

BỘ NÔNG NGHIỆP VÀ PHÁT TRIỂN NÔNG THÔN  
HỌC VIỆN NÔNG NGHIỆP VIỆT NAM  
KHOA CÔNG NGHỆ THÔNG TIN



**BÁO CÁO TỔNG KẾT**  
**ĐỀ TÀI SINH VIÊN NGHIÊN CỨU KHOA HỌC**  
**NĂM 2024**

**TÊN ĐỀ TÀI:**

**“ỨNG DỤNG DEEP LEARNING NHẬN DIỆN CẢM XÚC”**

**MÃ SỐ: SV 2024-02-06**

**Nhóm sinh viên thực hiện:**

<b>Phạm Văn Thắng</b>	<b>Mã SV:641213</b>	<b>K64ATTT</b>
<b>Ngô Sách Tiến</b>	<b>Mã SV:645211</b>	<b>K64ATTT</b>
<b>Nguyễn Minh Quý</b>	<b>Mã SV:6666480</b>	<b>K66CNTTC</b>

**Giáo viên hướng dẫn: TS.Nguyễn Trọng Kương**

**HÀ NỘI - 2024**

# **Tên đề tài: ỨNG DỤNG DEEP LEARNING NHẬN DIỆN CẢM XÚC**

**Tên tiếng Anh:** APPLICATION OF DEEP LEARNING IN EMOTION RECOGNITION

**Nhóm các thành viên nghiên cứu:**

Phạm Văn Thắng, Ngô Sách Tiến, Nguyễn Minh Quý.

## **Tóm tắt:**

Trí tuệ nhân tạo hiện nay được áp dụng thành công trong nhiều lĩnh vực, một trong số đó là thị giác máy tính. Trong nghiên cứu này, một mạng nơ-ron học sâu (DNN) được sử dụng cho nhận diện cảm xúc trên khuôn mặt (FER). Một trong những mục tiêu của nghiên cứu này là xác định các đặc điểm quan trọng trên khuôn mặt mà MTCNN tập trung vào để nhận diện cảm xúc. Cụ thể, nhóm nghiên cứu đã sử dụng mạng nơ-ron xếp tầng đa nhiệm (MTCNN), cùng với mô hình nhận diện khuôn mặt của OpenCV và tập dữ liệu cảm xúc của sinh viên Học viện Nông nghiệp Việt Nam để huấn luyện mô hình đào tạo, phát hiện các khuôn mặt trên màn hình và đưa ra kết quả cảm xúc khi nhận diện. Các kiểm tra cho thấy với tập dữ liệu chạy thử, mô hình của MTCNN được huấn luyện có độ chính xác là 73,3% và khi được kiểm tra trên hình ảnh thực tế có độ chính xác là 77,4%. Kết quả của nghiên cứu này sẽ giúp cải thiện các nguyên lý của học sâu và hỗ trợ nâng cao độ chính xác trong thị giác máy tính.

## **Từ khóa:**

Học sâu, mạng nơ-ron tích chập đa nhiệm, nhận diện cảm xúc trên khuôn mặt, mạng nơ-ron học sâu.

## **Tóm tắt tiếng Anh:**

Artificial intelligence is currently being successfully applied in various fields, one of which is computer vision. In this study, a deep neural network (DNN) is utilized for facial emotion recognition (FER), which aims to identify human emotions based on facial expressions. One of the primary goals of this research is to determine the key facial features that the Multi-task Cascaded Convolutional Neural Network (MTCNN) focuses on for effective emotion recognition. Specifically, we employed MTCNN alongside OpenCV's facial recognition model and the emotion datasets of the Vietnam National University of Agriculture students to train the model, detect faces on the screen, and provide real-time emotion recognition results. The model was rigorously tested, and results show that using the test datasets, the MTCNN-trained model achieved an accuracy of 73.3%. When tested on real-world images, it demonstrated a higher accuracy of 77.4%. These findings will contribute to

improving deep learning techniques and advancing accuracy in computer vision applications, particularly in emotion recognition systems.

**Keywords:** deep learning, Multi-task Cascaded Convolutional Neural Network, facial emotion recognition, deep neural network.

## 1. Đặt vấn đề

Trong giao tiếp của con người, các biểu cảm khuôn mặt chứa đựng những thông tin phi ngôn ngữ quan trọng có thể cung cấp thêm manh mối và ý nghĩa cho giao tiếp bằng lời nói. Một số nghiên cứu (Paul Ekman & cộng sự., 2007) đã cho rằng 60–80% giao tiếp là phi ngôn ngữ. Thông tin phi ngôn ngữ này bao gồm biểu cảm khuôn mặt, giao tiếp bằng mắt, giọng điệu, cử chỉ tay và khoảng cách vật lý [1]. Đặc biệt hơn, phân tích biểu cảm khuôn mặt đã trở thành một chủ đề nghiên cứu phổ biến trong giai đoạn hiện nay. Bài toán FER là những bài toán thú vị và thu hút nhiều nghiên cứu với kết quả tích cực trong lĩnh vực thị giác máy tính, ứng dụng rộng rãi của các bài toán này như giám sát trạng thái người lái xe, giám sát người dùng điện thoại, hệ thống giám sát tại các cơ sở y tế và trong giáo dục [2]. Tuy nhiên, bài toán FER vẫn còn nhiều thách thức do sự đa dạng của những người có nét mặt giống nhau và sự thể hiện biểu cảm trên khuôn mặt của mỗi người có thể thay đổi liên tục theo nhiều thời gian.

Trong tâm lý học và thị giác máy tính, cảm xúc được phân loại theo mô hình phân loại hoặc mô hình chiều (cảm xúc và mức độ kích thích). Trong mô hình phân loại, Ekman và cộng sự đã định nghĩa các cảm xúc cơ bản của con người là hạnh phúc, tức giận, ghê tởm, sợ hãi, buồn bã và ngạc nhiên. Trong mô hình chiều, cảm xúc được đánh giá bằng các thang đo số liên tục để xác định cảm xúc và mức độ kích thích. FER là một nhiệm vụ quan trọng trong thị giác máy tính với nhiều ứng dụng thực tiễn, và số lượng nghiên cứu về FER đã tăng lên trong những năm gần đây, nhờ vào sự tiến bộ của mạng nơ-ron sâu.

Nghiên cứu chung của hệ thống FER có thể chia theo hai giai đoạn chính. Đầu tiên là thực hiện trích xuất các đặc trưng của hình ảnh khuôn mặt đại diện cho các điểm trên khuôn mặt và biểu cảm tương ứng. Tiếp theo là phân loại các đặc trưng đó vào các đặc điểm khuôn mặt và biểu cảm. Việc trích xuất đặc trưng khuôn mặt cho nghiên cứu FER đặc biệt quan trọng vì nó ảnh hưởng đến độ chính xác khi nhận dạng.

Một trong những thách thức khi nhận diện cảm xúc là sự phức tạp của cảm xúc của con người và sự biến đổi của các yếu tố ngoại cảnh (ánh sáng, góc nhìn, đồ họa máy quay,...). Mặc dù các phương pháp Deep Learning (DL) đã đạt được nhiều kết quả ấn tượng, vẫn còn tồn tại những vấn đề về tính chính xác, tốc độ xử lý, và khả năng khái quát hóa của mô hình [3]. Vì vậy, cần có nhu cầu nghiên cứu và phát triển các mô hình Deep Learning hiệu quả hơn trong việc nhận diện cảm xúc khuôn mặt, đồng thời cải thiện khả năng xử lý trên nhiều điều kiện khác nhau của dữ liệu [4].

Về ứng dụng nhận diện cảm xúc, trong nghiên cứu gần đây (Shang Li & cs., 2020) nhóm tác giả đã nghiên cứu về Deep FER cho hai vấn đề khi nhận diện cảm xúc là huấn luyện quá vừa với sự thiếu hụt về dữ liệu đào tạo và các biến thể không liên quan đến biểu thức, như độ chiếu sáng, tư thế đầu và dự đoán danh tính [5].

Hiện nay, (Phạm Minh Quyền & cộng sự., 2021) các nhà nghiên cứu chủ yếu tiếp cận các vấn đề này dựa trên mạng nơron tích chập (CNN) với các mô hình hiện đại như ResNet, GoogleNet và chúng đều cho ra kết quả khả quan trong huấn luyện mô hình. Mặc dù các kết quả nhận dạng dùng CNN có tỷ lệ cao khi các kiến trúc mô hình ngày càng được cải tiến, tuy nhiên vẫn còn một số hạn chế cần phải được cải thiện, đặc biệt khi áp dụng vào ứng dụng thực tế [6]. Các mô hình CNN thường được thiết kế chỉ duy nhất cho từng bài toán và có độ phức tạp rất lớn đối với các tham số khổng lồ trong tài nguyên tính toán dữ liệu của ứng dụng trên thực tế [7].

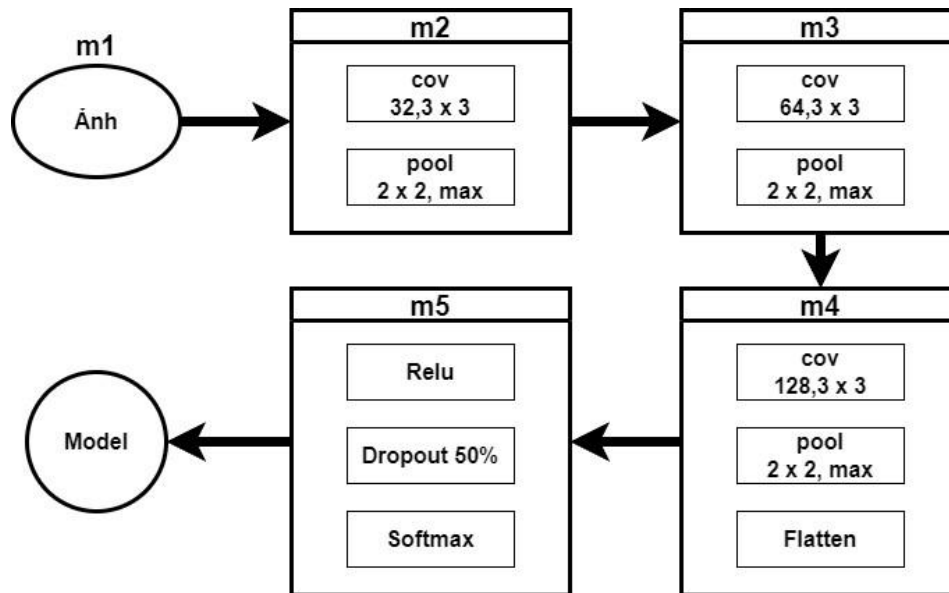
Trong nghiên cứu này tập trung thiết kế một mô hình DL cùng với mô hình MTCNN cho bài toán FER đồng thời điều chỉnh mức độ phức tạp vừa phải nhưng vẫn đảm bảo chất lượng nhận diện hiệu quả cho bài toán. Mô hình sẽ được chạy thử nghiệm với một số video camera được lưu trữ công khai và được thiết kế để hỗ trợ giám sát và quản lý quá trình đánh giá cảm xúc của người dùng.

## **2. Phương pháp nghiên cứu**

### **2.1 Mạng nơron học sâu**

#### **2.1.1 Mạng nơron tích chập đa nhiệm**

Mạng nơron tích chập đa nhiệm (Multi-task Cascaded Convolutional Networks - MTCNN)) là mô hình CNN sâu gồm 3 mạng nơron cải thiện chất lượng của đối tượng với một số nhiệm vụ có liên quan. Mô hình MTCNN thực hiện chia sẻ các tham số biểu diễn đặc trưng cho các lớp tích chập có mức sâu. Các mô hình MTCNN được nghiên cứu và xây dựng được đánh giá có hiệu quả trong các nhiệm vụ máy tính khác nhau [8]. Trong nghiên cứu này nhóm nghiên cứu mô hình MTCNN (Sơ đồ 1) để thực hiện nhiệm vụ nhận dạng cảm xúc. Mô hình được thiết kế được chia làm hai giai đoạn chính gồm đầu tiên là các đặc trưng của khuôn mặt được định danh và các biểu cảm trên khuôn mặt, tiếp theo là phân loại các đặc trưng thành các nhãn để phân lớp tương ứng với các định dạng được thực hiện. Số lớp mô hình và số lượng các nơron mỗi lớp có ảnh hưởng đến chất lượng mô hình và độ phức tạp khi tính toán trọng số trong đó. Vì vậy nhóm nghiên cứu và thiết kế mô hình này với số lượng lớp vừa phải để phù hợp với bài toán xử lý.



Sơ đồ 1: Cấu trúc MTCNN.

### 2.1.2 Cấu trúc mô hình và kết quả phân lớp

Kiến trúc của mô hình MTCNN sử dụng phương pháp chia sẻ tham số nhằm giảm kích thước và độ phức tạp của mô hình để ứng dụng mô hình cho các bài toán có định dạng hạn chế. Để giảm kích thước mô hình, nhóm nghiên cứu thiết kế số lớp tích chập (Conv) là 3, sau mỗi loại tích chập sử dụng phép gộp tín hiệu ở dạng lớn nhất (MaxPool). Chia các lớp nơron này thành 3 khối gồm m1, m2, m3 có cấu trúc cơ bản như nhau, mỗi khối này có một lớp tích chập theo sau nó là một phép gộp tín hiệu. Ở khối m1 là ảnh đầu vào, để phù hợp với camera của hình ảnh thiết bị đầu cuối có độ phân giải vừa phải và giảm kích thước tham số của mô hình, hình ảnh được đặt kích thước đầu vào là H (cao) x W (rộng) x D (sâu) = 48 x 48 x 1.

Các bộ lọc của lớp Conv có kích thước là 3x3, và ở lớp MaxPool có kích thước là 2x2. Các lớp Conv sử dụng hàm kích hoạt “ReLU” nhằm cho phép loại bỏ các giá trị âm, tăng tốc độ huấn luyện của mô hình. Ngoài ra để giảm thiểu hiện tượng quá khớp trong trọng số của các nơron trong học sâu nhóm nghiên cứu sử dụng kỹ thuật loại bỏ ngẫu nhiên kết nối của các nơron theo tỷ lệ 50% trong mỗi lớp Conv tăng dần kích thước của bộ lọc từ 32, 64, 128 nhằm tăng cơ hội trích chọn được nhiều các đặc trưng ẩn sâu bên trong hình ảnh ở các nơron mức độ sâu hơn.

Khối m5 có nhiệm vụ phân loại ảnh đầu vào đến các lớp tiếp theo của bài toán. Khối này có 2 lớp nơron được kết nối đầy đủ cho mỗi bài toán cần thực hiện, lớp đầu vào ẩn sử dụng hàm kích hoạt phi tuyến “ReLU” và lớp đầu ra kích hoạt hàm “softmax” để tính xác suất thuộc từng lớp cho mỗi hình ảnh đầu vào. Khối m6 có thêm cơ chế chuyển đổi dữ liệu đầu ra

từ các lớp tích chập và gộp thành một vector phẳng đến khối phân loại. Công thức tính đầu ra neuron phân lớp theo hàm kích hoạt softmax có dạng.

$$O_j^t = \text{softmax}(y_j^t) = \frac{e^{y_j^t}}{\sum_{k=1}^{M^t} e^{y_k^t}} \quad (1)$$

trong đó  $O_j^t$  là đầu vào của neuron thứ  $j^t$  của lớp ra tương ứng với nhiệm vụ  $t \in \{1, 2, \dots, T\}$ ,  $y_j^t$  là tổng các tín hiệu đầu vào của neuron thứ  $j^t$  trong lớp phân loại tương ứng với nhiệm vụ  $t$ ,  $M^t$  là số neuron lớp ra của nhiệm vụ  $t$ . Ở đây nhóm nghiên cứu thấy rõ tổng các giá trị đầu ra của các neuron đầu ra thuộc khối m6 cho mỗi bài toán có kết quả bằng 1, hay tổng  $O_j^t = 1$ .

Mô hình huấn luyện được theo phương pháp tối ưu hóa Adam, đây là một kỹ thuật tối ưu hóa được sử dụng rộng rãi bằng cách sử dụng các giá trị bình phương gradient để chia tỷ lệ học và sử dụng trung bình động của bước thay đổi gradient [9]. Cơ chế điều chỉnh trọng số của Adam thể hiện theo công thức.

$$w_{ijt} = w_{ijt-1} - \frac{n}{\sqrt{v^t} + \epsilon} * \widehat{m}^t \quad (2)$$

trong đó  $m^t$  và  $v^t$  là giá trị trung bình giảm theo cấp số nhân của gradient và của các gradient bình phương tại thời điểm học thứ  $t$ ,  $n$  là hệ số học (tốc độ huấn luyện) có hệ số bằng  $10^{-8}$ .

## 2.2 Thu thập dữ liệu

Dữ liệu được sử dụng trong nghiên cứu dựa trên sự tham gia của 50 sinh viên trong Học viện Nông nghiệp Việt Nam, và được sự đồng ý để sử dụng hình ảnh cho nghiên cứu. Kết quả các mẫu thu được 50 video cảm xúc bao gồm bảy trạng thái biểu cảm cảm xúc của con người. Ngoài ra bộ dữ liệu còn kết hợp bộ dữ liệu của các thư viện hình ảnh công khai để tạo nên một bộ dữ liệu khái quát và hoàn chỉnh cho mục đích nghiên cứu đề tài.

## 2.3 Tiền xử lý dữ liệu

Trong nghiên cứu về thị giác máy tính, để tiền xử lý, trích xuất đặc trưng và chuẩn hóa dữ liệu để phục vụ cho việc học và phân lớp các loại, trước hết cần xử lý các mẫu video được thu thập sang các hình ảnh. Nếu như sử dụng các video có sẵn để huấn luyện mô hình, dữ liệu được đưa vào sẽ vô cùng lớn, chúng không thể thay đổi sang các trạng thái chuẩn hóa khác nhau như đổi màu, cân bằng góc cạnh và khiến máy tính không thể vận hành cho bài toán tiếp theo.

Để xử lý việc cắt ảnh và hiệu chỉnh hình ảnh khuôn mặt, chúng em đã xây dựng công cụ tự động nhận diện và cắt hình ảnh khuôn mặt trong hộp giới hạn trong thư viện của

OpenCV. Các hình ảnh thu được sẽ điều chỉnh định dạng thành 224x224 pixel và lưu trong thư mục dữ liệu của nghiên cứu.

### 3. Kết quả

Từ dữ liệu thu thập được trong phần tiền xử lý, các hình ảnh được truyền vào trong mô hình mạng để huấn luyện và kiểm tra. Dữ liệu cho đào tạo và kiểm tra được nhóm nghiên cứu chia ngẫu nhiên theo tỷ lệ 4:1 tương ứng.

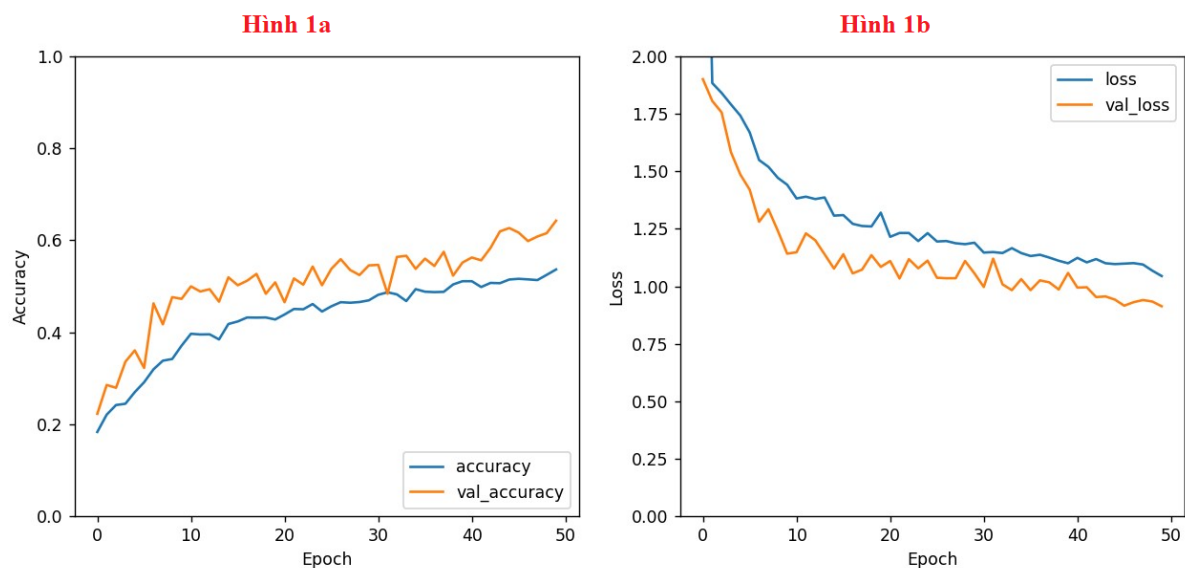
	Bình thường	Cười	Buồn	Tức giận	Ngạc nhiên
Tập huấn luyện	88%	67%	77%	43%	53%
Tập kiểm tra	82%	60%	49%	30%	45%

Bảng 1: Kết quả độ chính xác của hình ảnh cảm xúc trên tập dữ liệu

Để thấy được quá trình đào tạo của của mạng theo từng bước học và cập nhật thêm tham số, sau mỗi bước độ chính xác của phân lớp được kiểm tra và tính toán. Ở đây nhóm nghiên cứu tiến hành đào tạo trên 50 bước nhảy (Epoch), mỗi bước cập nhật lại tham số và đánh giá độ chính xác và hàm mất mát của mô hình (hình 1).

Sau khi huấn luyện, mô hình có thể được áp dụng để phân loại cảm xúc của một hình ảnh và video đầu vào. Một video thường bao gồm nhiều hơn một loại cảm xúc. Vì vậy để phát hiện những cảm xúc và tối ưu hóa tài nguyên khi nhận diện cảm xúc trong video, việc nhận diện được điều chỉnh và thay đổi nhân trong một khoản thời gian được nhập trên giao diện mà không cần thay đổi trong thời gian thực.

Kết quả thu được trên tập dữ liệu ảnh ngẫu nhiên, video và camera sẽ định dạng trên màn hình máy tính (hình 2).



Hình 1 a: Độ chính xác và b: Mất mát của mạng học sâu trong đào tạo





Hình 2: Mô tả một số thử nghiệm quá trình kiểm tra để đánh giá mô hình.

#### 4. Kết luận

Trong báo cáo này, nhóm nghiên cứu đã tìm hiểu, xây dựng các kỹ thuật về DL cũng như các kỹ thuật tiền xử lý dữ liệu. Dựa trên các kiến thức đã học, nhóm nghiên cứu kiểm chứng và lựa chọn mạng MTCNN cho bài toán nhận diện cảm xúc. Trong đó, các kết quả thực nghiệm tại (Bảng 1) cho thấy việc áp dụng mô hình DL từ dữ liệu có ảnh hưởng lớn đến độ chính xác khi sử dụng hình ảnh thực tế. Tuy nhiên nhóm nghiên cứu sử dụng mạng MTCNN đơn giản ở (Số đồ 1) để xây dựng hệ thống thử nghiệm nhận diện cảm xúc để phù hợp với cấu hình máy tính được nghiên cứu.

Các kết quả thử nghiệm trình bày ở (Hình 2) cũng cho thấy việc nhận diện khuôn mặt cho kết quả tốt trên hình ảnh được chụp và camera chạy trong môi trường với thời gian thực và có khả năng áp dụng trong những bài toán thực tế.

Trong nghiên cứu tiếp theo, nhóm nghiên cứu thử nghiệm trên nhiều bộ dữ liệu để đạt kết quả cao hơn như FER-2013, ResNet,... với hệ thống máy tính có cấu hình cao hơn. Ngoài ra nhóm cũng đề xuất xây dựng hệ thống nhận diện cảm xúc sinh viên có độ chính xác cao và đáp ứng được các yêu cầu về thời gian thực, phục vụ cho việc đánh giá mức độ giảng dạy và

cải thiện chất lượng học tập tại khoa Công nghệ Thông tin, Học viện Nông nghiệp Việt Nam để nâng cao trải nghiệm người dùng.

## **5. Tài liệu tham khảo**

- [2] Ayham Fayyouni1 & Anis Zarrad (2014), Novel Solution Based on Face Recognition to Address Identity Theft and Cheating in Online Examination Systems, Advances in Internet of Things.
- [7] Bùi Đức Thịnh, Trần Trọng Đạt, Nguyễn Thị Thủy (2018), Aerial Image Semantic Segmentation using Neural Search Network architecture.
- [9] Diederik P. Kingma & Jimmy Lei Ba (2015), Adam-A Method for Stochastic Optimization.
- [8] Dinh Viet Sang & Le Tran Bao Cuong (2018),Effective Deep Multi-source Multi-task Learning Frameworks for Smile Detection,Emotion Recognition and Gender Classification,Informatica.
- [4] Emad Barsoum,Sarah Bargal, Cristian Canton Ferrer, Cha Zhang (2016),Emotion Recognition in the Wild from Videos using Images.
- [1] Paul Ekman (2007), Emotions Revealed, Second Edition: Recognizing Faces and Feelings to Improve Communication and Emotional Life
- [6] Phạm Minh Quyền, Phùng Thanh Huy, Đỗ Duy Tân, Huỳnh Hoàng Hà, Trương Quang Phúc (2021), Nhận diện cảm xúc khuôn mặt dùng mạng nơ-ron tích chập CNN trên kit JETSON TX2.
- [5] Shang Li, Weihong Deng (2020). Deep facial expression recognition: A survey." IEEE Transactions on Affective Computing.
- [3] Võ Thị Hồng Nhung (2022).Phân tích biểu cảm mặt người dùng mạng nơron tích chập.

## Phụ lục

### 1. Chi tiết phần mềm

Phần mềm nhận diện cảm xúc được phát triển với sự tích hợp của các mô hình học sâu, cụ thể là mạng nơ-ron tích chập đa nhiệm (MTCNN) để phát hiện khuôn mặt và mô hình học sâu cho nhận diện cảm xúc. Dưới đây là các thành phần chính của phần mềm:

Ngôn ngữ lập trình: Python 3.x trở lên

#### Thư viện chính:

OpenCV: Để xử lý hình ảnh và phát hiện khuôn mặt.

TensorFlow/Keras: Để xây dựng và huấn luyện mô hình nhận diện cảm xúc.

MTCNN: Mô hình phát hiện khuôn mặt với độ chính xác cao.

NumPy, Matplotlib: Để xử lý dữ liệu và hiển thị kết quả.

#### Cấu trúc phần mềm:

Giao diện người dùng: Sử dụng Tkinter để tạo giao diện đồ họa.

Nhận diện cảm xúc: Dựa trên mô hình học sâu, phần mềm nhận diện các cảm xúc cơ bản như vui, buồn, tức giận, ngạc nhiên, tức giận, và bình thường.

Phân tích kết quả: Hiển thị kết quả nhận diện cảm xúc theo thời gian thực qua webcam, camera hoặc từ ảnh/video có sẵn.

#### Yêu cầu hệ thống:

Hệ điều hành: Windows

RAM: 4GB hoặc cao hơn.

GPU (tùy chọn): Hỗ trợ cho việc tăng tốc quá trình huấn luyện và nhận diện.

### 2. Hướng dẫn sử dụng phần mềm

Giao diện bên ngoài của phần mềm



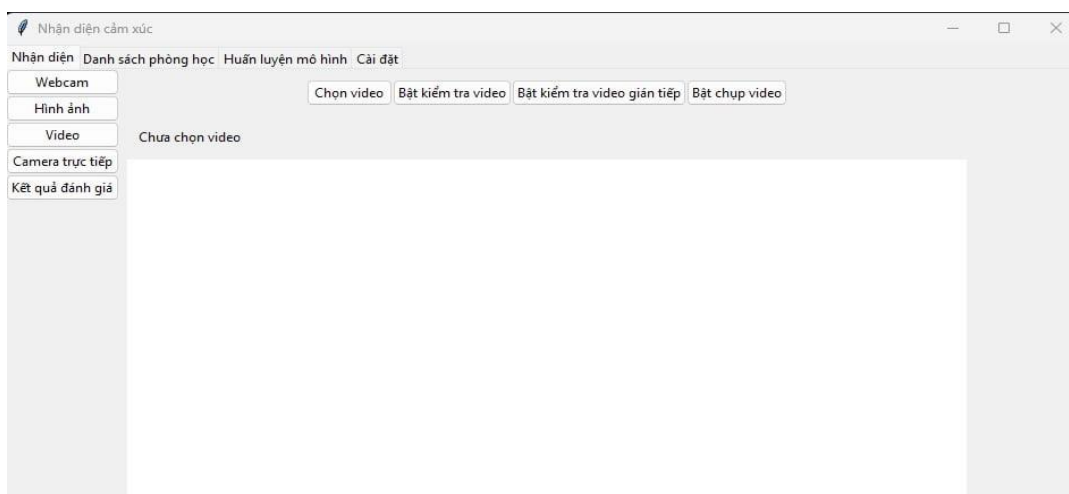
Giao diện chính của phần mềm: Các chức năng chính của webcam gồm bật webcam: mở webcam của máy tính và bật chụp webcam: chụp hình ảnh cảm xúc khuôn mặt khi đang nhận diện trên webcam



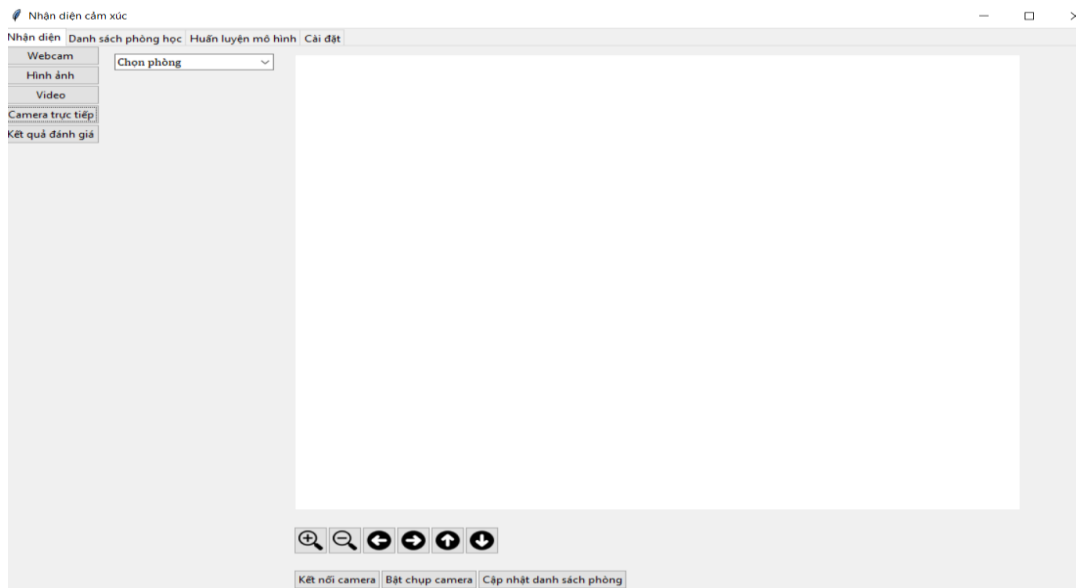
Giao diện nhận diện hình ảnh: Các chức năng chính của nhận diện hình ảnh bao gồm nhập số lượng ảnh cần kiểm tra, kiểm tra ảnh



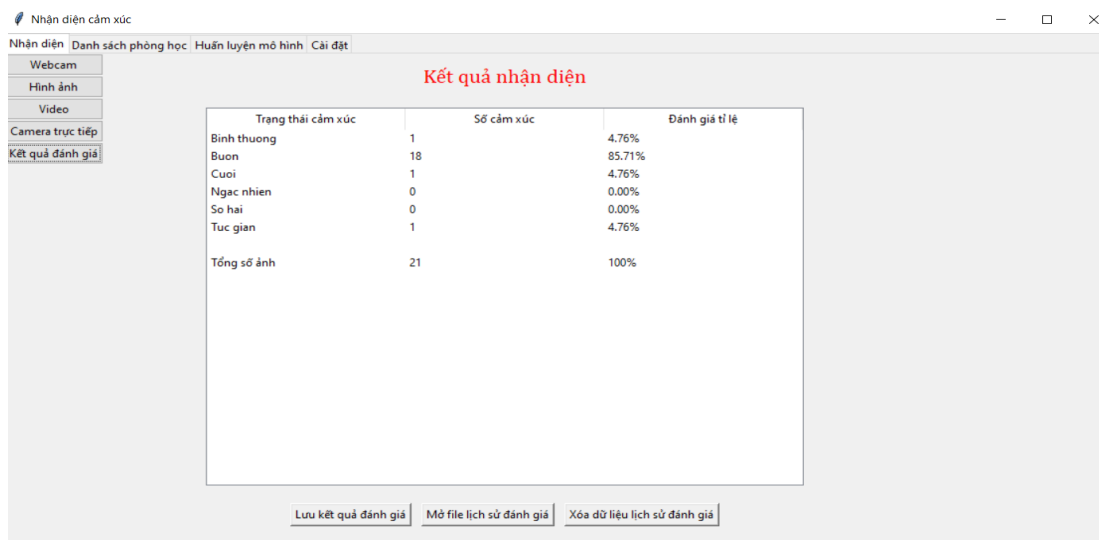
Giao diện video: Các chức năng bao gồm chọn video trong thư mục, bật video để kiểm tra video được chọn và xem trực tiếp, bật kiểm tra video gián tiếp dùng để kiểm tra video mà không cần xem trực tiếp trên màn hình.



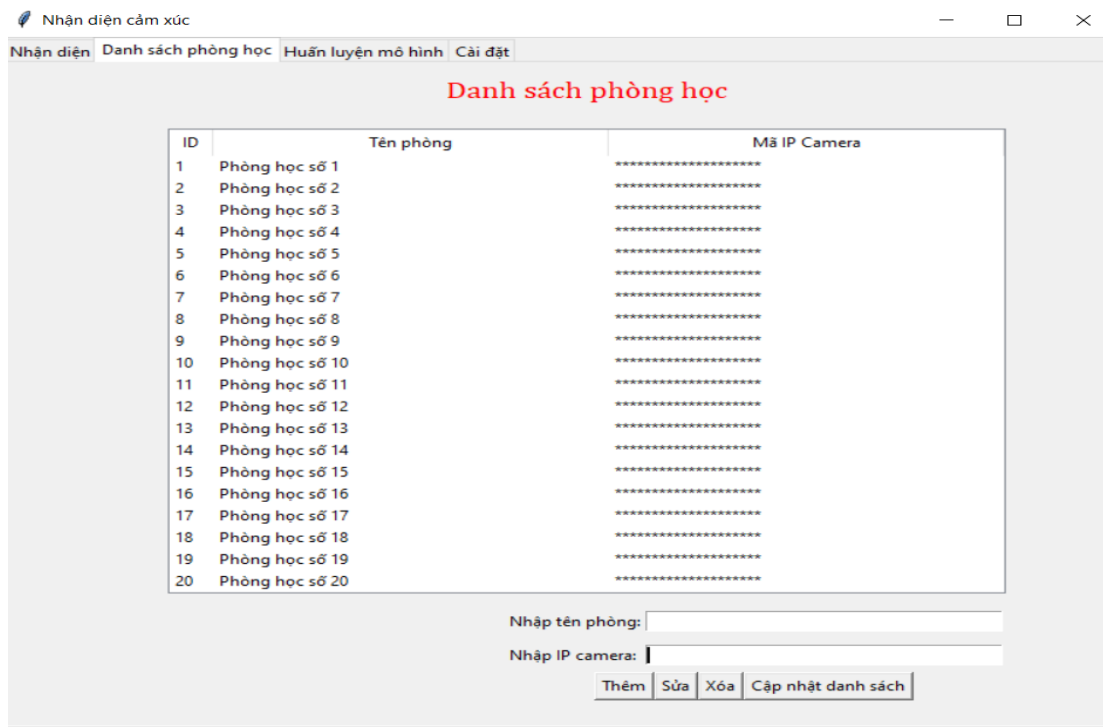
Giao diện camera trực tiếp: Các chức năng chính bao gồm lựa chọn phòng được cập nhật trong phần Danh sách phòng, kết nối camera với địa chỉ ip camera được kết nối.



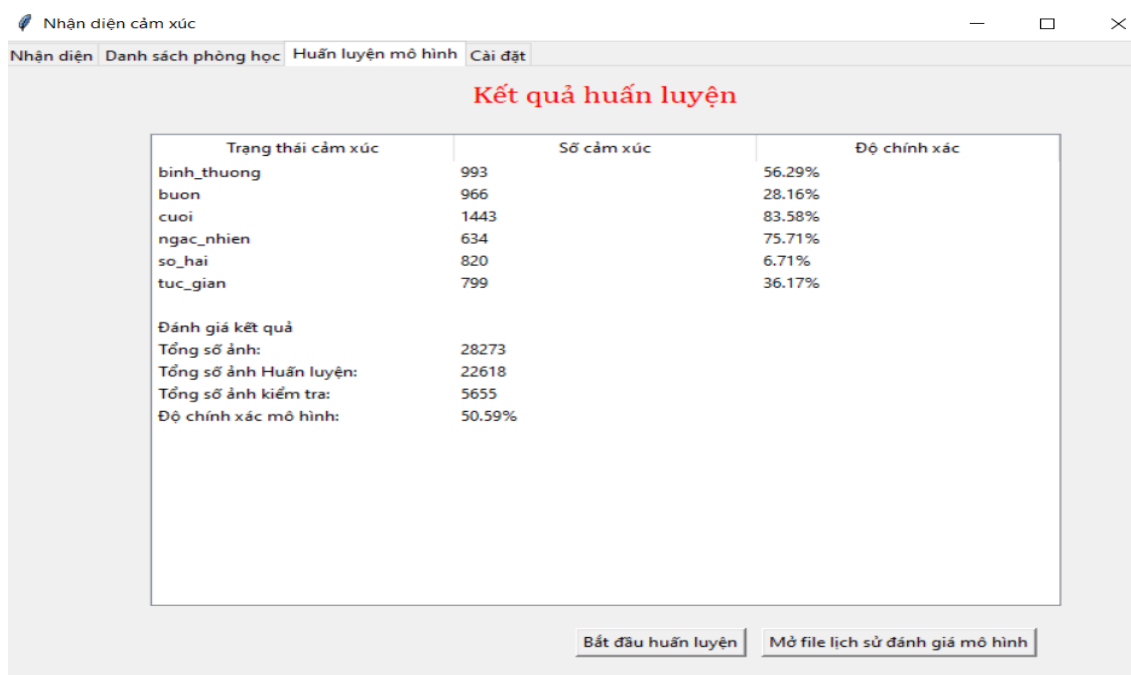
Giao diện kết quả đánh giá: trong giao diện hiển thị trạng thái cảm xúc nhận diện, số cảm xúc khuôn mặt được hiển thị trong tab nhận diện.



Giao diện danh sách phòng: Các chức năng này được sử dụng trong tab Camera trực tiếp của phần nhận diện, trong giao diện hiển thị hiển thị ID số thứ tự, tên phòng được khởi tạo, mã IP camera được kết nối. Ngoài ra chức năng thêm, sửa, xóa có nhiệm vụ thay đổi phần tên phòng và địa chỉ ip phòng và cập nhật danh sách dùng để cập nhật danh sách được thêm vào.



Giao diện huấn luyện mô hình: Hiển thị quá trình huấn luyện và các thông số kết quả của quá trình huấn luyện giúp người dùng có thể đánh giá được tỉ lệ chính xác của mô hình.



Giao diện Cài đặt: Để bật tắt lưu ảnh nhận diện và cài đặt thông số giờ và phút lưu ảnh trong quá trình nhận diện.

