KỊCH BẢN SEMINA

1. Mở đầu:

* Xin kính chào quý thầy cô và toàn thể các bạn sinh viên đã tham dự hội nghị cán bộ NCKH ngày hôm nay
* Sau đây em xin giới thiệu một đề tài vô cùng cấp thiết trong hoàn cảnh dịch bệnh như hiện nay đó là “nhận diện cảm xúc của sinh viên dựa trên biểu cảm khuôn mặt trong giáo dục trực tuyến”, dưới sự hướng dẫn của TS. ĐTT và sv thực hiện là em Phạm Quang Huy
* Nội dung trình bày của em gồm có 4 phần chính: ……

1. Giới thiệu

* Trong hoàn cảnh dịch bệnh nghiêm trọng như hiện nay, tính đến ngày 22/10 có tới 4.92tr ca tử vong do dịch bệnh covid 19, hầu hết các trường học đề phải đóng cửa để đảm bảo an toàn cho sinh viên cũng như cán bộ nhân viên trong trường.
* Chính vì vậy giáo dục trực tuyến được xem như là giải pháp tốt nhất trong hoàn cảnh dịch bệnh nghiêm trọng như hiện nay
* Vậy thì chúng ta có rất nhiều ứng dụng lên quan đến vấn đề biểu cảm trên khuôn mặt như quản lý sức khỏe, hay hỗ trợ lái xe
  + Và đã có nhiều chứng minh chỉ ra rằng cảm xúc có ảnh hưởng đến khả năng tiếp thu và nhận thức thông tin
  + Hai nhà tâm lý học là Ekman và Friesen đã chỉ ra rằng: con người chúng ta có thể nhận thức được 6 loại cảm xúc cơ bản, …, ngoài ra thì có thêm 1 loại cx nữa đó là cảm xúc bình thường
  + Bên dưới là 1 số hình ảnh đc gán nhãn cảm xúc trong csdl fer2013 là một bộ csdl ko kiểm soát được thu thập từ các môi trường phức tạp.
* Dưới sự phát triển của lĩnh vực trí tuệ nhân tạo và học sâu, nhiều thuật toán nhận diện cảm xúc được đề xuất để nhận diện các biểu cảm trên khuôn mặt, các mô hình trí tuệ nhân tạo đã cho thấy một hiệu suất tốt trong bài toán nhận diện nói chung
* Để tạo nên thành công đó thì các mô hình chủ yếu được xây dựng dựa trên cấu trúc CNN, là một mô hình phổ biến trong bài toán nhận dạng và phân loại hình ảnh nói chung và bài toán nhận diện cảm xúc nói riêng.
* Quay trở lại bài toán như đã giới thiệu: Vấn đề khó khăn nhất giữa lớp học trực tuyến với lớp học truyền thống là gì? Chắc hẳn chính bản thân chúng ta ngồi đây cũng đã và đang được trải qua những ngày tháng học trực tuyến
* Một trong những nhược điểm của hình thức học tt mà chắc hẳn thầy cô nào hay bất kỳ bạn sinh viên nào cũng đều có thể chỉ ra đó là: lớp học tt thường đc xem như là ít ràng buộc hơn và giao tiếp kém hq hơn so với lớp học tt, bên phải màn ảnh của quý thầy cô là một ví dụ vui về nhược điểm của học tt (bạn nữ trong ảnh có 1 khuôn mặt trầm tư và k nói gì khi đc thầy cô hỏi như rồi bạn ấy lại nhờ sự giúp đỡ của mẹ)
* Chính vì vậy, ở đây em xin đưa ra một kết luận: chúng ta cần có những công cụ đánh giá chủ quan và khách quan nhằm nâng cao chất lượng giáo dục trực tuyến.

PHƯƠNG PHÁP

* Nhận dạng cảm xúc tự động hỗ trợ nâng cao chất lượng đào tạo là một trong những giải pháp đánh giá một cách khách quan và hoàn toàn dựa trên những cơ sở khoa học về mối liên hệ chặt chẽ giữa nét mặt và cảm xúc của con người
* Phương pháp mà em đưa ra gồm có bốn bước chính:
  + Thu thập hình ảnh đầu vào từ lớp học trực tuyến: hình ảnh bên dưới là minh họa một sinh viên trong lớp học
  + Sau đó ta đi tiến hành Phát hiện khuôn mặt trong ảnh
  + Tiền xử lý ảnh trước khi đưa vào nhận diện cảm xúc
  + Hình ảnh sẽ đi qua Một mô hình đánh giá cảm xúc đc xây dựng dựa trên mô hình cnn và đc đào tạo qua các bộ csdl có quy mô vừa và lớn
  + Ta thu được kết quả là một biểu đồ thống kê cảm xúc của sinh viên
* Những tiến bộ trong công nghệ đã ta ra một số lượng lớn các nền tảng giáo dục trực tuyến và tăng tính linh hoạt trong đào tạo, những nền tảng công nghệ này cho phép giáo viên áp dụng các phương tiện công nghệ cao mà không lo lắng về số lượng sinh viên trong lớp như các lớp học truyền thống, thậm chí là có thể ở trên bất kỳ vị trí địa lý nào cũng hoàn toàn có thể giao tiếp trong thời gian thực mà ko cần đến lớp. Hiện tại có 2 nền tảng học tt được sử dụng phổ biến tại trường đhsp hn là zoom và gg meet. Khi đó giảng viên có thể dễ dàng tg tác với sinh viên trong thời gian thực và cũng dễ dàng thu đc hình ảnh khuôn mặt của sv dựa trên camera thích hợp trên tbi như lap, đt, máy tính bảng. Các hình ảnh này có thể đc sử dụng như là tập dữ liệu đầu vào cho hệ thống đề xuất để có thể đánh giá và nhận diện cảm xúc của ng học theo thời gian thực
* Các hình ảnh đầu vào có thể chứa nhiều tt khác nhau ngoài hình ảnh khuôn mặt cần nhận diện, do đó, cần phải xác định đúng vị trí khuôn mặt trong ảnh trc khi tiến hành nhận diện. Trong nghiên cứu này để có thể phát hiện và cắt được chính xác vị trí khuôn mặt trong ảnh, Phương pháp haar-cascade đc ứng dụng dựa trên các đặt trưng haar cho phép phát hiện các khuôn mặt trong ảnh 1 cách nhanh chóng mà ko phụ thuộc vào tỷ lệ ảnh, các đặc điểm chính của khuôn mặt bao gồm lông mày, mắt, đầu mũi và miệng có thể đc phát hiện bằng các đường viền HCN cho phù hợp, các đặc điểm này đc xây dựng dựa trên các điểm đặc trưng của mọi mặt bao gồm cả mặt trên mặt dưới xác định chiều rộng dọc, ngoài cùng bên phải , ngoài cùng bên trái, xác định chiều ngang của hình ảnh khuôn mặt. Để tránh bỏ sót tt khuôn mặt cũng như giảm nhiễu nều, các đg viền của hcn xác định vị trí khuôn mặt sẽ đc đặt là 3px. Hình bên dưới là minh họa 1 ví dụ về khuôn mặt đc phát hiện dựa trên pp haar-cascade và đc tô viền xung quanh khuôn mặt.
* Sau phát hiện khuôn mặt trong ảnh đầu vào dựa trên phương pháp Haar-Cascade thì việc thực hiện nhận diện cảm xúc là hoàn toàn khả thi, Một ảnh mới (chỉ có khuôn mặt) sẽ được cắt ra để làm hình ảnh đầu vào cho bước nhận diện tiếp theo. Việc cắt hình ảnh khuôn mặt sẽ làm giảm bớt các chi tiết dư thừa trong ảnh, nâng cao hiệu suất nhận diện. Tuy nhiên, trong quá trình thực nghiệm, các kết quả cho thấy việc nhận diện cảm xúc vẫn chưa thực sự hiệu quả một phần là do chất lượng ảnh đầu vào chưa tốt (quá tối, hoặc nhiễu, …), một phần là do kích thước hình ảnh đầu vào khác nhau, nên kích thước ảnh khuôn mặt sau khi được phát hiện cũng sẽ khác nhau. Do đó, cần phải tiến hành thêm bước tiền xử lý để chuẩn hoá các ảnh khuôn mặt đầu vào trước khi tiến hành nhận diện. Một số thao tác tiền xử lý được thực hiện trong lược đồ đề xuất bao gồm: nâng cấp hình ảnh (dựa trên việc cân bằng histogram), giảm nhiễu với bộ lọc Gaussian. Và chuẩn hóa hình ảnh về cùng 1 kích thước đầu vào của mô hình nhận diện ảnh
* Có thể nói linh hồn của bài báo nghiên cứu này đó chính là mô hình nhận diện cảm xúc dựa trên mạng cnn do sự vượt trội về hiệu suất và độ chính xác của nó so với các cách tiếp cận khác.
  + Khối thứ nhất chứa 2 lớp tích chập mỗi lớp gồm 64 bộ lọc (channel); mỗi bộ lọc có kích thước cỡ 3 x 3 và kích thước ảnh đầu vào của bộ lọc có kích thước 48x48x1. Theo sau đó là hai lớp tổng hợp (pooling) có kích cỡ 2x2, bước nhảy là 2x2 và lớp dropout có tỷ lệ là 0.5 nhằm loại bỏ một vài trường hợp trong quá trình huấn luyện mạng. Việc bỏ các điểm đầu vào được thực hiện bằng cách lấy ngẫu nhiên nhưng đảm bảo một ngưỡng xác suất nào đó. Việc bổ sung thêm lớp dropout nhằm trách trường hợp overfiting trong quá trình huấn luyện.
  + Khối thứ hai có cấu trúc tương tự như khối thứ nhất bao gồm 2 lớp tích chập gồm 64 bộ lọc cỡ 3x3, một lớp tổng hợp pooling cỡ 2x2 với bước nhảy 2x2 và cuối cùng là một lớp dropout với tỷ lệ 0.5. Tuy nhiên, khác với khối thứ nhất, kích thước ảnh đầu vào bộ lọc khối thứ 2 sẽ giảm một nửa còn 24x24 để giảm độ phức tạp của thuật toán và tăng độ chính xác về việc trích chọn đặc trưng của ảnh.
  + Khối thứ ba về cơ bản cũng có cấu trúc tương tự như hai khối trước. Kích thước ảnh đầu vào cũng được tiếp tục giảm đi một nửa còn 12x12. Ngoài ra, hai lớp tích chập trong khối này được tăng cường số lượng kênh lên là 128 đồng thời bổ sung thêm lớp flatten nhằm làm phẳng dữ liệu và kết hợp các đặc trưng của ảnh đề có được đầu ra cho mô hình.
  + Khối cuối cùng bao gồm các lớp kết nối đầy đủ (fully connected layer) gồm 3 lớp. Lớp đầu tiên có 1024 nơ-ron, trong đó sử dụng hàm kích hoạt ReLUs. Lớp kết nối đầu đủ sau cùng gồm 7 nơ-ron và sử dụng hàm softmax làm hàm kích hoạt để phân loại các biểu cảm bao gồm: Tức giận, ghê tởm, sợ hãi, vui vẻ, buồn, ngạc nhiên, bình thường.
* Nhờ những trình tự thực hiện các bước nêu trên nghiên cứu đã thu lại được một số thành quả nhất định.
* Trước tiên là 2 bộ csdl được sử dụng để đào tạo cũng như sử dụng để kiểm thử đó là bộ csdl FER2013 một bộ cơ sở dữ liệu quy mô lớn được giới thiệu trong IMCL 2013 gồm 32298 ảnh và CK Plus là 1 bộ csdl có kiểm soát được tạo ra trong phòng thí nghiệm với 981 ảnh. 2 bộ dl này đc gán nhãn đầy đủ với 7 loại cảm xúc khác nhau là: tức giận,
* Trong quá trình đào tạo mô hình với 2 bộ csdl này thì công nghệ đc sd ở đây là ngôn ngữ lập trình python và được chạy trên gg colab với cấu hình phần cứng đc mô tả như hình dưới
* Cuối cùng nhờ có 2 bộ csdl trên đã cho ra một kết quả vô cùng khả thi như sau: với fer2013 mô hình cho ra tỷ lệ nhận diện đúng là 95,9% Với CK+ là 92,5% với thời gian trùng bình nhận diện của 2 bộ csdl này sấp xỉ khoảng 50-55ms
* Tuy nhiên để đảm bảo tính khách quan của cũng như chứng minh được tính khả thi của mô hình để có thể đưa vào áp dụng thực tiễn thì trong báo cáo nghiên cứu này em có thu thập được các hình ảnh đầu vào từ các lớp học trực tuyến tại khoa cntt trường đh sp hn và bảng thông tin bên dưới là kết quả của một vài lớp học tiêu biểu như lớp học 1 số vđe xhoi của cô kiều pthuyf và 2 lớp phần mền nhúng và di động, phát triển pmem cho tbi di độgn k69 của giảng viên trần hải long mà nhờ mô hình nhận diện cảm xúc nên trên đã cho ra kết quả như sau: đối với môn 1 số vde kết quả là:, ….
* Chúng ta cùng nhìn một ví dụ ảnh thực tế trong quá trình nghiên cứ đánh giá ở đây em lấy ví dụ lớp học 1 số vde xã hội của cntt, hình ảnh bên trái là input đầu vào được chụp từ lớp học trực tuyến trên nền tảng google meet, hình ảnh bên phải là kết quả ảnh đầu ra sau khi nhận diện cảm xúc và đánh nhãn
* Và cuối cùng chúng ta sẽ tổng kết lại các kết quả từ lớp học ví dụ trên bằng một biểu đồ thống kê như sau:
* Có 22 học sinh là có cảm xúc bình thường, 4 học sinh là vui vẻ 1 học sinh là sợ hãi, điều này cho thấy tỷ lệ cảm xúc là bình thường chiếm tới 81.48% trong tổng số cảm xúc khuôn mặt đã đánh nhãn được.
* Từ những kết quả thực nghiệm với 2 bộ csdl và các lớp học trực tuyến tại khoa cntt trường đhsp hn nên trên đã cho thấy đây là một kết quả vô cùng khả thi kết quả của thí nghiệm này có thể hỗ trợ thuật lợi cho hoạt động của mô hình khi áp dụng vào môi trường thực tế.
* Tổng kết lại thì báo cáo nghiên cứ này đã đề xuất được một lược đồ hỗ trợ việc nhận diện cảm xúc của người học trong lớp học trực tuyến một cách tự động từ đó giúp cải thiện chất lượng lớp học trực tuyến dựa trên những cảm xúc được phát hiện và giáo viên có thể kịp thời theo dõi và thay đổi phương pháp giảng dạy sao cho hiệu quả nhất.
* Phát hiện mặt người dựa trên các đặc trưng haar và mạng tích chập cnn trong việc nhận diện cảm xúc
* Các kết quả thu được với độ chính xác trên 90% với các bộ csdl kiểm thử, phát hiện hiệu quả 59 trên tổng số 123 sinh viên trong 3 lớp học trực tuyến tại khoa cntt trường đhsp nêu trên
* Tuy nhiên thì vần có đâu đó những vấn đề bất cập và mặt hạn chế còn tồn tại trong đề tài.
* Số lượng khuôn mặt phát hiện trong ảnh lớp học hạn chế do đk ánh sáng, góc chụp hay việc nhận diện với tỷ lệ chính xác cao nhưng còn có thể cải tiến thêm.
* Trong thời gian tới: việc nâng cấp khả năng phát hiện khuôn mặt trong đk hạn chế là vấn đề cấp thiết cần được giải quyết, tiếp theo sau đó là một mô hình phát triển ưu việt hơn có độ chính xác cao hơn sẽ được tạo ra nhằm tạo ra một công cụ thực sự có thể áp dụng vào thực tiễn với độ phủ sóng cao trong các trường học nhằm cải thiện chất lượng lớp học online và giải quyết những vấn đề nhức nhối được xem như là nhược điểm lớn của lớp học online so với lớp học tt. Ngoài ra các thuật toán xử lý ảnh cũng như tốc độ nhận diện ảnh sẽ được tối ưu hơn trong thời gian tới.
* Trên đây là phần trình bày của em về đề tài “nhận diện ….” Xin trân trọng cảm ơn quý thầy cô cùng toàn thể các bạn sv trong hội nghị đã chú ý lắng nghe. Em xin kết thúc phần trình bày của mình tại đây, xin mời thầy cô và các bạn cùng đưa ra câu hỏi đề chúng ta thảo luận và đưa ra ý kiến về đề tài này.

**Bộ lọc Gauss** được cho **là bộ lọc** hữu ích nhất, được thực hiện bằng cách nhân chập ảnh đầu vào với một ma trận **lọc Gauss** sau đó cộng chúng lại để tạo thành ảnh đầu ra. Ý tưởng chung **là** giá trị mỗi điểm ảnh sẽ phụ thuộc nhiều vào các điểm ảnh ở gần hơn **là** các điểm ảnh ở xa.