MỤC LỤC

[MỞ ĐẦU 3](#_Toc154607318)

[1. Bối Cảnh Và Mục Tiêu 3](#_Toc154607319)

[2. Ý Nghĩa Của Nghiên Cứu 3](#_Toc154607320)

[3. Cấu Trúc Báo Cáo 3](#_Toc154607321)

[CHƯƠNG I. MÔ TẢ BÀI TOÁN 4](#_Toc154607322)

[1. Xác Định Bài Toán 4](#_Toc154607323)

[2. Môi Trường Nghiên Cứu 4](#_Toc154607324)

[3. Phương Pháp Dự Kiến 4](#_Toc154607325)

[CHƯƠNG II. PHƯƠNG PHÁP 5](#_Toc154607326)

[1. Landmark extractor 5](#_Toc154607327)

[2. Graph construction 6](#_Toc154607328)

[3. Mạng nơ-ron đồ thị có hướng 8](#_Toc154607329)

[CHƯƠNG III. Quá trình huấn luyện 10](#_Toc154607330)

[1. Bộ dữ liệu 10](#_Toc154607331)

[2. Quy trình tiền xử lý 10](#_Toc154607332)

[3. Mô hình đào tạo 11](#_Toc154607333)

[CHƯƠNG IV. KẾT QUẢ THỬ NGHIỆM 13](#_Toc154607334)

[1. Kết quả đánh giá 13](#_Toc154607335)

[1. Điểm Mạnh 13](#_Toc154607336)

[2. Điểm Cần Cải Thiện 14](#_Toc154607337)

[CHƯƠNG V. Hướng phát triển tiếp theo 15](#_Toc154607338)

[TÀI LIỆU THAM KHẢO 16](#_Toc154607339)

# MỞ ĐẦU

Báo cáo này đưa ra một cái nhìn chi tiết và đầy đủ về quá trình phát triển mô hình phân loại cảm xúc dựa trên đồ thị được xây dựng trên landmark sử dụng mô hình Directional Graph Neural Network (DGNN). Bài toán này không chỉ đặt ra thách thức trong việc hiểu rõ mối quan hệ phức tạp giữa các landmark trên ảnh mà còn đưa ra nhu cầu xây dựng biểu diễn đồ thị có hướng để tận dụng thông tin này.

## Bối Cảnh Và Mục Tiêu

Bài toán của chúng tôi xuất phát từ sự nhận thức về sự quan trọng của việc hiểu biểu hiện cảm xúc trong hình ảnh và video. Điều này không chỉ mang lại những đóng góp quan trọng trong lĩnh vực nhận dạng hình ảnh mà còn có ứng dụng rộng rãi trong lĩnh vực như phân tích thị trường, giao tiếp trực tuyến, và nhiều lĩnh vực khác.

Mục tiêu của chúng tôi là phát triển một mô hình có khả năng hiệu quả phân loại cảm xúc từ ảnh, đặc biệt là khi thông tin đồ thị có hướng từ key-point được tích hợp một cách chặt chẽ. Điều này không chỉ là một thách thức khoa học mà còn là nền tảng cho việc ứng dụng trong các ứng dụng thực tế.

## Ý Nghĩa Của Nghiên Cứu

Việc hiểu biểu hiện cảm xúc trong hình ảnh không chỉ hỗ trợ trong việc tối ưu hóa trải nghiệm người dùng mà còn mở ra nhiều cơ hội mới trong lĩnh vực trí tuệ nhân tạo và máy học. Bằng cách tận dụng thông tin đồ thị và key-point, chúng tôi hy vọng mở ra những cánh cửa mới trong việc hiểu biết về tương tác người-máy.

## Cấu Trúc Báo Cáo

Báo cáo sẽ bao gồm các phần chính như Mô Tả Về Bài Toán, Phương Pháp, Quá Trình Huấn Luyện, Kết Quả Thử Nghiệm, và Hướng Phát Triển Tiếp Theo. Mỗi phần sẽ chi tiết về quá trình phát triển, các quyết định quan trọng, và những bài học từ quá trình nghiên cứu.

Trong quá trình thực hiện, chúng tôi sử dụng thuật ngữ "điểm mốc" và "landmarks" một cách linh hoạt để mô tả các điểm đặc trưng trên khuôn mặt. Điều này giúp chúng tôi truyền đạt thông tin một cách chính xác và linh hoạt tùy thuộc vào ngữ cảnh cụ thể của mô hình.

# CHƯƠNG I. MÔ TẢ BÀI TOÁN

## Xác Định Bài Toán

Bài toán chúng tôi nghiên cứu nhằm phân loại cảm xúc từ hình ảnh khuôn mặt, sử dụng thông tin từ các key-point và xây dựng đồ thị có hướng để mô hình hóa mối quan hệ giữa chúng. Thách thức đặt ra không chỉ là hiểu rõ mối quan hệ phức tạp giữa các điểm trên khuôn mặt mà còn đòi hỏi sự hiểu biết vững về cách biểu hiện cảm xúc được thể hiện.

## Môi Trường Nghiên Cứu

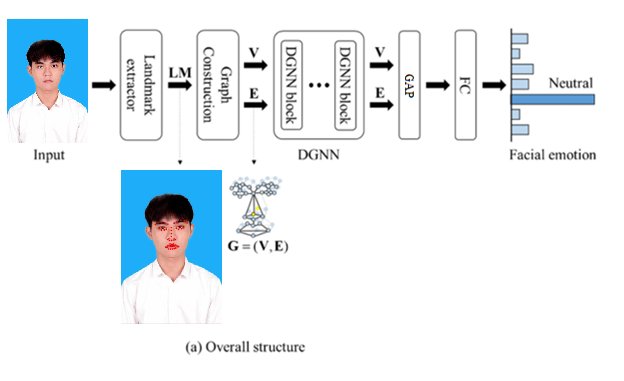
Chúng tôi đã sử dụng một bộ dữ liệu đa dạng chứa ảnh khuôn mặt với các nhãn cảm xúc tương ứng, đảm bảo rằng nó bao gồm sự đa dạng về biểu hiện cảm xúc. Môi trường nghiên cứu của chúng tôi được xây dựng trên nền tảng của các công cụ và thư viện phổ biến như PyTorch, OpenCV và NetworkX, nhằm hỗ trợ quá trình xây dựng và huấn luyện mô hình. Sự linh hoạt và tiện ích của các công cụ này giúp chúng tôi dễ dàng thực hiện các thử nghiệm và nghiên cứu một cách hiệu quả trong quá trình tối ưu hóa mô hình nhận diện cảm xúc trên ảnh khuôn mặt.

## **Phương Pháp Dự Kiến**

Chúng tôi sử dụng thư viện dlib để phát hiện và trích xuất key-point trên khuôn mặt. Dữ liệu từ key-point được sử dụng để xây dựng đồ thị có hướng, mô hình hóa mối quan hệ giữa các điểm trên khuôn mặt và biểu hiện cảm xúc. Mô hình chính của chúng tôi sử dụng Directional Graph Neural Network (DGNN) để tận dụng thông tin đồ thị này.

# CHƯƠNG II. PHƯƠNG PHÁP

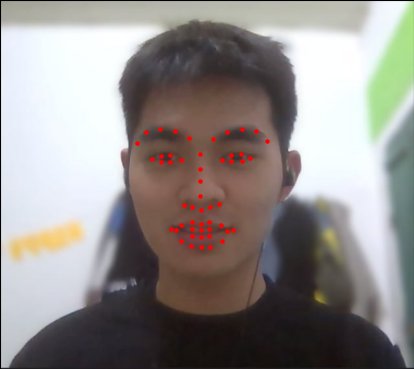
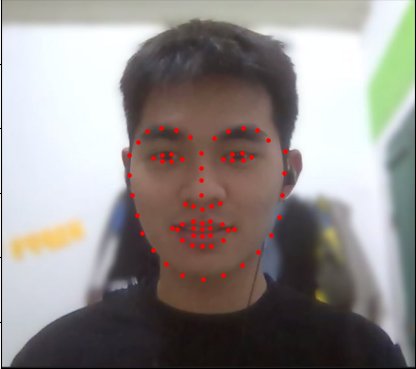
Phần này trình bày mạng nơ-ron đồ thị có hướng mà chúng tôi đề xuất để tận dụng tính năng của các mốc. Kiến trúc tổng thể của phương pháp được minh họa trong hình dưới đây. Đầu tiên, chúng tôi sử dụng công cụ trích xuất mốc dựa trên deep learning thông thường để thu được các mốc từ mỗi bức ảnh (Phần II.1). Tiếp theo, chúng tôi xây dựng cấu trúc biểu đồ dựa trên vị trí của các mốc. Để nắm bắt thông tin giữa các nút ở xa, chúng tôi đặt nút chính (trung tâm của mũi) để tạo ra đồ thị cấu trúc (Phần II.2). Cuối cùng, chúng tôi áp dụng mạng nơ-ron đồ thị có hướng (DGNN) để phân tích cấu trúc đồ thị.



Hình 1: Tổng quan về phương pháp đề xuất

## Landmark extractor

Chúng tôi khởi đầu mô tả bằng việc giới thiệu công cụ trích xuất mốc, một phần quan trọng trong quá trình nhận diện cảm xúc trên khuôn mặt. Để thực hiện nhiệm vụ này, chúng tôi đã chọn sử dụng thư viện dlib để nắm bắt các đặc điểm quan trọng trên khuôn mặt. Chi tiết hơn, chúng tôi đã trích xuất 68 điểm mốc trên khuôn mặt con người, bao gồm mũi, mắt, lông mày và miệng. Tuy nhiên, với mục tiêu cải thiện hiệu suất nhận diện cảm xúc, chúng tôi quyết định sử dụng chỉ 51 điểm mốc từ mỗi bức ảnh, loại bỏ các điểm mốc ở khu vực bên ngoài mà chúng tôi nhận thấy có thể ảnh hưởng tiêu cực đến quá trình nhận dạng.



Hình 2: Phần bên trái hiển thị 68 điểm mốc, trong khi phần bên phải chỉ bao gồm 51 điểm mốc.

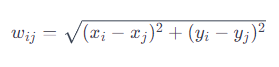
## Graph construction

Mỗi điểm mốc trên khuôn mặt thể hiện mối quan hệ chặt chẽ với các điểm mốc khác, được đặt ở vị trí tương tự. Điều này dẫn đến việc tập hợp các vị trí mốc có khả năng tạo ra một cấu trúc đồ thị có khả năng truyền tải thông tin về mối quan hệ cấu trúc, được định nghĩa là:

G = (V, E)

Trước tiên, chúng tôi gán số thứ tự từ 1 đến 51 cho mỗi vị trí mốc, xác định chúng là các đỉnh trong đồ thị. Điều này giúp đơn giản hóa mô hình và giảm tải tính toán bằng cách sử dụng số thứ tự thay vì toạ độ của mỗi đỉnh.

Tiếp theo, chúng tôi tiến hành xây dựng các cạnh E bằng phương pháp Delaunay. Phương pháp Delaunay giúp tạo ra các lưới tam giác giữa tất cả các điểm mốc và đây là một phương tiện hữu ích trong việc phân tích cảm xúc trên khuôn mặt. Tuy nhiên, cấu trúc lưới chỉ cung cấp thông tin về việc cạnh có được kết nối hay không. Do đó, chúng tôi đã gán trọng số cho các cạnh bằng cách tính toán khoảng cách Euclid giữa hai đỉnh trên khuôn mặt, theo công thức dưới đây:



Trong đó:

* *wij*​ là trọng số của cạnh giữa đỉnh *i* và *j*.
* *(xi*​,*yi*​) và (*xj*​,*yj*​) là tọa độ của hai đỉnh trên khuôn mặt.

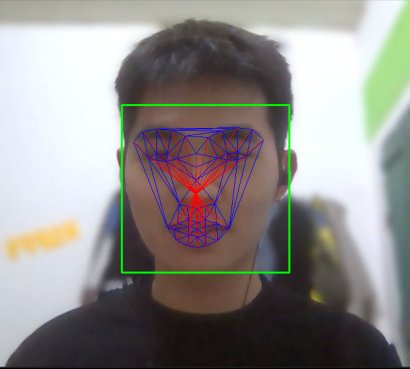
Quyết định sử dụng trọng số này giúp chúng tôi tận dụng thông tin về khoảng cách giữa các điểm mốc, làm giàu thông tin cấu trúc của đồ thị.

Xây dựng các cạnh bằng phương pháp Delaunay tạo ra một cấu trúc đồ thị cơ bản, nhưng trong bài toán nhận diện cảm xúc từ khuôn mặt, Delaunay không đảm bảo đủ hiệu suất. Việc thêm master node vào đồ thị giúp cải thiện thông tin toàn cảnh, quản lý hiệu quả kích thước đầu vào và tăng cường mối quan hệ giữa các vị trí trên khuôn mặt. Điều này giúp nâng cao khả năng biểu diễn và hiệu suất của mô hình trong việc nhận diện cảm xúc từ dữ liệu khuôn mặt.

Theo định nghĩa, nút chính liên kết với tất cả các đỉnh khác trong cấu trúc đồ thị. Do đó, thông tin được truyền bá một cách hiệu quả qua kết nối toàn bộ. Trong phương pháp của chúng tôi, chúng tôi chọn điểm trung tâm của mũi làm nút chính.

Chúng tôi ưu tiên sử dụng một chiều của đồ thị vì tin rằng nó đủ để đáp ứng yêu cầu của bài toán và giúp đơn giản hóa cấu trúc. Quyết định này cũng nhằm tối ưu hóa hiệu suất tính toán và giữ cho mô hình trở nên dễ quản lý hơn.

Kết quả của đồ thị G là một cấu trúc biểu đồ mô tả đầy đủ thông tin về biểu hiện cảm xúc trên khuôn mặt được mô tả như hình dưới đây.



Hình 3: Mô tả đồ thị của các cạnh và các đỉnh trên khuôn mặt người. Nút mũi được xác định là nút chính, và sau đó, tất cả các nút khác đều được kết nối với nút chính. Trong hình này, các cạnh màu đỏ được tạo ra dựa trên nút chính, trong khi các cạnh màu xanh được xây dựng theo phương pháp Delaunay.

## Mạng nơ-ron đồ thị có hướng

Trong bối cảnh lĩnh vực học máy, Mô hình Đồ thị có Hướng (DGNN) đặc biệt được thiết kế để giải quyết các vấn đề liên quan đến dữ liệu có cấu trúc đồ thị. Điểm độc đáo của DGNN là khả năng linh hoạt trong việc tận dụng thông tin từ cả đỉnh và cạnh của đồ thị để thực hiện hiệu quả các nhiệm vụ học máy. Điều này cho phép mô hình tự động học và hiểu các mối quan hệ phức tạp giữa các thành phần trong dữ liệu đồ thị.

Trong ngữ cảnh cụ thể của phân loại cảm xúc từ các landmarks trên khuôn mặt, DGNN có thể được áp dụng để tận dụng cấu trúc đồ thị có hướng. Các landmarks trên khuôn mặt được biểu diễn bằng các đỉnh, trong khi mối quan hệ giữa chúng được thể hiện qua các cạnh có hướng. Mô hình DGNN giúp tổ chức và hiểu rõ các mối quan hệ phức tạp này, đồng thời cải thiện khả năng phân loại cảm xúc từ dữ liệu hình ảnh.

Sự linh hoạt của DGNN trong việc xử lý dữ liệu đồ thị làm cho nó trở thành một công cụ mạnh mẽ trong nhiều ứng dụng, đặc biệt là khi thông tin không chỉ xuất hiện ở cấp độ điểm mà còn ở cấp độ mối quan hệ giữa chúng. Điều này giúp DGNN tự động học được các đặc trưng phức tạp và đồng thời cải thiện khả năng biểu diễn và nhận dạng các biểu hiện cảm xúc trên khuôn mặt người.

Vì vậy, chúng tôi tin rằng DGNN là một lựa chọn đáng tin cậy cho bài toán phân loại cảm xúc từ các landmarks trên khuôn mặt.

# CHƯƠNG III. Quá trình huấn luyện

Phần này sẽ tập trung trình bày về quá trình chọn lọc bộ dữ liệu, tiền xử lý dữ liệu và cấu trúc cũng như quá trình huấn luyện mô hình của chúng tôi.

## Bộ dữ liệu

Bộ dữ liệu CK+ (Cohn-Kanade) đóng vai trò quan trọng trong lĩnh vực nhận diện cảm xúc, tập trung vào nghiên cứu biểu hiện cảm xúc trên khuôn mặt con người. Bộ dữ liệu này chứa các bức ảnh ghi lại biểu hiện khuôn mặt của các tham gia trong nhiều tình huống khác nhau.

Dữ liệu CK+ bao gồm tổng cộng 680 bức ảnh trong tập huấn luyện (70%), 192 bức ảnh trong tập kiểm định (20%), và 99 bức ảnh trong tập kiểm tra (10%). Bộ dữ liệu tập trung chủ yếu vào 7 loại cảm xúc cơ bản, và mỗi cảm xúc được gán nhãn dựa trên tiếng Việt như sau: Tức giận, Khinh thường, Ghê tởm, Sợ hãi, Hạnh phúc, Buồn bã, Ngạc nhiên.

Đây là những biểu cảm cơ bản được sử dụng để nghiên cứu và phân loại cảm xúc trong bối cảnh khác nhau.



Hình 4: Một vài bức ảnh trong bộ dữ liệu CK+

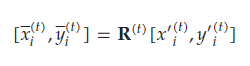
## Quy trình tiền xử lý

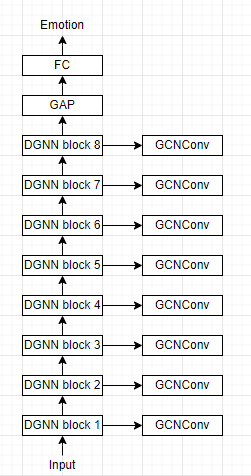
Nhận thức về thách thức của việc không đa dạng về nguồn sáng trong dữ liệu, chúng tôi đã thực hiện các cải tiến bằng cách thêm yếu tố nhiễu ngẫu nhiên vào tập dữ liệu. Điều này giúp mô phỏng và làm phong phú điều kiện ánh sáng trong các hình ảnh khuôn mặt. Việc thực hiện quay ngẫu nhiên và lật ảnh ngẫu nhiên giúp mô hình học nhận diện cảm xúc trên độ nghiêng và góc nhìn đa dạng.

Chúng tôi đã tìm được một phương pháp tăng cường dữ liệu để mở rộng kích thước của tập dữ liệu. Ban đầu, chúng tôi thêm nhiễu Gaussian vào vị trí các điểm mốc theo công thức sau:



Ở đây 𝜎 là độ lệch chuẩn của nhiễu Gaussian. Tiếp theo, chúng tôi áp dụng xoay ngẫu nhiên cho mỗi hình ảnh :



Ở đây𝐑(𝑡) là ma trận xoay 2 × 2. Phạm vi của góc quay là [−𝜋/10,𝜋/10]. Ngoài ra, chúng tôi thực hiện lật theo ảnh theo trục Oy đối với mỗi bức ảnh. Chúng tôi đã tạo tập dữ liệu gốc, tập dữ liệu bị nhiễu ngẫu nhiên (ba lần), tập dữ liệu được xoay ngẫu nhiên (ba lần) và các phiên bản lật của chúng. Kết quả là chúng tôi đã mở rộng bộ dữ liệu gốc lên 14 lần.

## Mô hình đào tạo

Hình 5: Tổng quan về mô hình DGNN được sử dụng

Đối với tất cả các thử nghiệm, chúng tôi đã áp dụng một kiến trúc mạng chung, được trình bày như hình dưới đây :

Một Directional Graph Neural Network (DGNN) bao gồm 8 khối tích chập đồ thị có hướng, thực hiện việc tổng hợp thông tin về các đỉnh và cạnh. Mỗi khối DGNN bao gồm một lớp tích chập đồ thị (Graph Convolutional Layer - GCN) để xử lý thông tin trên đồ thị. Trong mỗi lớp tích chập đồ thị, đầu vào là đặc trưng của các đỉnh trong đồ thị, và lớp này sử dụng cạnh và trọng số cạnh để cập nhật và học biểu diễn mới cho từng đỉnh.

Thông qua việc kết hợp các lớp tích chập đồ thị, mỗi khối DGNN có thể hiệu quả học biểu diễn phức tạp và chứa đựng thông tin từ toàn bộ đồ thị. Các khối được xây dựng trên nhau để tăng sự phức tạp của mô hình và nắm bắt những mối quan hệ sâu hơn và phức tạp hơn giữa các đỉnh trong đồ thị.

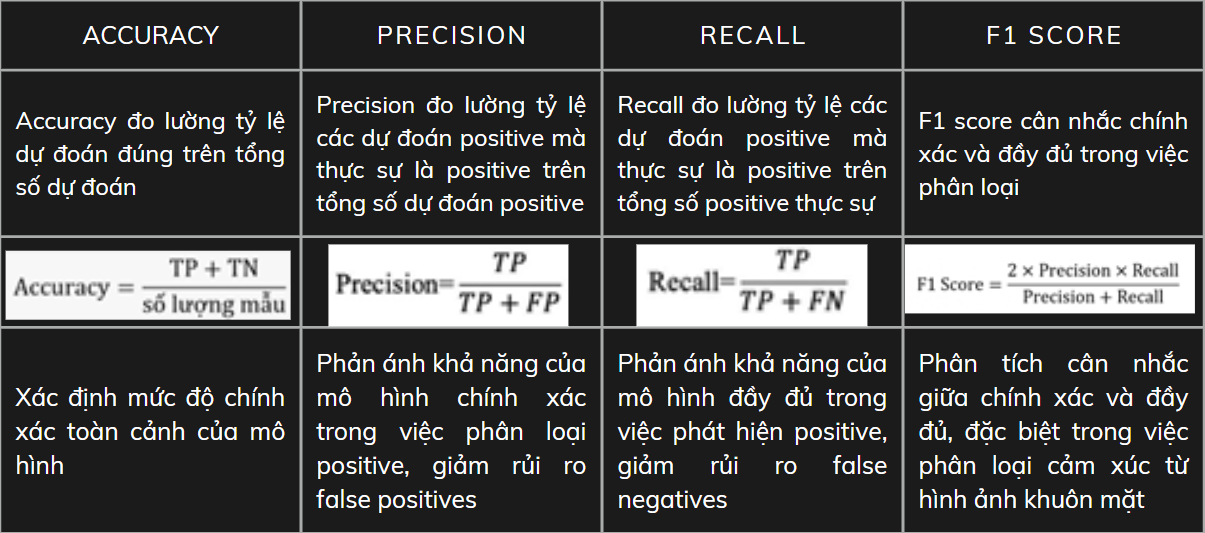
Sử dụng global mean pooling để tạo biểu diễn tổng hợp của đồ thị. Và 2 lớp Fully connected được áp dụng để phân loại dựa trên biểu diễn được học, với việc sử dụng dropout để tránh overfitting.

Trong trường hợp bộ dữ liệu biểu hiện khuôn mặt Cohn-Kanade (CK+), tốc độ học và kích thước lô được đặt tương ứng là 0,05 và 64. Quá trình đào tạo được thực hiện trong 43 epoch để đạt được hiệu suất tốt nhất. Chúng tôi đã sử dụng hàm mất mát nn.CrossEntropyLoss để đánh giá hiệu suất cho bài toán phân loại nhiều lớp.

# CHƯƠNG IV. KẾT QUẢ THỬ NGHIỆM

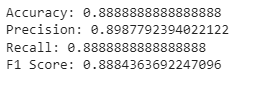
## Kết quả đánh giá

Chúng tôi sử dụng accuracy, precision, recall và F1 score làm các thước đo chính để đánh giá hiệu suất tổng thể của mô hình trên tập kiểm tra. Thông tin chi tiết về các phép đánh giá được trình bày chi tiết trong bảng dưới đây:



Hình 6: Thông tin chi tiết về các phép đánh giá

Kết quả khi đánh giá mô hình trên tập kiểm thử:

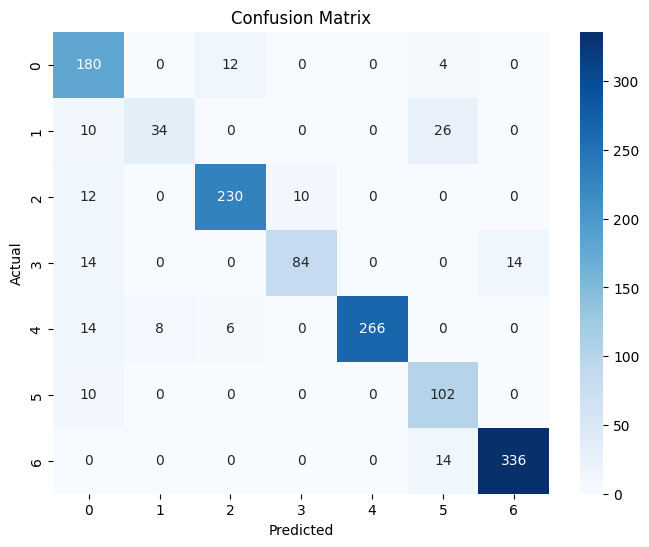


Hình 7: Kết quả mô hình sử dụng các phép đánh giá trên tập kiểm thử

## Điểm Mạnh

* Khả Năng Chính Xác và Đầy Đủ: Mô hình có khả năng phân loại cảm xúc một cách chính xác và đầy đủ, tăng cường sự tin cậy khi triển khai trong các ứng dụng thực tế.
* Xử Lý Keypoint Hiệu Quả: Việc sử dụng keypoint để trích xuất đặc trưng đã được chứng minh là một phương pháp hiệu quả, đặc biệt là trong bài toán nhận diện cảm xúc từ hình ảnh khuôn mặt.

## Điểm Cần Cải Thiện



Hình 8: Confused matrix

* Chưa phân biệt được chính xác “khinh thường” và “buồn bã”: Ma trận nhầm lẫn trong hình trên cho thấy DGNN thường không phân biệt được sự khinh miệt với nỗi buồn. Điều này là do các dấu hiệu như hình dạng lông mày và môi liên quan đến sự khinh miệt và cảm xúc buồn bã đều giống nhau
* Thời Gian Tính Toán: Mô hình đòi hỏi một lượng thời gian đáng kể để xây dựng đồ thị từ các key-point và sau đó thực hiện quá trình dự đoán. Điều này có thể được giải quyết thông qua việc tối ưu hóa hiệu suất tính toán hoặc sử dụng các kỹ thuật và thư viện được tối ưu hóa để tăng tốc quá trình này, đặc biệt là khi áp dụng mô hình trên các tập dữ liệu lớn.

# 

# CHƯƠNG V. Hướng phát triển tiếp theo

Tinh Chỉnh và Mở Rộng Mô Hình: Tiếp tục tinh chỉnh siêu tham số và mô hình để đạt được hiệu suất tốt hơn, vượt qua ngưỡng 90%. Mở rộng khả năng xử lý dữ liệu và trích xuất đặc trưng cho các trường hợp đặc biệt như ánh sáng yếu hay biểu cảm khuôn mặt phức tạp hơn.

Dự Đoán Cảm Xúc Dựa Trên Video: Nghiên cứu và phát triển mô hình để thực hiện dự đoán cảm xúc dựa trên dữ liệu video thay vì chỉ từ các ảnh tĩnh. Điều này sẽ mở ra ứng dụng rộng lớn trong việc theo dõi cảm xúc qua thời gian.

Nghiên Cứu Về Tốc Độ Tính Toán: Tối ưu hóa quy trình tính toán để giảm thời gian xử lý, đặc biệt là trong quá trình xây dựng đồ thị từ key-point và dự đoán. Điều này có thể bao gồm sử dụng kỹ thuật tối ưu hóa mã hóa GPU hoặc phát triển các phương pháp tính toán hiệu quả hơn.

# 

# TÀI LIỆU THAM KHẢO

1. Jung, Y., & Lee, S. (2015). Joint Fine-Tuning in Deep Neural Networks for Facial Expression Recognition. In Proceedings of the IEEE International Conference on Computer Vision (ICCV) (pp. 2983-2991). Link: <https://www.cv-foundation.org/openaccess/content_iccv_2015/html/Jung_Joint_Fine-Tuning_in_ICCV_2015_paper.html>
2. VIBLO. (n.d.). "Computer Vision: Phát Hiện Gương Mặt và Nhận Diện Nụ Cười Đơn Giản Cho Người Mới Bắt Đầu.". Link: <https://viblo.asia/p/computer-vision-phat-hien-guong-mat-va-nhan-dien-nu-cuoi-don-gian-cho-nguoi-moi-bat-dau-vyDZOwbGZwj>
3. Gurjar, M. G., & Patil, M. M. (2018). Emotion Recognition using Spatiotemporal Features from Facial Expression Landmarks. Link: https://www.researchgate.net/publication/328676113\_Emotion\_Recognition\_using\_Spatiotemporal\_Features\_from\_Facial\_Expression\_Landmarks
4. H. Nguyen, T. Nguyen, L. D. Nguyen, V. T. Le, & C. N. Do. (2019). "Emotion Recognition using Spatiotemporal Features from Facial Expression Landmarks." Electronics, 9(5), 764. Link: https://www.mdpi.com/2079-9292/9/5/764#B11-electronics-09-00764
5. Neptune.ai. (n.d.). "Graph Neural Network and Some of GNN Applications." Link: https://neptune.ai/blog/graph-neural-network-and-some-of-gnn-applications