BÀI BÁO NGHIÊN CỨU

NHẬN BIẾT CÁM XÚC CỦA HỌC SINH DƯA TRÊN BIỂU CẢM KHUÔN MẶT TRONG GIÁO DỤC TRỰC TUYẾN DỰA TRÊN QUAN ĐIỂM CỦA MÔ PHỎNG MÁY TÍNH

**Weiqing Wang,1 Kunliang Xu,1 Hongli Niu,1 and Xiangrong Miao 2**

Trường Kinh tế và Quản trị, Đại học Bách Khoa Bắc Kinh, Bắc Kinh 100083, Trung Quốc

Trung tâm Bảo vệ Môi trường và Bảo tồn Năng lượng Bắc Kinh, Bắc Kinh 101199, Trung Quốc

Thư từ nên được gửi đến Xiangrong Miao; [miaoxr@bjbeec.cn](mailto:miaoxr@bjbeec.cn)

Nhận ngày 14 tháng 7 năm 2020; Sửa đổi ngày 26 tháng 8 năm 2020; Được chấp nhận thông qua ngày 28 tháng 8 năm 2020; Xuất bản ngày 11 tháng 9 năm 2020

Biên tập viên học thuật: Zhihan Lv Bản quyền © 2020 Weiqing Wang et al. , là một bài báo truy cập mở được phân phối theo Giấy phép Ghi công Creative Commons, cho phép sử dụng, phân phối và tái sản xuất không hạn chế ở bất kỳ phương tiện nào, miễn là tác phẩm gốc được trích dẫn chính xác.

Giáo dục trực tuyến đã phát triển nhanh chóng do sự tiện lợi không thể thay thế của nó. Trong các trường hợp nghiêm trọng do COVID- 19 gần đây, nhiều trường học trên thế giới đã trì hoãn việc mở cửa và áp dụng giáo dục trực tuyến như một trong những phương pháp giảng dạy chính. Tuy nhiên, hiệu quả của các lớp học trực tuyến từ lâu đã bị đặt nhiều dấu hỏi. So với các lớp học trực tiếp truyền thống, các khóa học trực tuyến thiếu sự giao tiếp và phản hồi trực tiếp, kịp thời và hiệu quả giữa giáo viên và học viên. Các nghiên cứu trước đây đã chỉ ra rằng có một mối quan hệ chặt chẽ và ổn định giữa nét mặt và cảm xúc của một người nói chung. Từ góc độ mô phỏng máy tính, một khuôn khổ kết hợp thuật toán nhận dạng biểu cảm khuôn mặt (FER) với các nền tảng khóa học trực tuyến được đề xuất trong công trình này. , máy ảnh điện tử trong thiết bị được sử dụng để thu thập hình ảnh khuôn mặt của học sinh và các biểu hiện trên khuôn mặt được phân tích và phân loại thành 8 loại cảm xúc bằng thuật toán FER. Một khóa học trực tuyến bao gồm 27 sinh viên được thực hiện trên Tencent Meeting được sử dụng để kiểm tra phương pháp được đề xuất và kết quả đã chứng minh rằng phương pháp này hoạt động mạnh mẽ trong các môi trường khác nhau. , là khuôn khổ cũng có thể được áp dụng cho các tình huống tương tự khác như các cuộc họp trực tuyến.

1. GIỚI THIỆU

Biểu cảm trên khuôn mặt là một trong những tín hiệu mạnh mẽ, tự nhiên và phổ biến nhất để con người truyền tải các trạng thái cảm xúc và ý định của họ dẫu bất kể biên giới quốc gia, chủng tộc và giới tính [1, 2], và có nhiều ứng dụng liên quan đa dạng như quản lý sức khỏe [3], hỗ trợ lái xe [4, 5], và các ứng dụng khác [6–9]. Trong các nghiên cứu trước đó về biểu hiện cảm xúc trên khuôn mặt, Ekman và Friesen lập luận rằng con người nhận thức một số cảm xúc cơ bản theo cùng một cách bất kể nền tảng văn hóa của họ, và họ đã xác định các biểu hiện trên khuôn mặt điển hình thành 6 loại: tức giận, ghê tởm, sợ hãi, hạnh phúc, nỗi buồn, và sự ngạc nhiên [10, 11]. Dựa trên các nghiên cứu về Ekmanand Friesen, Ekman và Heider [12–14], Matsumoto [15] đã cung cấp đủ bằng chứng cho một biểu cảm khuôn mặt phổ biến khác, đó là sự khinh bỉ.

Ngoài ra, FER2013 [16], một cơ sở dữ liệu quy mô lớn và không bị giới hạn được giới thiệu trong những thách thức ICML 2013 trong Representation Learning, đã gắn nhãn các hình ảnh trên khuôn mặt của nó thành giận dữ, ghê tởm, sợ hãi, hạnh phúc, buồn bã, ngạc nhiên và trung lập, đã được sử dụng rộng rãi trong thiết kế hệ thống nhận dạng nét mặt (FER). Trong các nghiên cứu tiếp theo, mặc dù các nhà nghiên cứu đã giới thiệu nhiều mô hình có thể cung cấp nhiều loại cảm xúc hơn để đối phó với sự phức tạp và tinh tế của nét mặt [17–20], phân loại mô tả cảm xúc cơ bản một cách riêng biệt vẫn là phương pháp được sử dụng rộng rãi nhất trong FER do tính tổng quát và định nghĩa trực quan của biểu cảm khuôn mặt [21], và Hình 1 hiển thị 8 kiểu biểu cảm cơ bản trên khuôn mặt từ bộ dữ liệu CK + [22] và FER2013 [16]. Để xác định nét mặt, Ekman và Friesen [23] đã đề xuất Hệ thống mã hóa hành động trên khuôn mặt (FACS) để xác định nét mặt, dựa trên thực tế là biểu cảm là kết quả của sự thay đổi các bộ phận trên khuôn mặt. Với sự hỗ trợ của máy tính, các phương pháp tiên tiến hơn đã được đề xuất trong những thập kỷ qua [24], và các điểm đặc trưng có thể được nhìn thấy trong Hình 2.

Hình 1: Các kiểu biểu cảm cơ bản trên khuôn mặt từ CK + và FER2013: từ trái sang phải là giận dữ, khinh thường, ghê tởm, sợ hãi, hạnh phúc, trung tính, buồn bã và ngạc nhiên

Hình 2: Các điểm phân tích khuôn mặt được sử dụng để biểu hiện trên khuôn mặt [25]

Với sự phát triển của trí tuệ nhân tạo và việc học sâu rộng, nhiều thuật toán FER đã được đề xuất để xử lý thông tin biểu hiện trong các biểu diễn trên khuôn mặt, điều này đã cải thiện dần độ chính xác của nhận dạng và đạt được hiệu suất tốt hơn so với các phương pháp truyền thống [26, 27]. , các tác vụ điện tử của FER chủ yếu có thể được chia thành hai loại: hình ảnh tĩnh (được biểu thị bằng ảnh) [28–30] và chuỗi động (được biểu thị bằng video) [31–33] có tính đến mối quan hệ động giữa các hình ảnh thay đổi liên tục và do đó đặt ra những thách thức bổ sung hơn so với trước đây. Ngoài các phương pháp dựa trên thị lực, các kỹ thuật sinh trắc học khác [34, 35] cũng có thể được áp dụng để hỗ trợ nhận dạng biểu hiện.

Cơ sở dữ liệu đào tạo được dán nhãn đầy đủ bao gồm nhiều biến thể của quần thể và môi trường nhất có thể là điều quan trọng đối với các nhà nghiên cứu để thiết kế và thử nghiệm mô hình hoặc hệ thống FER; các cơ sở dữ liệu hiện có chủ yếu được chia thành có kiểm soát và không kiểm soát. Mặt khác, các cơ sở dữ liệu được kiểm soát, được đại diện bởi CK + [22], Jaffe [36], và MMI [37], được thu thập từ môi trường phòng thí nghiệm với đủ ánh sáng và nền đơn giản. Ngày nay, vì hầu hết các cảnh thực tế đều phức tạp và có thể thay đổi do các yếu tố như ánh sáng, FER trong phòng thí nghiệm hoặc môi trường được kiểm soát thường được coi là ít có ý nghĩa thực tế và được sử dụng chủ yếu để chứng minh khái niệm về các phương pháp khai thác và phân loại tính năng. Mặt khác, các cơ sở dữ liệu không được kiểm soát, chẳng hạn như FER2013 [15] và AFEW [38], được thu thập từ các môi trường phức tạp với nền tảng, khớp cắn và độ chiếu sáng rất khác nhau; những cảnh này giống với tình huống thực tế hơn và ngày càng được sử dụng nhiều hơn trong các nghiên cứu.

Bị giới hạn bởi phần cứng và khả năng xử lý không đủ, phần lớn các phương pháp truyền thống cho FER sử dụng các tính năng thủ công hoặc học nông, chẳng hạn như các mẫu nhị phân cục bộ (LBP) [28] và phân tích nhân tử ma trận không âm (NMF) [39]. Với sự phát triển của khả năng xử lý và mô phỏng máy tính, tất cả các loại thuật toán học máy, chẳng hạn như Mạng thần kinh nhân tạo (ANN), Máy vectơ hỗ trợ (SVM) và bộ phân loại Bayes, đã được áp dụng cho FER và độ chính xác cao đã được xác minh trong môi trường được kiểm soát để có thể phát hiện các khuôn mặt một cách hiệu quả. Tuy nhiên, các phương pháp này yếu về khả năng khái quát hóa trong khi đây là chìa khóa để đánh giá tính thực tiễn của một mô hình [40]. Các thuật toán học sâu có thể giải quyết vấn đề này và nó cũng mạnh mẽ trong các môi trường không được kiểm soát. Các công trình gần đây đã chỉ ra rằng mạng nơ-ron tích tụ (CNN), do tính hiệu quả của chúng trong các tác vụ trích xuất và phân loại tính năng, đã hoạt động tốt trong việc giải quyết các vấn đề về thị giác máy tính, đặc biệt là trong FER [41, 42], và nhiều mô hình dựa trên cấu trúc CNN được đề xuất liên tục và đã đạt được kết quả tốt hơn so với các phương pháp trước đây. Simonyan và Zisserman [43] đã thông qua kiến ​​trúc của các bộ lọc tích chập rất nhỏ (3 × 3) để tiến hành đánh giá toàn diện các mạng với độ sâu ngày càng tăng và hai mô hình ConvNet hoạt động tốt nhất đã được công bố công khai để tạo điều kiện cho các nghiên cứu sâu hơn trong lĩnh vực này. Bằng cách tăng chiều sâu và chiều rộng của mạng trong khi vẫn giữ ngân sách tính toán không đổi, Szegedy et al. [44] đã giới thiệu một kiến ​​trúc mạng nơ-ron phức hợp sâu có tên là “Khởi đầu” trong đó việc sử dụng tài nguyên máy tính có thể được cải thiện đáng kể, và Jahandad et al. [45] đã hoạt động trên 2 kiến ​​trúc mạng nơ-ron phức hợp (Inceptionv1 và Inception-v3) dựa trên “Inception” và chứng minh rằng 2 mô hình này hoạt động tốt hơn các mô hình khác và Inceptionv1 với mạng sâu 22 lớp hoạt động tốt hơn mạng Inception-v3 42 lớp sâu khi đối mặt với hình ảnh đầu vào có độ phân giải thấp và hình ảnh 2D của chữ ký; tuy nhiên, Inception-v3 hoạt động tốt hơn trong thử thách ImageNet. , xu hướng chung của mạng nơ-ron là tăng độ sâu của mạng và độ rộng của lớp. Về lý thuyết, các mô hình mạng nơ-ron càng sâu thì khả năng học càng mạnh nhưng càng khó đào tạo. Ông và cộng sự. [46] đề xuất một khung học tập dư để giảm độ khó đào tạo của các mạng sâu hơn và đã chứng minh kỹ lưỡng rằng các mạng dư này dễ tối ưu hóa hơn trong khi tăng độ chính xác từ độ sâu tăng lên đáng kể. Ngoài ra, một bộ phận các nhà nghiên cứu đề xuất rằng độ chính xác của nhận dạng có thể được cải thiện hơn nữa bằng cách kết hợp CNN với mạng nơ-ron tái diễn (RNN) trong đó CNN được sử dụng làm đầu vào cho RNN [47, 48].

Trong những thập kỷ qua, giáo dục trực tuyến đã phát triển nhanh chóng cho dù ở các trường đại học hay cơ sở đào tạo [49], mang lại cơ hội ứng dụng tiềm năng cho FER.

Khác biệt đáng kể so với các khóa học trực tiếp truyền thống, các khóa học trực tuyến thường được coi là ít ràng buộc hơn và giao tiếp hiệu quả, điều này chắc chắn sẽ dẫn đến sự nghi ngờ của giảng viên đối với phương pháp giáo dục mới lạ này [50, 51] trong khi có một số nghiên cứu cho rằng kết quả học tập của sinh viên đạt được bằng giáo dục trực tuyến có thể tương đương với các khóa học trực diện truyền thống [52, 53], ngoại trừ các kỹ năng đòi hỏi độ chính xác tối ưu và mức độ nhận thức xúc giác cao hơn [54]. Không thể phủ nhận rằng tốc độ phát triển nhanh chóng của giáo dục trực tuyến có thể mang lại sự thuận tiện và linh hoạt cho nhiều học sinh hơn, vì vậy nó cũng có không gian phát triển rộng rãi trong tương lai; do đó, làm thế nào để đảm bảo rằng học sinh giữ được mức độ tập trung và hiệu quả học tập như các khóa học truyền thống trong quá trình giáo dục trực tuyến là rất quan trọng để thúc đẩy sự phát triển hơn nữa của giáo dục trực tuyến.

Tóm lại, đóng góp chính của bài báo này như sau. Bằng cách kết hợp các nền tảng giáo dục trực tuyến hiện có với mô hình nhận dạng nét mặt dựa trên kiến trúc của mạng nơ-ron phức hợp, công trình này đã đề xuất một khuôn khổ cho phép theo dõi thời gian thực cảm xúc của sinh viên trong các khóa học trực tuyến và đảm bảo rằng phản hồi được thể hiện bằng nét mặt có thể cung cấp cho giáo viên kịp thời để họ có thể linh hoạt điều chỉnh chương trình dạy học và cuối cùng là nâng cao chất lượng và hiệu quả của giáo dục trực tuyến.

2. Khung đề xuất

Khung chủ yếu bao gồm hai phần: nền tảng các khóa học trực tuyến, trong bài báo này, chúng tôi lấy Cuộc họp Tencent làm ví dụ cho thử nghiệm chế độ và mô hình học sâu dựa trên CNN, lấy cảm hứng từ Kuo và cộng sự. [27], trước đó có lưu ý rằng các hình ảnh gốc được thu thập từ các khóa học trực tuyến cần phải được xử lý trước, bao gồm nhận diện khuôn mặt, căn chỉnh, xoay và thay đổi kích thước, tùy theo các yếu tố khác nhau trong hình ảnh gốc. Hình 3 trình bày quy trình của FER, và các bước chi tiết của khung đề xuất như sau: đầu tiên, các máy ảnh tích hợp trong các thiết bị điện tử được sử dụng để chụp ảnh khuôn mặt của các sinh viên đang theo học. Thứ hai, thuật toán nhận dạng nét mặt được đào tạo bởi cơ sở dữ liệu biểu cảm khuôn mặt tiêu chuẩn là được sử dụng để phát hiện các khuôn mặt và phân loại các biểu hiện trên khuôn mặt như tức giận, ghê tởm, sợ hãi, hạnh phúc, buồn bã, ngạc nhiên, khinh thường và trung tính. Thứ ba, biểu đồ phân bố xác suất của biểu thức được vẽ và cung cấp cho giáo viên để có thể điều chỉnh kế hoạch dạy học kịp thời.

*2.1. Nền tảng Giáo dục Trực tuyến*. Những tiến bộ trong phương thức phân phối công nghệ đã tạo ra một số lượng lớn các nền tảng giáo dục trực tuyến và cải thiện đáng kể tính linh hoạt của giáo dục, cho phép giáo viên áp dụng các phương pháp kỹ thuật đa dạng để hỗ trợ giảng dạy mà không phải lo lắng về giới hạn số lượng học sinh trong khóa học dựa trên lớp học truyền thống và học sinh ở các khu vực khác nhau có thể giao tiếp trong thời gian thực mà không cần phải xem xét đến lưu lượng truy cập và các vấn đề khác. Các tài liệu giảng dạy tương tự như các lớp học truyền thống có thể được tải lên các nền tảng này để sinh viên tham khảo. Hiện tại, trong các nền tảng có chức năng giảng dạy trực tuyến, chẳng hạn như DingTalk, Zoom và Rain lớp học, giáo viên có thể áp dụng phương pháp họp video và tận dụng camera tích hợp trong thiết bị để chụp và nhận dạng nét mặt của học sinh trong thời gian thực. Các hình ảnh được chụp sẽ được xử lý trước và sau đó được sử dụng làm đầu vào của CNN.

*2.2. Tiền xử lý dựa trên IntraFace*. Xử lý trước hiệu quả có thể làm giảm sự can thiệp của các vật thể giống như khuôn mặt trong nền khi phát hiện khuôn mặt trong một hình ảnh và sau đó chuẩn hóa các hình ảnh khuôn mặt theo kiến ​​thức heuristic, điều này sẽ giúp nâng cao hiệu quả của mô hình học sâu. Chúng tôi đã sử dụng IntraFace [55], một gói phần mềm có sẵn công khai tích hợp các thuật toán để theo dõi đặc điểm khuôn mặt, ước tính tư thế đầu, phát hiện thuộc tính khuôn mặt, v.v., làm công cụ xử lý trước. Như trong Hình 4, IntraFace cũng có thể được sử dụng để phát hiện nhiều khuôn mặt cùng một lúc. , Các đặc điểm chính của từng khuôn mặt bao gồm lông mày, mắt, đầu mũi và miệng có thể được nhận ra một cách hiệu quả, và biểu hiện trên khuôn mặt có thể được phát hiện bằng các đường viền hình chữ nhật cho phù hợp; những đường viền này được xây dựng bởi các điểm đặc trưng ở cạnh của mọi mặt, bao gồm cả mặt trên và mặt dưới, xác định chiều rộng dọc, ngoài cùng bên phải và ngoài cùng bên trái, xác định chiều ngang của hình ảnh khuôn mặt. Để tránh bỏ sót thông tin trên khuôn mặt đồng thời giảm nhiễu nền, chúng tôi phóng to đường viền hình chữ nhật lên 1,05 lần để che được nhiều nội dung trên khuôn mặt hơn. Hơn nữa, xem xét kích thước hình ảnh đầu vào mô hình học tập được đặt trước là 48 × 48, hình ảnh được phát hiện sẽ được xoay với đầu mũi làm trung tâm và được thay đổi kích thước phù hợp để phù hợp với kích thước đầu vào.

2.3. Mô hình học tập dựa trên CNN. Kiến trúc của mô hình học sâu ứng dụng dựa trên CNN được minh họa trong Hình 5, dựa trên kết quả nghiên cứu do Kuo et al đề xuất. [27], và hiệu suất trước đây của mô hình này trong FER so với mô hình tương tự khác cũng đã được chứng minh. Sau lớp tích hợp gồm 32 bản đồ đối tượng, lớp đầu vào được theo sau bởi 2 khối, bao gồm 2 lớp tích tụ và 1 lớp tổng hợp tối đa với 64 bản đồ đối tượng riêng biệt. Và kích thước của các hạt nhân trong lớp tích chập đầu tiên được đặt thành 3 × 3, lớp thứ hai là 5 × 5, các lớp gộp tối đa đều bao gồm một hạt nhân có kích thước 2 × 2 và bước 2, và do đó, đầu vào hình ảnh sẽ được nén thành một phần tư. Và có 2 lớp được kết nối đầy đủ sau đây gồm 2048 và 1024 nơ-ron, trong đó Đơn vị tuyến tính chỉnh lưu (ReLU) [56–59] được sử dụng làm chức năng kích hoạt. Để ngăn chặn việc trang bị quá mức, một Dropout được thêm vào sau mỗi lớp trong số 2 lớp được kết nối đầy đủ, lớp này sẽ giải phóng một phần tế bào thần kinh theo xác suất rơi cài đặt trước; trong bài báo này, cả 2 giá trị đều được đặt thành 0,5. , e lớp đầu ra sau bao gồm 8 đơn vị và softmax [60] được sử dụng làm hàm kích hoạt để phân loại các biểu hiện được kiểm tra theo các khía cạnh giận dữ, ghê tởm, sợ hãi, hạnh phúc, buồn bã, ngạc nhiên, khinh thường và trung tính.

Hình 3: Quá trình của FER.

Hình 4: Đầu ra tự động của IntraFace để đo phản ứng của khán giả. [55]

Hình 5: Cấu trúc của mô hình CNN được áp dụng.

Mô hình đề xuất được đào tạo bởi cơ sở dữ liệu Jaffe, CK + và FER2013 bao gồm 8 biểu thị cơ bản trên cùng một lúc. Bởi vì cơ sở dữ liệu FER nhỏ thường chỉ chứa vài trăm hình ảnh, rõ ràng điều này là không đủ cho việc đào tạo mô hình; chúng tôi áp dụng chiến lược tăng cường trực tuyến với cả lật ngang và dịch chuyển ngẫu nhiên để tăng hình ảnh của các tập huấn luyện. Chi tiết hơn về mô hình CNN được đưa ra trong Bảng 1.

Và trong mô hình này, kích thước đầu ra N của mỗi lớp tích chập có thể được xây dựng dưới dạng

N = I - F + 2P/S + 1, (1)

trong đó I, F, P và S biểu thị kích thước đầu vào, kích thước hạt nhân, kích thước đệm và kích thước sải chân, tương ứng.

Trong mỗi lớp tổng hợp tối đa, kích thước phần đệm là 0 và kích thước đầu ra cũng có thể được biểu thị bằng

N =I-F/S + 1. (2)

Các đơn vị tuyến tính đã được chỉnh lưu (ReLU) được sử dụng làm chức năng kích hoạt trong các lớp chập và các lớp chia tối đa để tránh bùng nổ gradient và đảm bảo nhanh hơn tốc độ hội tụ trong quá trình lan truyền ngược, có thể được xây dựng dưới dạng

f (x) = 0, x≤0; 􏼨 x, x> 0. (3)

Softmax được sử dụng làm chức năng kích hoạt trong lớp đầu ra, đầu vào là ma trận zi =wT i x + bi

Bảng 1: Các thông số của mô hình CNN áp dụng

đầu ra từ các lớp được kết nối đầy đủ. Công thức là đưa ra như

Si = ……, (4)

trong đó K đại diện cho kích thước đầu ra của lớp, nghĩa là có K loại kết quả và Si đại diện cho xác suất của kết quả i, i = 1, 2 ,. . . , K.

Và mất Softmax, được sử dụng để lấy độ dốc và cập nhật, có thể được tính bằng cách sau:

CÔNG THỨC SỐ (5)

trong đó L biểu thị hàm mất mát và yi là biến nhãn, giá trị của nó là 1 hoặc 0 tùy theo đầu ra có phù hợp với giá trị thực tế hay không.

3. THÍ NGHIỆM VÀ KẾT QUẢ

Để kiểm tra hiệu suất của khung được đề xuất trong các ứng dụng thực tế, chúng tôi đã chụp một hình ảnh bao gồm 27 người từ một cuộc họp trực tuyến được tổ chức trên Tencent Meeting và nhập nó vào mô hình CNN. , là hình ảnh được chụp trước khi kết thúc cuộc họp; người điều hành đã thực hiện một bài phát biểu kết luận trong một bầu không khí vui vẻ. Ngoài ra, mọi người đều được thông báo rằng cuộc họp sắp kết thúc, theo thí nghiệm được thực hiện bởi Tonguç và Ozkara [61], mức độ hạnh phúc của học sinh sẽ được cải thiện đáng kể trong vòng vài phút trước khi kết thúc bài giảng, do đó, tương tự hoàn cảnh, có thể suy ra rằng cảm xúc của hầu hết các khuôn mặt trong hình ảnh này là vui vẻ hoặc trung tính. Hình 6 cho thấy đầu vào (bên trái) và đầu ra (phải) hình ảnh của mô hình CNN. Có thể thấy rõ ràng từ kết quả rằng tất cả các khuôn mặt đã được nhận dạng và đánh dấu bằng các đường viền ngoài hình chữ nhật, và các biểu cảm khuôn mặt đáp ứng cũng được gắn nhãn. Trong tổng số 27 khuôn mặt, 10 khuôn mặt được gắn nhãn “vui vẻ”, 15 khuôn mặt được gắn nhãn “trung tính” và 2 khuôn mặt được gắn nhãn “buồn”, lưu ý rằng hình ảnh thứ 2 ở dòng cuối cùng và hình ảnh thứ 3 ở dòng thứ 4 từ cuối cùng, được đánh dấu bằng các đường viền màu đỏ, không được phát hiện bởi các đường viền chính xác; nguyên nhân có thể là hình ảnh 2 khuôn mặt chưa hoàn thiện nên các đặc điểm hiển thị không đủ để nhận diện. Hình 7 cho thấy biểu đồ phân bố xác suất của cảm xúc, từ đó chúng ta có thể quan sát tổng thể các cảm xúc một cách trực quan và phán đoán trạng thái cảm xúc của lớp cho phù hợp. Cần lưu ý rằng xác suất hạnh phúc cao hơn đáng kể so với khuôn mặt trung tính trong hình này, trong khi những khuôn mặt được dán nhãn "hạnh phúc" ít hơn "trung tính" như được thể hiện trong Hình 6. Sự khác biệt có thể được giải thích như sau: có thể có các tính năng của nhiều biểu cảm trên một khuôn mặt cùng một lúc; biểu cảm hiển thị trên khuôn mặt này sẽ được gắn nhãn theo biểu cảm có khả năng xảy ra nhất do các đặc điểm này quyết định, nhưng biểu cảm tổng thể của một hình ảnh bao gồm nhiều khuôn mặt được quyết định bởi tổng số các đặc điểm biểu cảm khác nhau có trong mỗi khuôn mặt. Ở một số khuôn mặt được đánh dấu là "hạnh phúc", xác suất hạnh phúc có thể cao hơn nhiều so với trung tính, trong khi ở một số khuôn mặt được đánh dấu là "trung tính", xác suất hạnh phúc có thể chỉ thấp hơn một chút so với trung tính. Nhìn chung, kết quả của thí nghiệm này có thể hỗ trợ thuận lợi cho hoạt động của mô hình khi áp dụng vào môi trường thực tế.

Hình 6:, hình ảnh đầu vào (a) và đầu ra (b) của mô hình FER.

Hình 7: Phân bố xác suất của cảm xúc.

Possibility: xác suất

Anger: sự tức giận Disgust: sự kinh tởm Fear: Nỗi sợ

Happiness: hạnh phúc Sadness: nỗi buồn Surprise: bất ngờ

Neutral: trung tính Contempt: sự khinh thường

4. KẾT LUẬN VÀ THẢO LUẬN:

Trong nghiên cứu này, bằng cách kết hợp các nền tảng khóa học trực tuyến và mô hình học sâu nhỏ gọn dựa trên kiến ​​trúc của CNN, chúng tôi xây dựng một khung phân tích cảm xúc của sinh viên theo nét mặt của họ từ góc độ mô phỏng máy tính. Kết quả tổng thể có thể được trình bày dưới dạng biểu đồ một cách trực quan và giáo viên có thể điều chỉnh chiến lược giảng dạy của mình cho phù hợp để nâng cao hiệu quả của việc giảng dạy trực tuyến. Với việc tham khảo các nghiên cứu của Ekman et al. và năm 2013, các cảm xúc được phân loại thành tức giận, ghê tởm, sợ hãi, hạnh phúc, buồn bã, ngạc nhiên, khinh thường và trung tính trong khuôn khổ đề xuất. Để xác minh khả năng áp dụng của khuôn khổ này trong môi trường thực tế, chúng tôi đã chụp một hình ảnh bao gồm hình ảnh khuôn mặt của tất cả những người tham gia cùng một lúc trong một cuộc họp trực tuyến thực sự; có 12 người tham gia cuộc họp này và thời gian bắt được xác định vào cuối cuộc họp. Tổng cộng 12 khuôn mặt đã được chụp, trong đó 11 khuôn mặt có thể nhận dạng hiệu quả chứa đủ các điểm đặc trưng. Bằng cách nhập hình ảnh này vào mô hình CNN được áp dụng, chúng tôi đã thu được các thẻ cảm xúc cho từng khuôn mặt hợp lệ và có được cảm xúc tổng thể tại thời điểm đó. Người ta đã chứng minh rằng khung có khả năng ứng dụng tốt trong hoạt động thực tiễn và đóng vai trò tích cực trong việc giải quyết các vấn đề như thiếu tính ràng buộc đối với học sinh, giáo viên không thể phản hồi kịp thời. Cuối cùng, nó sẽ góp phần nâng cao chất lượng giáo dục trực tuyến.

Bất chấp những lợi ích nêu trên, vẫn còn nhiều chỗ để cải thiện trong khung này và các ứng dụng của nó. Từ góc độ công nghệ, với sự phát triển của mô phỏng máy tính, các thuật toán có hiệu suất tốt hơn và thời gian hoạt động ngắn hơn, bao gồm các mô hình tiền xử lý và học sâu, sẽ liên tục được phát triển theo thời gian. Ví dụ: tiền xử lý hình ảnh bao gồm nhận diện khuôn mặt, căn chỉnh, xoay và thay đổi kích thước, nhưng khi gặp các vấn đề, chẳng hạn như ngược sáng, bóng đổ và sự không hoàn chỉnh của khuôn mặt do môi trường phức tạp gây ra, các phương pháp hiện tại này luôn bất lực và những thiếu sót này có thể được giải quyết trong tương lai. Hơn nữa, mặc dù mô hình CNN trong khung đề xuất hiện đang hoạt động tốt, nó sẽ được thay thế bằng các mô hình có khả năng học tập cao hơn và độ chính xác phân loại cao hơn trong tương lai. Để đảm bảo tính cạnh tranh của khuôn khổ trong thời gian dài hơn, nó cần được điều chỉnh và duy trì thường xuyên, đồng thời áp dụng các thuật toán và công nghệ tiên tiến hơn để cập nhật nó.

Ngoài ra, với số lượng lớn người tham gia các khóa học trực tuyến, chúng tôi không có cách nào để đảm bảo rằng tất cả mọi người đều giữ được mức độ tập trung cao và khi đó biểu hiện của học viên có thể không thể hiện đầy đủ cảm xúc của họ do những yếu tố chủ quan này. Thực hiện các biện pháp như đặt ngưỡng có thể lọc ra một số thông tin không hợp lệ và làm nổi bật cảm xúc chính trong ảnh. Cuối cùng, kết quả là hiệu quả giảng dạy có thể được cải thiện.

**Tính sẵn có của dữ liệu**

Dữ liệu được sử dụng trong bản thảo này có thể được truy cập bởi độc giả qua BaiduPan của các tác giả tại https://pan.baidu.com/s/ 1dbKUfeKp5joeYh4wSOU7Qw với mã trích xuất "Qin6."

**Xung đột lợi ích**

Các tác giả tuyên bố rằng họ không có xung đột lợi ích.

**Góc nhìn nhận**

Công việc được hỗ trợ bởi Nghiên cứu về Ảnh hưởng thị trường chứng khoán Trung Quốc trong điều kiện kinh tế không chắc chắn (Số FRF-DF-20-11).