đường thẳng ngang và dọc từ 2 ảnh khác nhau thì tiến hành ghép 2 ảnh đó lại. Sau đó dùng hàm findContours() để phát hiện các hình chữ nhật lưu vào một mảng và tiến hành cắt ô. Vì các ô được cắt từ dưới lên nên chúng ta cần đảo ngược mảng để lấy các ô điểm theo thứ tự từ trên xuống.

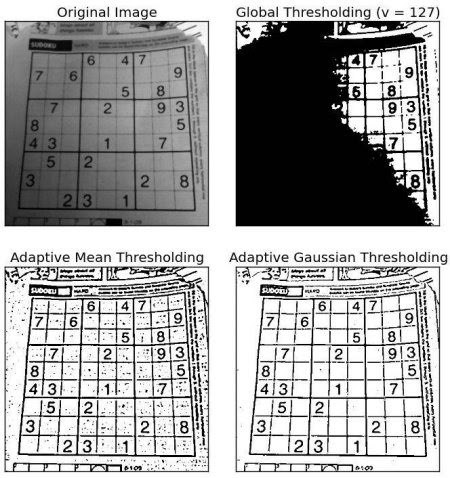
Cuối cùng, tiến hành cắt các ô chứa điểm viết tay bằng cách duyệt qua mảng vừa được đảo ngược ở trên dùng hàm boudingRect() để xác định chiều cao và chiều rộng cho mỗi đường viền của ô, chỉ chọn các ô nằm trong phạm vi chiều cao và chiều rộng quy định.

1. **Nhận dạng điểm số viết tay**

Để nhận dạng điểm số viết tay ta cần thực hiện qua 3 giai đoạn là tiền xử lý các ô điểm, trích xuất đặc trưng với GIST và phân lớp với máy học CNN.

1. **Tiền xử lý các ô điểm**

Các ô sau khi được cắt ra từ bảng điểm gốc, ta tiến hành xử lý ảnh để chuẩn bị cho việc nhận dạng chữ số. Tiến hành tiền xử lý xóa các nhiễu áp dụng phương pháp lọc Gauss (Gaussian Filter) để làm mịn ảnh. Trong đó sử dụng mà trận lọc (kernel) 7x7 để đưa ra được hiệu quả phù hợp cho bài toán này. Sau đó nhị phân hóa ảnh với ngưỡng thích nghi (Adaptive Threshold). Phương pháp phân ngưỡng đơn giản không phù hợp cho nhiều trường hợp, như là ánh sáng không đồng đều trên ảnh. Trong trường hợp này chúng ta dùng hàm adaptiveThreshold(). Phương thức này tính giá trị trung bình của các n điểm xung quanh pixel đó rồi trừ cho C chứ không lấy ngưỡng cố định (n thường là số lẻ, còn C là số nguyên bất kỳ). Sau khi xóa các nhiễu tiến hành chuẩn hóa kích thước các ảnh về kích thước 28x28.



**Hình 6.** Hình ảnh minh họa cách phân ngưỡng khác nhau



**Hình 7.** Ảnh sau khi làm mịn và phân ngưỡng



**Hình 8.** Ảnh sau khi cắt

1. **Trích chọn đặc trưng**

Sau khi xử lý các bước trên, ảnh kết quả vẫn chưa phù hợp. Vì các con số trong ảnh chưa thực sự nằm ở chính giữa của ảnh. Vì vậy, tiến hành tìm kiếm điểm chính giữa và đưa các con số trong ảnh vào đúng vị trí chính giữa đó.

Trong xử lý ảnh, moment của ảnh được dùng để nêu bật đặc trưng hình dạng của một ảnh. Những moment này ghi lại những thuộc tính của hình ảnh, bao gồm diện tích của đối tượng (the area), trọng tâm (centroid), hướng (orientation) và những thuộc tính liên quan khác. Chúng ta đã tính được điểm chính giữa của ảnh, là tọa độ (cX, cY), sau đó áp dụng hàm warpAffine() dịch chuyển các con số trong ảnh sang điểm trung tâm đó.

1. **Phân lớp**

MNIST là một dataset gồm 70.000 ảnh số viết tay có kích thước 28x28 và có thể sử dụng thông qua Keras. Mỗi ảnh trong dataset là một ảnh xám hiển thị cho một chữ số viết tay từ 0-9. Trong đó, được chia ra thành training set 60.000 ảnh và test set 10.000 ảnh. Hàm load\_data() được sử dụng để tải bộ dữ liệu từ keras.



**Hình 9.** Mnist Dataset

Mô hình mạng neural tích chập hoặc CNN (Convolutional Neural Network) là 1 trong những mô hình để nhận dạng và phân loại hình ảnh. Trong đó, CNN tập hợp các thuật toán để có mô hình dữ liệu trừu tượng hóa ở mức cao bằng cách sử dụng nhiều lớp xử lý cấu trúc phức tạp và được sử dụng rộng rãi trong các bài toán xác định và nhận dạng đối tượng.

* Các lớp cơ bản của Convolutional Neural Network
* Convolutional Layer

Convolutional được hiểu là tích chập. Khi gặp hình ảnh mới, Convolutional sẽ thử chúng với tất cả các vị trí khác nhau và tạo thành một bộ lọc gọi là Filter.

* Pooling Layer

Khi đầu vào quá lớn, các lớp Pooling Layer sẽ được dịch chuyển vào giữa những lớp Convolutional Layer nhằm giảm các Parameter.

Pooling Layer được biết đến với hai loại phổ biến là: Max Pooling và Average Pooling.

Tại Pooling Layer, khi bạn sử dụng lớp Max Pooling thì số lượng Parameter có thể sẽ giảm đi. Vì vậy, Convolutional Neural Network sẽ xuất hiện nhiều lớp Filter Map, mỗi Filter Map đó sẽ cho ra một Max Pooling khác nhau.

* Relu Layer

Đây chính là một hàm kích hoạt trong Neural Network. Chúng ta có thể biết đến hàm kích hoạt này với một tên gọi khác là Activation Function. Nhiệm vụ chính của hàm kích hoạt là mô phỏng lại các Neuron có tỷ lệ truyền xung qua Axon. Trong đó, hàm kích hoạt sẽ bao gồm các hàm cơ bản như: Sigmoid, Tanh, Relu, Leaky Relu, Maxout.

Hiện nay, hàm Relu đang được sử dụng khá phổ biến và thông dụng. Đặc biệt, Relu sở hữu những ưu điểm nổi bật như: hỗ trợ tính toán nhanh nên rất được ưa chuộng sử dụng trong việc huấn luyện các mạng Neuron.

Khi sử dụng Relu, bạn cần lưu ý đến việc tùy chỉnh các Learning Rate và theo dõi Dead Unit. Lớp Relu Layer này được sử dụng sau khi Filter Map được tính toán ra và áp dụng hàm Relu lên tất cả các giá trị trên Filter Map.

* Fully Connected Layer

Fully Connected Layer thường sử dụng để đưa ra các kết quả.

Ví dụ: Sau khi các lớp Convolutional Layer và Pooling Layer nhận được các ảnh đã truyền qua chúng, bạn sẽ thu được kết quả là Model đã đọc được khá nhiều thông tin về ảnh. Do đó, để có thể liên kết các đặc điểm này lại và cho ra Output, bạn cần dùng đến Fully Connected Layer.

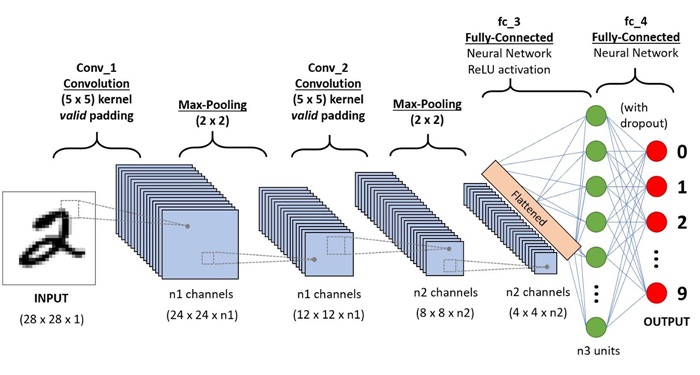
Bên cạnh đó, khi có được các dữ liệu về hình ảnh, Fully Connected Layer sẽ chuyển đổi chúng thành những mục có phân chia chất lượng. Tương tự như kiểu chia chúng thành các phiếu bầu và đánh giá để chọn ra hình ảnh đạt chất lượng tốt nhất. Dù vậy, quá trình này không được coi là quá trình dân chủ nên rất ít sử dụng.

* **Cấu trúc của mạng CNN**

Mạng Convolutional Neural Network là tập hợp nhiều lớp Convolutional chồng lên nhau, sử dụng các hàm Nonlinear Activation và tanh để kích hoạt các trọng số trong các node. Ở mỗi lớp CNN, sau khi được các hàm kích hoạt sẽ tạo ra các thông tin trừu tượng hơn cho những lớp tiếp theo. Mỗi Layer kế tiếp sẽ là kết quả Convolution từ Layer trước đó nên chúng ta có được các kết nối cục bộ.

Thông qua quá trình huấn luyện mạng, các lớp Layer CNN tự động học các giá trị được thể hiện qua các lớp Filter.

Ví dụ: Trong tác vụ phân lớp ảnh, CNNs sẽ cố gắng tìm kiếm những thông số tối ưu cho các Filter tương ứng theo một thứ tự: Raw Pixel => Edges => Shapes => Facial => High – level Features. Layer cuối cùng được dùng để phân lớp ảnh.



**Hình 9.** Phân lớp ảnh

1. **Cài đặt**

Sau khi nghiên cứu các phương pháp ở trên, nhóm đã tiến hành cài đặt triển khai demo trên Google Colab. Thời gian thực hiện chương trình demo chạy trên GPU cho kết quả khả quan. Thời gian thực hiện chương trình nhận dạng phụ thuộc khá nhiều vào chất lượng bảng điểm đầu vào, số lượng các ô điểm trên mỗi bảng điểm,…

Các công cụ và thư viện được nhóm sử dụng:

* Google colab
* Thư viện Opencv-python
* Bộ ảnh Mnist của Tensorflow để xây dựng model nhận diện ảnh số viết tay.

1. **Kết quả và đánh giá**
2. **Ưu điểm**

* Áp dụng được một số kiến thức đã học như: đọc ảnh, phân ngưỡng, âm bản, làm mịn,…
* Nhóm đã thực hiện được phương pháp định vị ô ứng dụng trong nhiều loại bảng điểm khác nhau và đạt độ chính xác 99,9 %.
* Train model đạt độ chính xác 99,9 %
* Nhận dạng chữ viết tay có kết quả khả quan với bộ test 41 dòng thì số nhận dạng không chính xác là 9 đạt độ chính xác 78%.

1. **Hạn chế**

Vì phục vụ cho quá trình nghiên cứu và học tập nên dự án có một số hạn chế như:

* Chưa có nguồn dữ liệu ảnh bảng điểm viết tay đủ lớn.
* Độ chính xác còn khá chủ quan vì dữ liệu test không đủ lớn để đánh giá được dự án một cách đúng nhất
* Nguồn ảnh còn hạn chế nên chưa triển khai xây dựng mô hình, đang sử dụng nguồn ảnh từ Tensorflow (Mnist) để xây dựng mô hình nhận diện chữ số viết tay.
* Một số nguyên nhân khách quan ảnh hưởng tới độ chính xác của mô hình: điều kiện thu nhận ảnh, dữ liệu bị nhiễu nhiều do chữ số viết tay chưa đồng bộ,…

1. **Hướng phát triển**

* Khắc phục được những hạn chế còn tồn tại ở trên.
* Thu thập đủ nguồn dữ liệu ảnh bảng điểm để xây dựng mô hình nhận diện chữ số viết tay.
* Mở rộng nhận diện thêm các trường như mã số sinh viên, mã môn học,.. để đúng với mục tiêu đề ra của dự án.

1. **Tài liệu tham khảo**
2. <https://stackoverflow.com/>
3. <https://tailieumienphi.vn/doc/nhan-dang-diem-viet-tay-tren-bang-diem-voi-bien-doi-hough-va-dac-trung-gist-xay9tq.html>
4. <https://docs.opencv.org/4.x/d6/d00/tutorial_py_root.html>
5. <https://www.analyticsvidhya.com/blog/2021/07/classification-of-handwritten-digits-using-cnn/>
6. <https://www.analyticsvidhya.com/blog/2020/02/learn-image-classification-cnn-convolutional-neural-networks-3-datasets/>