

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN



BÀI TẬP LỚP NHÀ

Nhóm CKT

Ứng dụng xử lý ảnh số và video số 20_23

Giảng viên – TS.Lý Quốc Ngọc

Thành phố Hồ Chí Minh – 2023

MỤC LỤC

THÔNG TIN CÁC THÀNH VIÊN TRONG NHÓM	4
I. Tuần 1.....	5
II. Tuần 2.....	7
III. Tuần 3.....	9
IV. Tuần 4.....	10
V. SEMINAR.....	13
1. Nhóm DC - Truy vấn ảnh y phục.....	13
a) Động lực nghiên cứu	13
b) Phương pháp	13
2. Nhóm 2T - Định vị và tái tạo môi trường xung quanh	14
a) Động lực nghiên cứu	14
b) Phương pháp	15
3. Nhóm VIP2 - Xây dựng hệ thống điểm danh lớp học và lưu giữ đường cong cảm xúc khuôn mặt của sinh viên	16
a) Động lực nghiên cứu	16
b) Phương pháp	16
4. Nhóm Cực Cháy - Xây dựng hệ chuẩn đoán khả diễn bệnh ung thư xương (XAI).....	17
a) Động lực nghiên cứu	17
b) Phương pháp	17
5. Nhóm Chiến thần - Tạo sinh ảnh dựa vào ảnh mẫu quảng cáo và ảnh đích	18
a) Động lực nghiên cứu	18
b) Phương pháp	18

6. Nhóm TT - Xây dựng hệ thống truy vấn ảnh dựa vào hỏi đáp	19
a) Động lực khoa học.....	19
b) Động lực ứng dụng.....	19

THÔNG TIN CÁC THÀNH VIÊN TRONG NHÓM

MSSV	Họ Tên	Email	Ghi chú
20120201	Phạm Gia Thông	20120201@student.hcmus.edu.vn	
20120259	Nguyễn Văn Chung	20120259@student.hcmus.edu.vn	Nhóm trưởng

I. Tuần 1

1/ Ứng dụng 1.1, 1.2 sẽ thúc đẩy được gì vào tiếp thị số, thương mại điện tử?

Styles	
Tên đồ án	Nội dung
<u>Tạo sinh ảnh dựa vào mẫu</u> <u>quảng cáo và ảnh đích.</u> <u>Ứng dụng: hỗ trợ ghép mẫu</u> <u>quảng cáo vào ảnh người mẫu </u> <u>ảnh nền.</u>	Applications in digital marketing and e- commerce 1.1
<u>Xây dựng hệ thống Truy vấn y</u> <u>phục, đối tượng dựa vào ảnh</u> <u>truy vấn và câu chú thích.</u> <u>Ứng dụng: hỗ trợ truy vấn ảnh</u> <u>thông minh, kết hợp thông tin</u> <u>từ ảnh và câu chú thích.</u>	Applications in digital marketing and e- commerce 1.2

- Những ứng dụng này kéo người dùng lại gần với sản phẩm hơn, liên kết thông tin, giá thành, hình ảnh của sản phẩm đến khách hàng một cách chi tiết và nhanh nhất.

- Ngoài ra nó còn đem lại sự trải nghiệm tiện nghi cho người dùng, có thể cắt ghép ảnh nhanh chóng mà không cần phải tốn quá nhiều thời gian và công sức, có thể lựa chọn sản phẩm phù hợp với người tiêu dùng bằng cách ướm thử sản phẩm trong một môi trường ảo hoặc tìm kiếm sản phẩm đó một cách nhanh chóng chỉ thông qua mô tả, hình ảnh mà không cần phải có được thông tin cụ thể của sản phẩm đó (tên, nhãn hàng,...)

2/ Tình huống nào thì cần dùng hỏi đáp, cần dùng truy vấn hay dùng kết hợp cả truy vấn và hỏi đáp?

VQA	VIR	V.IR+VQA
- input: * số người ra vào dãy nhà * tập ảnh giám sát	- input: ảnh đối tượng tập ảnh giám sát	- input: ảnh đối tượng * đối tượng rời khỏi dãy nhà vào buổi tối tập ảnh giám sát
- out put: số người rút ra được từ tập ảnh	- Out put: tập ảnh giám sát có đối tượng	- Out put: tập ảnh rút ra được có chứa đối tượng rời khỏi dãy nhà vào buổi tối

II. Tuần 2

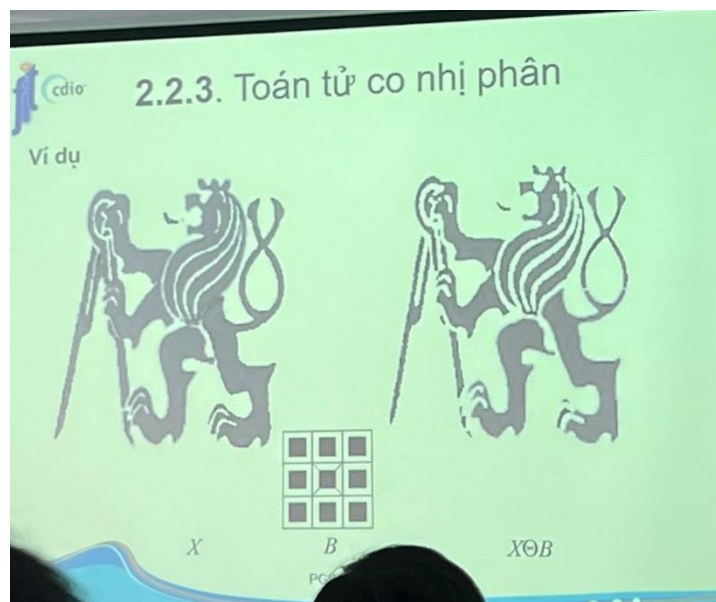
1/ Làm sao để từ ma trận X , B có thể nhìn ra vector gia số $X \oplus B$ cho nhanh?

- Ta lấy điểm tham chiếu ở bên B sau đó đặt lên trên X , tịnh tiến dần cho hết mọi điểm trong X , hình thái ảnh của B dần hiện ra và hằn lên theo các điểm tham chiếu trong X

2/ Viết toán tử dilation, toán tử erosion

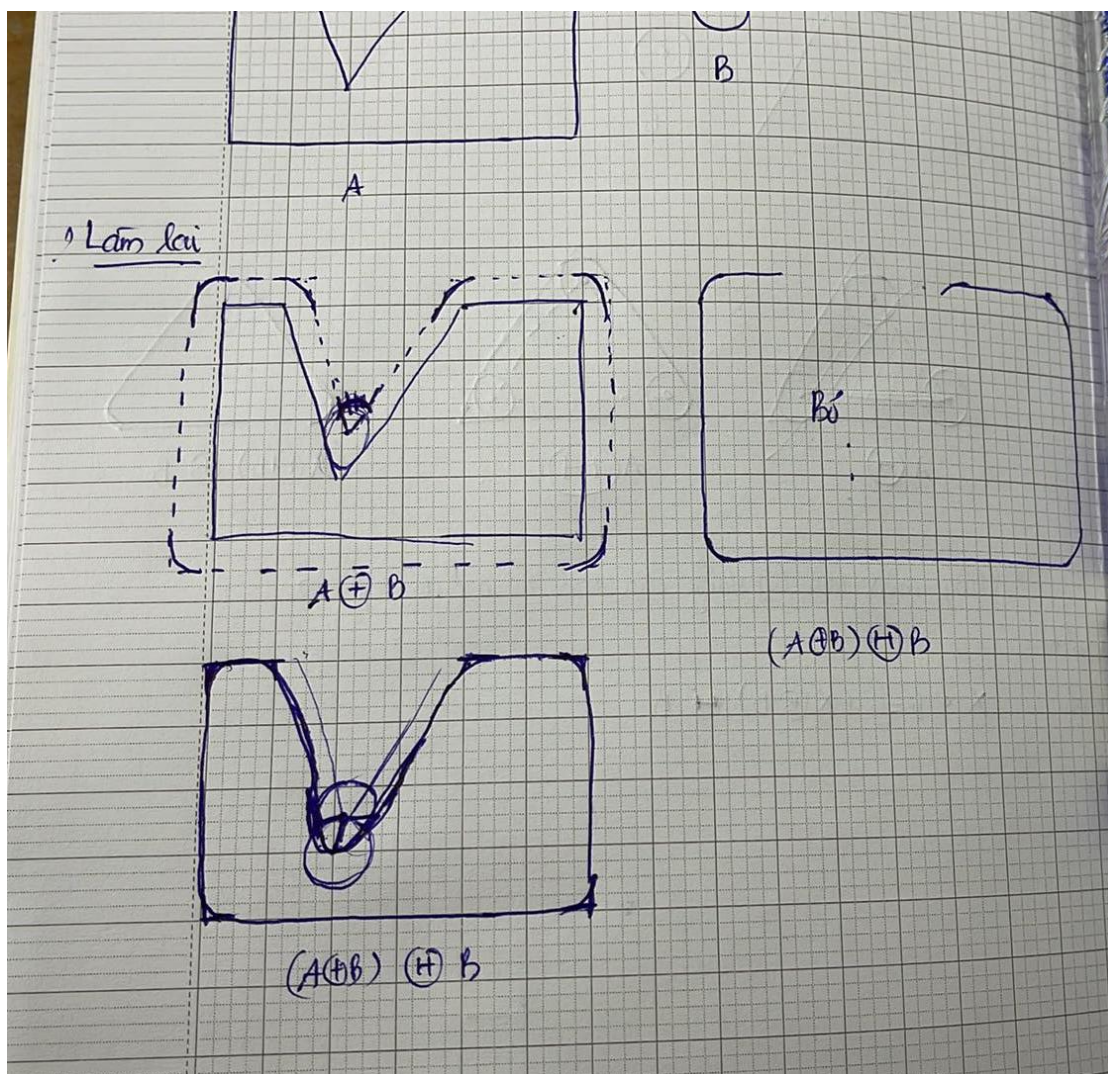
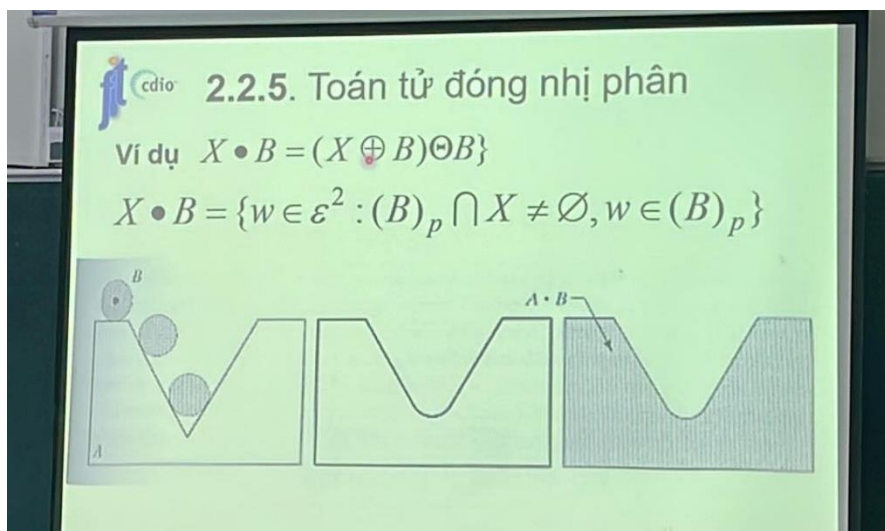
- Tập dữ liệu y tế. Tập dữ liệu sẽ tồn tại các mẫu âm nhiều hơn các mẫu dữ liệu dương (mẫu không mắc bệnh nhiều hơn mẫu mắc bệnh)
- Tập dữ liệu về gian lận. Số người không gian lận sẽ lớn hơn nhiều số người gian lận.
- Tập dữ liệu về phát hiện lỗi trong quá trình sản xuất.
- Tập dữ liệu về khuôn mặt người bị biến dị.

3/ Tại sao nó ra hình bên phải?



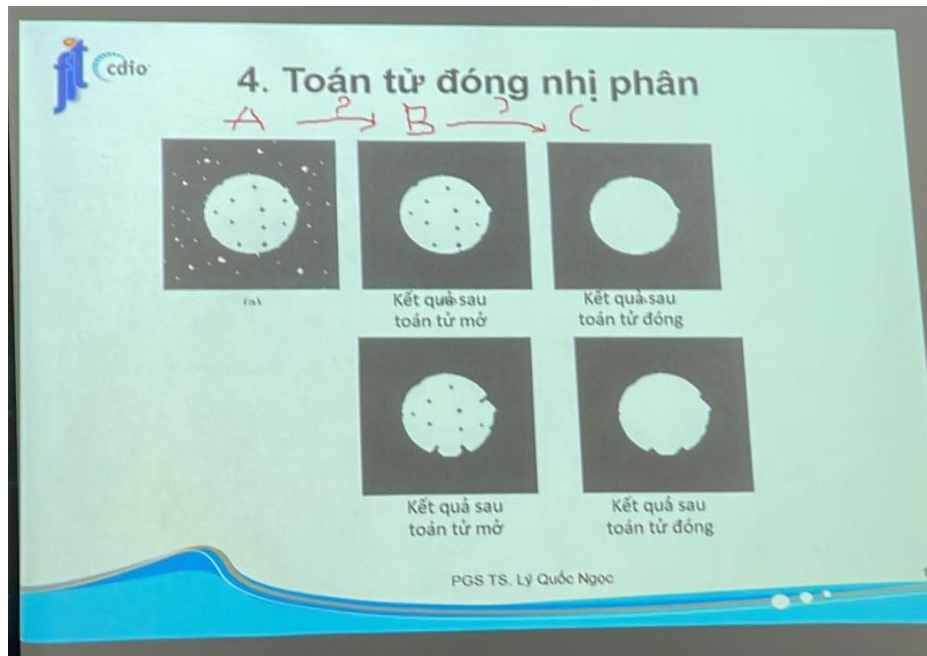
- Vì đây là toán tử co nhị phân (erosion) nó sẽ hằn điểm tham chiếu lên đường kẻ của hình, điểm nào vượt ra ngoài đường kẻ nó sẽ không lấy, chỉ điểm thuộc trong phạm vi của đối tượng

4/ Chứng minh hình ở dưới



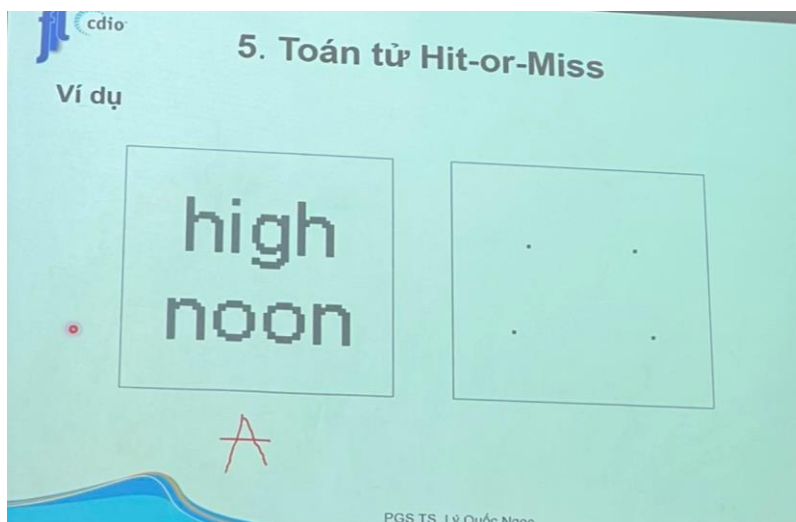
III. Tuần 3

1/ Làm sao để từ A -> B -> C?



Từ A qua B chúng ta sẽ dùng phép co (erosion) để loại bỏ những nhiễu trắng bên ngoài đối tượng, từ B qua C sẽ dùng phép giãn (dilation) để lọc bỏ những nhiễu đen bên trong, với ma trận kết cấu (kernel) có n là cạnh của hình vuông lớn nhất, ma trận quyết định kết quả sau cùng của các phép biến đổi

2/ Làm sao để từ A -> C?



Dùng toán tử hình thái học Hit or Miss

Lấy ma trận kết cấu giống chữ n: B1

Lấy X Erosion B1 sẽ giúp lấy được vị trí n, nhưng với h thì ma trận kết cấu n vẫn nằm bên trong nên chữ h bị detect thành chữ n.

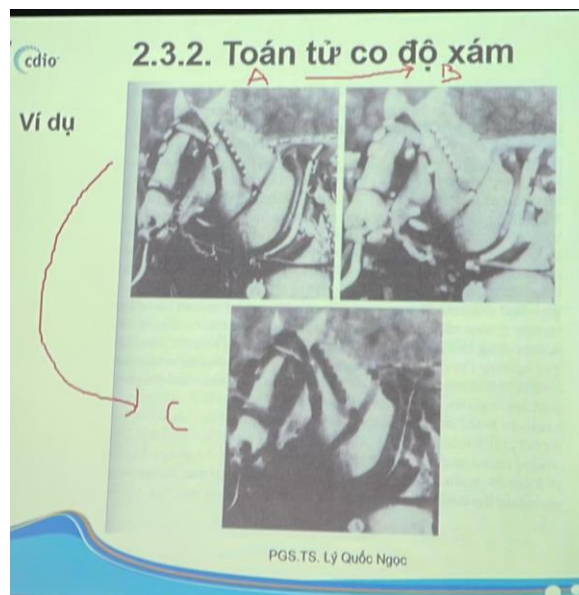
Do đó Ta lấy B2 là phần bù B1 $\Rightarrow B2 = W - B1 = B1^c$

Ta lấy X bù Erosion với B2 ra 2 chấm dưới.

Lấy phép giao của 2 bước trên ta được vị trí của n

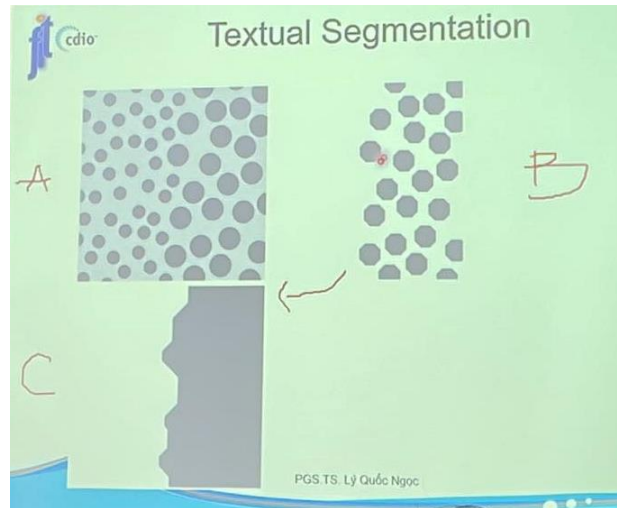
IV. Tuần 4

1/ Làm sao để từ A -> B, A -> C?



Từ A qua B chúng ta sẽ dùng phép giãn nở độ xám (dilation), từ A qua C chúng ta sẽ dùng phép co độ xám (erosion). Vùng nào bị tối bên ảnh gốc, sau khi dùng phép dilation thì những vùng tối được sáng lên bên ảnh mới, và ngược lại với phép erosion, vùng sáng sẽ bị tối lại khi qua ảnh mới

2/ Làm sao để từ A -> B -> C?



Đóng hình ảnh đầu vào bằng cách sử dụng các phần tử cấu trúc lớn hơn succ. Khi kích thước (phần tử cấu trúc) \approx kích thước (đốm màu nhỏ), các đốm màu bị xóa

Mở một lần với phần tử cấu trúc lớn liên quan đến sự phân cách giữa các đốm màu lớn
 -> các mảng sáng giữa các đốm màu bị loại bỏ -> vùng sáng ở bên trái, vùng tối ở bên phải

Ngưỡng -> ranh giới

3/ Nhìn toán tử này thì opening, closing nó sẽ làm cái gì?

Opening thì sẽ bào được đỉnh cao, Closing thì sẽ bào được đỉnh thấp

4/ Giải thích tại sao các toán tử lại có thể bảo toàn hình thái, không làm biến dạng đối tượng khi áp dụng các toán tử vào?

Các toán tử đều hoạt động trên các bản sao hoặc giá trị của đối tượng ban đầu, thay vì làm việc trực tiếp trên đối tượng gốc. Điều này đảm bảo rằng đối tượng gốc không bị thay đổi sau khi áp dụng toán tử. Thay vào đó, kết quả của toán tử sẽ được gán cho một biến mới hoặc sử dụng ngay lập tức trong một biểu thức.

5/ Nhìn toán tử này sẽ làm gì?

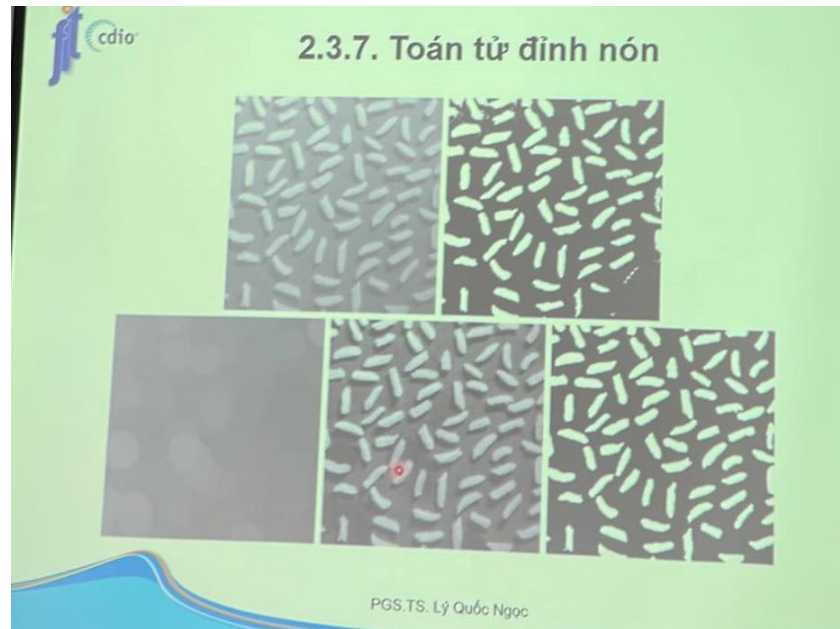
Giúp có thể đề lộ ra cái biên

6/ So sánh phép tìm biên cạnh Gradient với toán tử Grayscale Gradient?

Chúng ta có thể mô phỏng lại thế giới thực bằng các phương pháp ánh xạ các thực thể, đối tượng từ không gian vào màn hình vi tính, phương pháp phân dạng plasma và phân dạng đứt đoạn có thể đo vẽ lại địa hình với nhiều độ cao thay đổi.

Dùng các phép biến đổi hình học (tịnh tiến, quay, co giãn, đối xứng, trượt,..) và phép thể hiện thực thể hình học (phóng to, thu nhỏ, ViewToWin, WinToView, phép chiếu song song, phối cảnh,..) để có thể mô phỏng lại thế giới thực vào màn hình máy tính.

7/Tại sao hình ở trên xét thiếu còn hình ở dưới thì không?



Do opening xét từng vị trí điểm ảnh, phụ thuộc vào điểm lân cận của mỗi điểm ảnh đang xét, đánh đốm ra rồi mới thêm ngưỡng vào

Còn hình ở trên xét thiếu do là thêm ngưỡng vô trước ngay từ lúc bắt đầu, sau đó mới xét điểm lân cận

V. SEMINAR

1. Nhóm DC - Truy vấn ảnh y phục

a) Động lực nghiên cứu

Động lực: Nhận diện trang phục để nhận diện người giữa đám đông thay cho khuôn mặt vì đôi khi khuôn mặt không rõ:

Ý nghĩa khoa học:

- Giúp máy tính nhận biết và rút trích thông tin ở mức độ chi tiết.
- Góp phần vào nghiên cứu mô hình học máy.
- Góp phần vào bài toán truy vấn, đánh chỉ mục

Ý nghĩa thực tiễn:

- Tích hợp vào các ứng dụng thương mại điện tử
- Xác định, theo dõi người dùng trong khu vực đông người.
- Áp dụng vào nhiều lĩnh vực thị giác khác như: nhận dạng, định dạng đối tượng,...

b) Phương pháp

Nhận diện -> phân đoạn -> phân loại -> gán nhãn thuộc tính -> truy vấn

Giai đoạn	Tác vụ	Đầu vào	Đầu ra
Nhận diện	Tiên đoán sự xuất hiện của y phục	Tập dữ liệu N mẫu ảnh	Tập dữ liệu N được huấn luyện với độ chính xác cao
Phân đoạn	Phân đoạn ảnh thành các vùng	Ảnh chứa y phục dạng 2D	Ảnh được phân và đánh dấu khu vực chứa y phục
Phân loại	Tập dữ liệu N mẫu ảnh Tập dữ liệu được huấn	Hình ảnh y phục 2D	Tên loại y phục
Gán nhãn	Cho biết thuộc tính trong y phục cần truy vấn	Hình ảnh y phục 2D	Tên loại y phục và thuộc tính y phục
Truy vấn ảnh	Thực hiện việc tìm kiếm y phục	Hình ảnh y phục 2D	Những hình ảnh y phục tương đồng

Mô hình sử dụng: Faster RCNN

2. Nhóm 2T - Định vị và tái tạo môi trường xung quanh

a) Động lực nghiên cứu

Dữ liệu đầu vào: ảnh hoặc video RGB-D

Dữ liệu đầu ra: toạ độ và bản đồ của thiết bị trong môi trường

Ý nghĩa khoa học: Việc thu thập và trích xuất thông tin hình ảnh 3 chiều là một trong những vấn đề mà các nhà khoa học thị giác máy tính đã và đang hướng tới, nó là một xu hướng trong tương lai. Khi chúng ta thi được thông tin dạng 3 chiều chúng ta có thể trích xuất các thông tin như kích thước, toạ độ, ... của thế giới thực qua những bức hình một cách dễ dàng.

- Các nhà khoa học và nhà phát minh đã phát minh ra rất nhiều thiết bị chuyên dụng để làm điều này tiêu biểu là LiDar. Đây là những thiết bị chuyên dụng tương với độ chính xác cao, tuy nhiên nó lại rất đắt đỏ để sở hữu và nghiên cứu cá nhân.
- Từ đó chúng ta có giải pháp là tái tạo lại môi trường 3D từ môi trường 2D cụ thể phương pháp này là vSLAM. Tuy nhiên việc tái tạo lại chiều không gian thứ 3 chưa bao giờ là việc dễ dàng.
- Với bước phát triển của vSLAM gần đây đã đóng góp rất nhiều vào sự phát triển của các thuật toán và công nghệ khác. Nó phục vụ tốt hơn cho công việc xác nhận danh tính bằng khuôn mặt 3D thay vì 2D hay trong các hệ thống tạo dựng bản đồ của các robot tự hành.
- Bài báo cáo này sẽ tập trung vào nghiên cứu ưu nhược điểm, cách thức hoạt động của vSLAM và cách khắc phục nếu có.

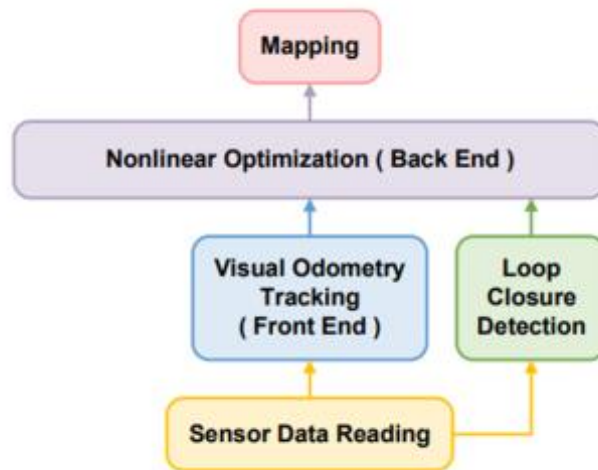
Thách thức:

- Lỗi về sai số trong các phép đo.
- Lỗi về mất thông tin điểm neo.
- Chi phí tính toán cao, phần cứng lại quá yếu.
- Sự thay đổi của môi trường
- Hạn chế về ánh sáng
- Hạn chế về quyền riêng tư

Ý nghĩa thực tiễn: Việc định vị và tái tạo môi trường xung quanh đóng vai trò rất lớn trong việc vận hành của các hệ thống tự hành, những hệ thống tự hành yêu cầu rất nhiều về

việc kiểm soát được vị trí trong môi trường như xe tự hành, robot tự hành. Bên cạnh đó SLAM còn được ứng dụng trong các robot tái tạo bản đồ của môi trường xung quanh, hay việc sử dụng việc tái tạo 3D để tái tạo cơ thể con người để đưa ra các đề xuất trang phục phù hợp với mỗi thân hình khác nhau.

b) Phương pháp



- **Sensor Data Reading:** bước này tập trung vào việc thu thập và tiền xử lý ảnh đầu vào từ camera của điện thoại.
- **Visual Odometry Tracking:** công việc của bước này là tính toán và ước tính chuyển động của camera và tìm cấu trúc của các hình gần kề nhau, liên kết những điểm trùng nhau lại và tạo thành một bản đồ cục bộ. Bước này còn có một tên gọi khác là front end.
- **Loop Closure Detection:** trong trường hợp camera quay lại vị trí cũ mà nó đã từng quét function này sẽ được kích hoạt và dừng vòng lặp, sau đó mọi thông tin camera sẽ được chuyển xuống back end.
- **Nonlinear Optimization:** ở bước này các hình dáng mà camera thu lại ở những thời điểm khác nhau sẽ được tập trung lại, kết hợp với thông tin về phát hiện đóng vòng lặp nó sẽ tối ưu hoá những thông tin trên. Sau đó xây dựng những toạ độ thống nhất trên phạm vi toàn cục. Vì nó được kết nối ngay sau bước front end nên nó còn được gọi là bước back end.
- **Mapping:** dựa vào những toạ độ đã được tính toán ở bước back end mà nó sẽ tạo ra một bản đồ hoàn chỉnh.

Mô hình sử dụng:

- Nhóm này không sử dụng mô hình máy học mà tập trung vào phương pháp toán và computer vision.
- Cụ thể: trong quá trình tái tạo và tự định vị. Thiết bị phải dùng tới 2 hệ toạ độ là hệ toạ độ thế giới (WCS) và hệ toạ độ camera (CCS). Mỗi khung hình khác nhau là một CCS khác nhau.
- Tựu chung lại việc chúng ta cần làm là tìm được phép biến đổi toạ độ camera từ thời điểm T sang thời điểm $T+1$. Có n phép biến đổi toạ độ như vậy. Và từ đó suy ngược lại vị trí từ thời điểm $T+1$ sang thời điểm T và từ thời điểm T về lại WCS để xây dựng một hệ toạ độ đồng nhất, xây dựng lên bản đồ cho thiết bị.
- Các bước thực hiện: Tìm đặc trưng trong từng ảnh -> so khớp đặc trưng giữa 2 ảnh liên quan nhau -> chuyển hoá các điểm đặc trưng vào toạ độ 3D -> tạo bản đồ và tự định vị -> loop closures -> tái tạo lại môi trường

3. Nhóm VIP2 - Xây dựng hệ thống điểm danh lớp học và lưu giữ đường cong cảm xúc khuôn mặt của sinh viên

a) Động lực nghiên cứu

Về khoa học:

- Tăng tính chính xác và hiệu quả của quá trình điểm danh
- Giám sát chuyên cần và hiệu suất học tập
- Phát triển công nghệ nhận dạng khuôn mặt và phân tích cảm xúc

Ứng dụng:

- Trong lĩnh vực giáo dục, hệ thống điểm danh tự động và lưu giữ đường cong cảm xúc có thể được sử dụng để giúp giáo viên và quản lý trường học quản lý chuyên cần của sinh viên
- Nghiên cứu về giảng dạy và học tập
- Quản lý nhân sự và đào tạo

b) Phương pháp

Sử dụng công nghệ nhận dạng khuôn mặt để xác định và nhận dạng sinh viên trong lớp học. Các thuật toán nhận dạng khuôn mặt có thể được sử dụng để so khớp khuôn mặt với cơ sở dữ liệu chứa thông tin của sinh viên.

Sử dụng camera để ghi lại hình ảnh và video trong lớp học. Camera có thể được đặt ở vị trí chiến lược để thu thập dữ liệu từ các góc độ khác nhau và đảm bảo tính chính xác trong việc nhận dạng khuôn mặt và phân tích cảm xúc.

Sử dụng công nghệ phân tích hình ảnh để xác định biểu hiện cảm xúc trên khuôn mặt của sinh viên. Các thuật toán và mô hình máy học có thể được áp dụng để phân loại và ghi lại các biểu hiện cảm xúc như vui, buồn, bất ngờ, lo lắng, và tức giận.

4. Nhóm Cực Cháy - Xây dựng hệ chuẩn đoán khả diễn bệnh ung thư xương (XAI)

a) Động lực nghiên cứu

Về khoa học:

- Tối ưu hóa quy trình chẩn đoán: Hệ chuẩn đoán XAI có thể tổng hợp và phân tích dữ liệu lâm sàng, hình ảnh y khoa, kết quả xét nghiệm và thông tin bệnh án từ nhiều nguồn khác nhau.
- Đồng bộ hóa thông tin: XAI giúp đồng bộ hóa thông tin giữa các cơ sở y tế và chuyên gia từ xa. Nhờ đó, các chuyên gia y tế có thể truy cập và phân tích dữ liệu từ xa, đưa ra những ý kiến và đề xuất điều trị cùng nhau, không bị ràng buộc bởi giới hạn địa lý.

Ứng dụng:

- Nhu cầu y tế: Ung thư xương là một loại ung thư hiếm gặp nhưng rất nghiêm trọng. Việc chẩn đoán sớm và chính xác là quan trọng để đưa ra phác đồ điều trị phù hợp.
- Cải thiện chẩn đoán: Bệnh ung thư xương là một loại ung thư hiếm gặp, khó chẩn đoán và thường được phát hiện ở giai đoạn muộn.

b) Phương pháp

- Thu thập dữ liệu: Đầu tiên, cần thu thập dữ liệu lâm sàng, hình ảnh y khoa và các tài liệu liên quan về bệnh ung thư xương từ các nguồn khác nhau. Các nguồn này có thể bao gồm hồ sơ bệnh án điện tử, hình ảnh chụp X-quang, MRI, CT scan và kết quả xét nghiệm.

- Tiền xử lý dữ liệu: Sau khi thu thập dữ liệu, quá trình tiền xử lý dữ liệu là cần thiết để làm sạch và chuẩn bị dữ liệu cho việc phân tích. Điều này có thể bao gồm việc loại bỏ dữ liệu nhiễu, đối chiếu và chuẩn hóa dữ liệu.
- Phân tích dữ liệu: Tiếp theo, sử dụng các phương pháp và thuật toán trí tuệ nhân tạo, bao gồm học máy và học sâu, để phân tích dữ liệu lâm sàng và hình ảnh y khoa. Các thuật toán này có thể áp dụng để xác định các đặc trưng quan trọng, phân loại bệnh nhân, dự đoán khả năng mắc bệnh ung thư xương và đề xuất phương pháp điều trị.
- Xây dựng mô hình XAI: Với việc áp dụng các phương pháp học máy giải thích, có thể xây dựng mô hình XAI để giải thích quyết định và tạo ra giải pháp khả thi cho chẩn đoán bệnh ung thư xương. Các mô hình này có thể cung cấp giải thích về cơ sở khoa học, đặc trưng quan trọng và mối liên hệ giữa các yếu tố trong việc chẩn đoán.

5. Nhóm Chiến thần - Tạo sinh ảnh dựa vào ảnh mẫu quảng cáo và ảnh đích

a) Động lực nghiên cứu

Về khoa học:

- Nghiên cứu hình ảnh và thị giác máy tính: Hệ thống Tạo sinh ảnh dựa trên ảnh mẫu quảng cáo và ảnh đích cung cấp một cơ sở để nghiên cứu và phát triển các thuật toán và công nghệ về xử lý hình ảnh và thị giác máy tính.
- Phát triển ứng dụng tạo ảnh tự động: Hệ thống Tạo sinh ảnh có thể được sử dụng để phát triển các ứng dụng tạo ảnh tự động trong nhiều lĩnh vực khác nhau.
- Gia tăng sự hiểu biết về học máy và học sâu: Hệ thống Tạo sinh ảnh là một ứng dụng của học máy và học sâu. Nghiên cứu về hệ thống này giúp chúng ta hiểu rõ hơn về cách mạng nơ-ron tích chập và các kiến trúc học sâu khác hoạt động, cũng như các vấn đề liên quan như quá khớp, tối ưu hóa và tạo ảnh sinh động.

Ứng dụng:

- Nâng cao hiệu quả quảng cáo: Bằng cách phát triển một hệ thống Tạo sinh ảnh thông minh, có thể tạo ra các hình ảnh quảng cáo tương tự với ảnh mẫu nhưng thích hợp với ảnh đích, hệ thống có thể giúp tăng cường hiệu quả của các chiến dịch quảng cáo.

b) Phương pháp

- Thu thập dữ liệu: Thu thập một bộ dữ liệu gồm các cặp ảnh mẫu quảng cáo và ảnh đích là bước đầu tiên. Bộ dữ liệu này cần đủ lớn và đa dạng để mô hình có thể học được các đặc trưng và quy luật trong quá trình tạo ảnh.
- Xây dựng mô hình học sâu: Sử dụng mô hình học sâu như mạng nơ-ron tích chập (CNN) và mạng sinh (GAN) để xây dựng hệ thống. Mạng CNN được sử dụng để trích xuất đặc trưng từ ảnh, trong khi mạng GAN được sử dụng để tạo ra ảnh mới dựa trên ảnh mẫu và ảnh đích.
- Huấn luyện mô hình: Sử dụng bộ dữ liệu đã thu thập, huấn luyện mô hình học sâu bằng cách cung cấp ảnh mẫu quảng cáo và ảnh đích làm đầu vào. Quá trình huấn luyện này nhằm tối ưu hóa các tham số của mô hình để nó có thể tạo ra các ảnh đích tương tự với ảnh thực tế.
- Đánh giá và điều chỉnh: Sau khi huấn luyện, đánh giá hiệu suất của mô hình bằng cách sử dụng các phép đo như độ tương đồng, độ tương tự hay độ chân thực. Nếu kết quả không đạt yêu cầu, có thể điều chỉnh kiến trúc mô hình, thay đổi các siêu tham số hoặc tăng cường bộ dữ liệu huấn luyện để cải thiện hiệu suất.

6. Nhóm TT - Xây dựng hệ thống truy vấn ảnh dựa vào hỏi đáp

a) Động lực khoa học

- Đưa ra các giải pháp truy vấn ảnh thông minh: Xây dựng một hệ thống truy vấn ảnh dựa vào hỏi đáp giúp người dùng có thể tương tác với hình ảnh một cách tự nhiên và thông minh hơn.
- Hiểu sâu hơn về mối quan hệ giữa ngôn ngữ và hình ảnh: Xây dựng hệ thống truy vấn ảnh dựa vào hỏi đáp yêu cầu khả năng hiểu sâu về mối quan hệ giữa ngôn ngữ và hình ảnh. Nghiên cứu trong lĩnh vực này có thể khám phá cách ngôn ngữ được sử dụng để mô tả hình ảnh, và từ đó xây dựng các mô hình đại diện ngôn ngữ và hình ảnh.

b) Động lực ứng dụng

- Tiềm năng thương mại: Xây dựng một hệ thống truy vấn ảnh dựa vào hỏi đáp có thể mang lại tiềm năng thương mại lớn. Công nghệ này có thể được ứng dụng trong các nền tảng thương mại điện tử, quảng cáo trực tuyến, và các dịch vụ tìm kiếm ảnh chuyên nghiệp. Điều này tạo ra động lực kinh tế để nghiên cứu và phát triển hệ thống truy vấn ảnh dựa vào hỏi đáp.

- Thách thức kỹ thuật: Xây dựng một hệ thống truy vấn ảnh dựa vào hỏi đáp đòi hỏi kỹ thuật cao, bao gồm xử lý ngôn ngữ tự nhiên, trích xuất đặc trưng ảnh, xếp hạng kết quả và hiểu ngữ cảnh. Các thách thức này thúc đẩy nghiên cứu và phát triển thuật toán và phương pháp mới để giải quyết những vấn đề phức tạp trong lĩnh vực này.
- Ứng dụng thực tế: Hệ thống truy vấn ảnh dựa vào hỏi đáp có thể được áp dụng trong nhiều lĩnh vực, bao gồm công nghệ, y tế, giáo dục, quảng cáo, và nhiều lĩnh vực khác. Việc nghiên cứu và phát triển hệ thống này sẽ mang lại lợi ích và ứng dụng rộng rãi cho nhiều ngành công nghiệp và cộng đồng.