**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ ĐÔNG Á**

**KHOA CÔNG NGHỆ THÔNG TIN**



**BÀI TẬP LỚN**

**HỌC PHẦN: XỬ LÝ ẢNH VÀ THỊ GIÁC MÁY TÍNH**

**Đề tài số 18: Phát triển hệ thống nhận diện cảm xúc từ hình ảnh khuôn mặt sử dụng mạng GAN**

**Giảng viên hướng dẫn: Lương Thị Hồng Lan**

|  |  |  |  |
| --- | --- | --- | --- |
| **TT** | **Mã sinh viên** | **Sinh viên thực hiện** | **Lớp hành chính** |
| **1** | 20211393 | Nguyễn Ngọc Tuyền | DCCNTT12.10.6 |
| **2** | 20211553 | Trương Khánh Linh | DCCNTT12.10.6 |
| **3** | 20211351 | Nguyễn Việt Tùng | DCCNTT12.10.6 |
| **4** | 20211651 | Nguyễn Đại Dương | DCCNTT12.10.6 |
| **5** | 20211732 | Nguyễn Trung Đức | DCCNTT12.10.6 |

**Bắc Ninh, năm 2024**

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ ĐÔNG Á**

**KHOA CÔNG NGHỆ THÔNG TIN**

**BÀI TẬP LỚN**

**HỌC PHẦN: XỬ LÝ ẢNH VÀ THỊ GIÁC MÁY TÍNH**

**Đề tài số 18: Phát triển hệ thống nhận diện cảm xúc từ hình ảnh khuôn mặt sử dụng mạng GAN**

**Giảng viên hướng dẫn: Lương Thị Hồng Lan**

|  |  |  |  |
| --- | --- | --- | --- |
| **TT** | **Mã sinh viên** | **Sinh viên thực hiện** | **Lớp hành chính** |
| **1** | 20211393 | Nguyễn Ngọc Tuyền | DCCNTT12.10.6 |
| **2** | 20211553 | Trương Khánh Linh | DCCNTT12.10.6 |
| **3** | 20211351 | Nguyễn Việt Tùng | DCCNTT12.10.6 |
| **4** | 20211651 | Nguyễn Đại Dương | DCCNTT12.10.6 |
| **5** | 20211732 | Nguyễn Trung Đức | DCCNTT12.10.6 |

|  |  |
| --- | --- |
|  |  |

**Bắc Ninh, năm 2024**

|  |  |
| --- | --- |
| TRƯỜNG ĐẠI HỌC CÔNG NGHỆ ĐÔNG Á  **KHOA CÔNG NGHỆ THÔNG TIN** | **KỲ THI KẾT THÚC HỌC PHẦN**  **HỌC KỲ 1, NĂM HỌC 2024** – **2025** |

|  |  |
| --- | --- |
| **PHIẾU CHẤM THI BÀI TẬP LỚN KẾT THÚC HỌC PHẦN**  **Mã đề thi: 18**  **Tên học phần: Xử lý ảnh và thị giác máy tính**  **Lớp Tín chỉ: XATGMT.03.K12.06.LH.C04.1\_LT** | |
| **Cán bộ chấm thi 1**  *(Ký và ghi rõ họ tên)*  **Lương Thị Hồng Lan** | **Cán bộ chấm thi 2**  *(Ký và ghi rõ họ tên)* |

| **TT** | **TIÊU CHÍ** | **THANG ĐIỂM** | **Nguyễn Ngọc Tuyền** | **Trương Khánh Linh** | **Nguyễn Việt Tùng** | **Nguyễn Đại Dương** | **Nguyễn Trung Đức** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 20211393 | 20211553 | 20211351 | 20211651 | 20211732 |
| **1** | **Nội dung báo cáo trên Word đầy đủ** | **3.5** |  |  |  |  |  |
| 1.1 | Có bố cục rõ ràng (mục lục, phần mở đầu, nội dung chính, kết luận). | 0,5 |  |  |  |  |  |
| 1.2 | Nội dung phân tích rõ ràng, logic. | 0,5 |  |  |  |  |  |
| 1.3 | Có dẫn chứng, số liệu minh họa đầy đủ. | 0,5 |  |  |  |  |  |
| 1.4 | Ngôn ngữ và trình bày chuẩn, không lỗi chính tả. | 0,5 |  |  |  |  |  |
| 1.5 | Có trích dẫn tài liệu tham khảo đúng quy cách. | 0,5 |  |  |  |  |  |
| 1.6 | Được trình bày chuyên nghiệp (canh lề, font chữ, khoảng cách dòng hợp lý). | 0,5 |  |  |  |  |  |
| 1.7 | Tài liệu đầy đủ, bám sát yêu cầu của đề bài. | 0,5 |  |  |  |  |  |
| **2** | **Nội dung thuyết trình đầy đủ** | **1.0** |  |  |  |  |  |
| 2.1 | Trình bày tự tin, phát âm rõ ràng, mạch lạc. | 0,5 |  |  |  |  |  |
| 2.2 | Nội dung thuyết trình đúng trọng tâm, không lan man. | 0,5 |  |  |  |  |  |
| **3** | **Slides báo cáo đầy đủ nội dung + Hỏi đáp** | **3.0** |  |  |  |  |  |
| 3.1 | Slides có bố cục rõ ràng (mở đầu, nội dung, kết luận). | 0,5 |  |  |  |  |  |
| 3.2 | Thiết kế slides đẹp, chuyên nghiệp (màu sắc, hình ảnh minh họa). | 0,5 |  |  |  |  |  |
| 3.3 | Nội dung trên slides ngắn gọn, dễ hiểu, súc tích. | 0,5 |  |  |  |  |  |
| 3.4 | Nội dung slides phù hợp với nội dung báo cáo. | 0,5 |  |  |  |  |  |
| 3.5 | Trả lời câu hỏi đầy đủ, chính xác. | 0,5 |  |  |  |  |  |
| 3.6 | Trả lời câu hỏi tự tin, thuyết phục. | 0,5 |  |  |  |  |  |
| **4** | **Code đầy đủ** | **2.5** |  |  |  |  |  |
| 1.1 | Code được trình bày rõ ràng, có chú thích đầy đủ. | 0,5 |  |  |  |  |  |
| 1.2 | Code chạy đúng, không lỗi. | 0,5 |  |  |  |  |  |
| 1.3 | Code tối ưu, không dư thừa. | 0,5 |  |  |  |  |  |
| 1.4 | Đáp ứng đầy đủ các yêu cầu chức năng theo đề bài. | 0,5 |  |  |  |  |  |
| 1.5 | Có tính sáng tạo hoặc cải thiện so với yêu cầu. | 0,5 |  |  |  |  |  |
| **TỔNG ĐIỂM BẰNG SỐ:** | | **10** |  |  |  |  |  |
| **TỔNG ĐIỂM BẰNG CHỮ:** | | *Mười tròn* |  |  |  |  |  |

**Mục Lục**

[**LỜI NÓI ĐẦU** 8](#_Toc184652715)

[**DANH MỤC BẢNG BIỂU VÀ SƠ ĐỒ** 9](#_Toc184652716)

[**PHẦN MỞ ĐẦU** 10](#_Toc184652717)

[**A. Giới thiệu tổng quan đề tài** 10](#_Toc184652718)

[**1. Lý do chọn đề tài** 10](#_Toc184652719)

[**2. Mục đích nghiên cứu** 10](#_Toc184652720)

[**3. Đối tượng và phạm vi nghiên cứu** 10](#_Toc184652721)

[**4. Phương pháp nghiên cứu** 11](#_Toc184652722)

[**B. Danh sách thành viên và phần công việc** 11](#_Toc184652723)

[**CHƯƠNG I: CƠ SỞ LÝ THUYẾT** 12](#_Toc184652724)

[**1.1. Giới thiệu chung về bài toán nhận diện cảm xúc** 12](#_Toc184652725)

[**1.1.1. Tầm quan trọng của nhận diện cảm xúc [2]** 12](#_Toc184652726)

[**1.1.2. Ứng dụng của hệ thống nhận diện cảm xúc [2]** 12](#_Toc184652727)

[**1.2. Các phương pháp nhận diện cảm xúc truyền thống [3]** 14](#_Toc184652728)

[**1.2.1. Dựa trên đặc trưng hình ảnh thủ công** 14](#_Toc184652729)

[**1.2.2. Dựa trên mô hình học máy truyền thống** 15](#_Toc184652730)

[**1.3. Giới thiệu mạng GAN (Generative Adversarial Network)** 17](#_Toc184652731)

[**1.3.1. Kiến trúc của GAN [1]** 17](#_Toc184652732)

[**1.3.2. Cơ chế hoạt động của GAN [1]** 19](#_Toc184652733)

[**1.3.3. Các biến thể phổ biến của GAN [4]** 21](#_Toc184652734)

[**CHƯƠNG II: XÂY DỰNG HỆ THỐNG NHẬN DIỆN CẢM XÚC** 24](#_Toc184652735)

[**2.1. Định nghĩa bài toán** 24](#_Toc184652736)

[**2.1.1. Mô tả bài toán** 24](#_Toc184652737)

[**2.1.2. Yêu cầu đầu vào và đầu ra** 29](#_Toc184652738)

[**2.2. Ứng dụng GAN trong nhận diện cảm xúc [5]** 30](#_Toc184652739)

[**2.2.1. Tổng hợp và khuếch đại dữ liệu** 30](#_Toc184652740)

[**2.2.2. Tăng cường độ chính xác mô hình nhận diện cảm xúc** 31](#_Toc184652741)

[**2.3. Xây dựng hệ thống** 32](#_Toc184652742)

[**2.3.1. Thu thập và xử lý dữ liệu** 32](#_Toc184652743)

[**2.3.2. Phân tích kiến trúc GAN được sử dụng** 33](#_Toc184652744)

[**2.3.3. Triển khai hệ thống** 38](#_Toc184652745)

[**2.3.4. Tích hợp và đánh giá** 42](#_Toc184652746)

[**CHƯƠNG III: KẾT QUẢ THỰC NGHIỆM** 46](#_Toc184652747)

[**3.1. Dữ liệu** 46](#_Toc184652748)

[**3.1.1. Mô tả dữ liệu** 46](#_Toc184652749)

[**3.1.2. Tiền xử lý dữ liệu** 48](#_Toc184652750)

[**3.1.3. Bổ sung dữ liệu với GAN** 49](#_Toc184652751)

[**3.2. Độ đo đánh giá** 50](#_Toc184652752)

[**3.2.1. Các chỉ số đánh giá** 50](#_Toc184652753)

[**3.2.2. Đánh giá hiệu suất** 51](#_Toc184652754)

[**3.3. Kết quả thực nghiệm** 51](#_Toc184652755)

[**3.3.1. Hiệu suất mô hình** 51](#_Toc184652756)

[**3.3.2. Ảnh chụp màn hình kết quả** 52](#_Toc184652757)

[**3.3.3. Thảo luận kết quả** 54](#_Toc184652758)

[**KẾT LUẬN** 55](#_Toc184652759)

[**4.1 Kết quả đạt được** 55](#_Toc184652760)

[**4.2 Hạn chế** 55](#_Toc184652761)

[**4.3 Hướng phát triển** 55](#_Toc184652762)

[**TÀI LIỆU THAM KHẢO** 56](#_Toc184652763)

# **LỜI NÓI ĐẦU**

Trong thời đại công nghệ số ngày nay, nhận diện cảm xúc từ hình ảnh khuôn mặt đã trở thành một lĩnh vực nghiên cứu quan trọng, với nhiều ứng dụng thực tiễn trong đời sống như giáo dục, y tế, giao tiếp người-máy và các hệ thống hỗ trợ tâm lý. Khả năng hiểu và phản hồi cảm xúc của con người không chỉ giúp cải thiện sự tương tác giữa người và máy mà còn mở ra tiềm năng trong việc nâng cao chất lượng cuộc sống.

Dự án "Phát triển hệ thống nhận diện cảm xúc từ hình ảnh khuôn mặt sử dụng mạng GAN" tập trung vào việc áp dụng mạng đối kháng tạo (GAN) để giải quyết bài toán nhận diện cảm xúc với độ chính xác cao. Mạng GAN, với khả năng tổng hợp và khuếch đại dữ liệu vượt trội, đóng vai trò quan trọng trong việc cải thiện hiệu suất của các hệ thống nhận diện cảm xúc hiện nay.

Báo cáo này được thực hiện với sự hướng dẫn tận tình từ Giảng viên Lương Thị Hồng Lan, cùng sự nỗ lực của nhóm nghiên cứu nhằm xây dựng một hệ thống hiệu quả, đáp ứng yêu cầu thực tiễn. Nội dung báo cáo được tổ chức thành ba chương: chương đầu tiên trình bày các cơ sở lý thuyết liên quan; chương hai mô tả chi tiết quá trình xây dựng hệ thống; và chương ba trình bày các kết quả thực nghiệm.

Chúng tôi xin gửi lời cảm ơn chân thành đến giảng viên hướng dẫn và những người đã hỗ trợ, đóng góp ý kiến trong suốt quá trình thực hiện đề tài. Hy vọng rằng báo cáo này sẽ cung cấp những giá trị hữu ích và mở ra những hướng nghiên cứu tiếp theo trong lĩnh vực nhận diện cảm xúc từ hình ảnh khuôn mặt.

Em xin chân thành cảm ơn.

Hà Nội, ngày 12 tháng 6 năm 2024

# **DANH MỤC BẢNG BIỂU VÀ SƠ ĐỒ**

|  |  |  |
| --- | --- | --- |
| Số hiệu | Tên | Trang |
| 1 |  |  |
| 2 |  |  |
| 3 |  |  |
| 4 |  |  |
| 5 |  |  |
| 6 |  |  |
| 7 |  |  |
| 8 |  |  |
| 9 |  |  |
| 10 |  |  |
| 11 |  |  |
| 12 |  |  |
| 13 |  |  |
| 14 |  |  |
| 15 |  |  |
| 16 |  |  |
| 17 |  |  |
| 18 |  |  |
| 19 |  |  |
| 20 |  |  |
| 21 |  |  |
| 22 |  |  |
| 23 |  |  |
| 24 |  |  |
| 25 |  |  |
| 26 |  |  |
| 27 |  |  |
| 28 |  |  |
| 29 |  |  |
| 30 |  |  |
| 31 |  |  |
| 32 |  |  |
| 33 |  |  |
| 34 |  |  |
| 35 |  |  |
| 36 |  |  |

# **PHẦN MỞ ĐẦU**

## **A. Giới thiệu tổng quan đề tài**

### **1. Lý do chọn đề tài**

Trong thời đại công nghệ 4.0, việc ứng dụng trí tuệ nhân tạo (AI) để nâng cao khả năng tương tác giữa con người và máy móc ngày càng được chú trọng. Nhận diện cảm xúc từ khuôn mặt là một trong những lĩnh vực nghiên cứu tiêu biểu với tiềm năng ứng dụng trong giáo dục, y tế, chăm sóc sức khỏe tinh thần và giao tiếp người-máy. Tuy nhiên, các phương pháp truyền thống thường gặp khó khăn trong việc xử lý dữ liệu hình ảnh phức tạp, dẫn đến hiệu suất chưa cao.

Mạng đối kháng tạo (GAN) đã chứng minh khả năng vượt trội trong việc tạo và khuếch đại dữ liệu hình ảnh, từ đó cải thiện đáng kể hiệu quả của các hệ thống AI. Việc áp dụng GAN để phát triển hệ thống nhận diện cảm xúc không chỉ giúp nâng cao độ chính xác mà còn mở ra hướng đi mới trong nghiên cứu và ứng dụng. Vì những lý do này, đề tài "Phát triển hệ thống nhận diện cảm xúc từ hình ảnh khuôn mặt sử dụng mạng GAN" đã được chọn để nghiên cứu.

### **2. Mục đích nghiên cứu**

1. Xây dựng một hệ thống nhận diện cảm xúc từ hình ảnh khuôn mặt với độ chính xác cao thông qua việc áp dụng mạng GAN.
2. Khai thác khả năng của GAN trong việc tổng hợp và khuếch đại dữ liệu hình ảnh nhằm giải quyết vấn đề thiếu hụt dữ liệu đào tạo.
3. Đánh giá hiệu suất của mô hình GAN trong bài toán nhận diện cảm xúc so với các phương pháp truyền thống.
4. Đề xuất một giải pháp khả thi và hiệu quả cho các ứng dụng thực tiễn, như phân tích tâm lý, hỗ trợ y tế và cải thiện trải nghiệm giao tiếp người-máy.

### **3. Đối tượng và phạm vi nghiên cứu**

1. **Đối tượng nghiên cứu:** Các mô hình học sâu, đặc biệt là mạng GAN và ứng dụng của nó trong nhận diện cảm xúc từ hình ảnh khuôn mặt.
2. **Phạm vi nghiên cứu:**

* Sử dụng các bộ dữ liệu phổ biến như FER2013, AffectNet để thực hiện huấn luyện và đánh giá mô hình.
* Tập trung vào việc nhận diện các cảm xúc cơ bản: vui, buồn, giận dữ, sợ hãi, ngạc nhiên và trung tính.
* Nghiên cứu các kỹ thuật tiền xử lý dữ liệu, thiết kế và tối ưu hóa mạng GAN để phục vụ bài toán nhận diện cảm xúc.

### **4. Phương pháp nghiên cứu**

* **Nghiên cứu lý thuyết:**
  + Tìm hiểu các kiến thức cơ bản về mạng GAN, nhận diện cảm xúc từ hình ảnh khuôn mặt và các phương pháp học sâu liên quan.
  + Phân tích các nghiên cứu trước đây để lựa chọn hướng đi phù hợp.
* **Phương pháp thực nghiệm:**
  + Thu thập và xử lý dữ liệu từ các bộ dữ liệu có sẵn.
  + Xây dựng mô hình GAN tùy chỉnh, kết hợp với mạng học sâu để nhận diện cảm xúc.
  + Huấn luyện, kiểm tra và so sánh hiệu suất của mô hình với các phương pháp truyền thống.
* **Phương pháp đánh giá:**
  + Sử dụng các chỉ số đánh giá như Accuracy, Precision, Recall, F1-Score để đánh giá độ chính xác của mô hình.
  + So sánh kết quả thực nghiệm với các phương pháp truyền thống nhằm làm rõ hiệu quả của hệ thống.

## **B. Danh sách thành viên và phần công việc**

|  |  |  |
| --- | --- | --- |
| STT | Họ và tên | Công việc |
| 1 | Nguyễn Việt Tùng | word |
| 2 | Nguyễn Ngọc Tuyền | Code |
| 3 | Trương Khánh Linh | Pp |
| 4 | Nguyễn Đại Dương | word |
| 5 | Nguyễn Trung Đức | word |

# **CHƯƠNG I: CƠ SỞ LÝ THUYẾT**

## **1.1. Giới thiệu chung về bài toán nhận diện cảm xúc**

Nhận diện cảm xúc là một lĩnh vực trong xử lý ảnh và trí tuệ nhân tạo, tập trung vào việc phân tích và xác định trạng thái cảm xúc của con người dựa trên các đặc trưng sinh học như khuôn mặt, giọng nói, hoặc hành vi cử chỉ. Đây là một lĩnh vực quan trọng vì cảm xúc là yếu tố trung tâm trong giao tiếp của con người và là nền tảng để cải thiện tương tác giữa người và máy móc.

Hệ thống nhận diện cảm xúc sử dụng các thuật toán xử lý ảnh và học máy để trích xuất đặc điểm từ hình ảnh khuôn mặt, phân tích chúng để nhận dạng cảm xúc như vui, buồn, tức giận, sợ hãi, ngạc nhiên hoặc trung tính. Điều này tạo ra tiềm năng ứng dụng rộng rãi trong các lĩnh vực giáo dục, y tế, và tương tác người-máy.

### **1.1.1. Tầm quan trọng của nhận diện cảm xúc [2]**

**Cải thiện giao tiếp tự nhiên**

Con người giao tiếp không chỉ qua ngôn ngữ mà còn qua các biểu cảm khuôn mặt, cử chỉ và giọng nói. Nhận diện cảm xúc giúp các hệ thống máy móc trở nên nhạy bén hơn với các tín hiệu phi ngôn ngữ, tạo ra các phản hồi tương tác tự nhiên và phù hợp hơn với trạng thái của người dùng.

**Ứng dụng đa lĩnh vực**

Nhận diện cảm xúc không chỉ giới hạn ở nghiên cứu mà còn được áp dụng rộng rãi trong các lĩnh vực như chăm sóc sức khỏe, giáo dục, tâm lý học, và thương mại, mang lại những lợi ích thiết thực như nâng cao hiệu quả công việc, tối ưu hóa trải nghiệm người dùng, và hỗ trợ ra quyết định.

**Hướng tới xã hội thông minh**

Trong bối cảnh công nghệ ngày càng phát triển, các hệ thống thông minh có khả năng nhận diện cảm xúc sẽ trở thành yếu tố cốt lõi trong xây dựng thành phố thông minh, hệ thống chăm sóc khách hàng tự động và các thiết bị gia dụng thông minh.

### **1.1.2. Ứng dụng của hệ thống nhận diện cảm xúc [2]**

1. **Trong giáo dục**

**Phân tích cảm xúc học sinh:**

Hệ thống nhận diện cảm xúc hỗ trợ giáo viên theo dõi trạng thái cảm xúc của học sinh trong lớp học trực tuyến hoặc trực tiếp. Ví dụ, nếu học sinh có biểu hiện chán nản hoặc mất tập trung, giáo viên có thể thay đổi cách tiếp cận để tạo hứng thú và khuyến khích học tập.

**Đánh giá hiệu quả giảng dạy:**

Các lớp học có thể được trang bị hệ thống nhận diện cảm xúc để đánh giá phản ứng của học sinh với từng phần nội dung bài giảng. Điều này giúp điều chỉnh giáo án hoặc phương pháp dạy phù hợp hơn với từng nhóm học sinh.

**Cá nhân hóa việc học:**

Dựa trên cảm xúc nhận diện được, hệ thống có thể gợi ý các tài liệu học tập hoặc bài tập phù hợp, giúp tăng cường hiệu quả học tập cho từng cá nhân.

1. **Trong y tế**

**Theo dõi sức khỏe tâm lý:**

Nhận diện cảm xúc giúp bác sĩ hoặc chuyên gia tâm lý phát hiện và theo dõi trạng thái cảm xúc của bệnh nhân trong các liệu pháp điều trị, đặc biệt là các bệnh liên quan đến tâm thần như trầm cảm, lo âu.

**Chẩn đoán và điều trị:**

Hệ thống này có thể phân tích cảm xúc từ khuôn mặt bệnh nhân để phát hiện các dấu hiệu tâm lý bất thường, hỗ trợ việc chẩn đoán chính xác và điều trị kịp thời.

**Giám sát từ xa:**

Trong các trường hợp bệnh nhân không thể đến khám trực tiếp, nhận diện cảm xúc qua video có thể giúp bác sĩ theo dõi tình trạng sức khỏe từ xa, cải thiện khả năng chăm sóc liên tục.

1. **Trong tương tác người-máy**

**Cải thiện trải nghiệm người dùng:**

Các hệ thống thông minh, như trợ lý ảo (Siri, Alexa), có thể nhận diện cảm xúc để phản hồi một cách đồng cảm và tự nhiên. Ví dụ, nếu người dùng có biểu hiện căng thẳng, hệ thống có thể điều chỉnh giọng nói nhẹ nhàng hơn hoặc đề xuất các biện pháp thư giãn.

**Hỗ trợ trong môi trường thông minh:**

Các thiết bị trong nhà thông minh có thể phản ứng với cảm xúc của người dùng, chẳng hạn như điều chỉnh ánh sáng và âm nhạc phù hợp với tâm trạng của họ.

**Ứng dụng trong giao thông và xe tự hành:**

Hệ thống nhận diện cảm xúc có thể được tích hợp trong xe hơi để theo dõi trạng thái cảm xúc của tài xế, cảnh báo khi phát hiện sự mất tập trung hoặc mệt mỏi, từ đó tăng cường an toàn giao thông.

**Robot tương tác xã hội:**

Trong lĩnh vực robot xã hội, các robot có khả năng nhận diện cảm xúc sẽ đóng vai trò như người bạn đồng hành, hỗ trợ con người trong các công việc thường nhật hoặc cung cấp sự an ủi trong các tình huống căng thẳng.

## **1.2. Các phương pháp nhận diện cảm xúc truyền thống [3]**

Nhận diện cảm xúc truyền thống tập trung vào việc sử dụng các đặc trưng thủ công và các mô hình học máy cổ điển để phân loại cảm xúc từ khuôn mặt. Mặc dù các phương pháp này không hiệu quả bằng các mô hình học sâu hiện đại, chúng vẫn có vai trò quan trọng trong việc đặt nền móng cho nghiên cứu và ứng dụng trong lĩnh vực xử lý ảnh và thị giác máy tính.

### **1.2.1. Dựa trên đặc trưng hình ảnh thủ công**

1. **Đặc trưng Haar-like**

**Khái niệm:**

* + Đặc trưng Haar-like là một phương pháp trích xuất đặc trưng dựa trên cường độ ánh sáng trong các vùng hình chữ nhật của hình ảnh.
  + Các đặc trưng này so sánh sự khác biệt về cường độ sáng giữa các vùng sáng và tối trong ảnh, từ đó phát hiện các chi tiết quan trọng trên khuôn mặt như mắt, mũi, miệng.

**Ưu điểm:**

* + Tính toán đơn giản và nhanh chóng.
  + Hiệu quả cao trong việc phát hiện các cấu trúc đặc trưng trên khuôn mặt.

**Hạn chế:**

* + Dễ bị ảnh hưởng bởi ánh sáng và góc chụp.
  + Đòi hỏi nhiều công sức trong việc thiết kế đặc trưng và lựa chọn các tham số phù hợp.

**Ứng dụng trong nhận diện cảm xúc:**

* + Kết hợp với các thuật toán phân loại như SVM để phân loại cảm xúc.

1. **Phương pháp LBP (Local Binary Pattern)**

**Khái niệm:**

* + Local Binary Pattern (LBP) là một phương pháp trích xuất đặc trưng cục bộ dựa trên việc mã hóa mối quan hệ giữa mỗi pixel và các pixel lân cận.
  + Phương pháp này chuyển đổi hình ảnh thành các biểu đồ biểu diễn cấu trúc bề mặt, giúp phân biệt các đặc trưng khuôn mặt.

**Cách hoạt động:**

* + So sánh giá trị của mỗi pixel với các pixel lân cận.
  + Mỗi pixel được biểu diễn bằng một mã nhị phân dựa trên sự khác biệt về cường độ sáng.
  + Các mã nhị phân được tổng hợp thành một biểu đồ để làm đặc trưng đầu vào cho các thuật toán phân loại.

**Ưu điểm:**

* + Hiệu quả trong việc trích xuất đặc trưng trên các ảnh có điều kiện ánh sáng thay đổi.
  + Dễ triển khai và tính toán nhanh.

**Hạn chế:**

* + Không nắm bắt được thông tin về cấu trúc toàn cục.
  + Nhạy cảm với nhiễu trong dữ liệu.

**Ứng dụng trong nhận diện cảm xúc:**

* + LBP thường được kết hợp với các thuật toán như SVM hoặc Random Forest để xây dựng các hệ thống nhận diện cảm xúc.

### **1.2.2. Dựa trên mô hình học máy truyền thống**

1. **SVM (Support Vector Machines)**

**Khái niệm:**

* + SVM là một thuật toán học máy giám sát được sử dụng rộng rãi trong phân loại và hồi quy. Trong nhận diện cảm xúc, SVM hoạt động bằng cách tìm một siêu phẳng tối ưu để phân biệt các biểu cảm khuôn mặt dựa trên đặc trưng đã trích xuất.

**Cách hoạt động:**

* + Sử dụng các đặc trưng như LBP hoặc Haar-like làm đầu vào.
  + SVM tìm kiếm một siêu phẳng trong không gian đặc trưng để phân tách các nhãn cảm xúc (vui, buồn, giận dữ, ngạc nhiên, v.v.).
  + Với các bài toán không tuyến tính, SVM áp dụng các kernel (như RBF hoặc Polynomial) để mở rộng không gian đặc trưng.

**Ưu điểm:**

* + Hiệu quả cao với các bộ dữ liệu nhỏ hoặc trung bình.
  + Có khả năng xử lý các bài toán tuyến tính và phi tuyến tính.

**Hạn chế:**

* + Độ phức tạp tính toán cao với dữ liệu lớn.
  + Hiệu quả giảm nếu đặc trưng đầu vào không được trích xuất tốt.

**Ứng dụng trong nhận diện cảm xúc:**

* + Phân loại cảm xúc với độ chính xác cao khi sử dụng các đặc trưng hình ảnh mạnh mẽ như HOG hoặc LBP.

1. **Random Forest**

**Khái niệm:**

* + Random Forest là một mô hình học máy dựa trên việc xây dựng nhiều cây quyết định độc lập và sử dụng cơ chế bỏ phiếu để dự đoán nhãn cuối cùng.
  + Trong nhận diện cảm xúc, Random Forest tận dụng sự ngẫu nhiên trong việc chọn mẫu và đặc trưng để giảm thiểu overfitting và tăng độ chính xác.

**Cách hoạt động:**

* + Trích xuất các đặc trưng như LBP hoặc Haar-like từ ảnh.
  + Random Forest tạo ra một tập hợp các cây quyết định dựa trên các đặc trưng đầu vào.
  + Kết quả cuối cùng được quyết định dựa trên kết quả bỏ phiếu từ các cây.

**Ưu điểm:**

* + Khả năng xử lý dữ liệu đa chiều tốt.
  + Độ bền cao với nhiễu và dữ liệu không đồng nhất.
  + Hiệu quả trong việc tránh overfitting nhờ tính ngẫu nhiên.

**Hạn chế:**

* + Đòi hỏi nhiều tài nguyên tính toán hơn so với các thuật toán đơn giản khác.
  + Hiệu quả phụ thuộc vào số lượng cây và độ sâu của mỗi cây.

**Ứng dụng trong nhận diện cảm xúc:**

* + Random Forest được sử dụng như một thuật toán phân loại hiệu quả khi kết hợp với các đặc trưng hình ảnh cục bộ như LBP hoặc GLCM (Gray-Level Co-occurrence Matrix).

## **1.3. Giới thiệu mạng GAN (Generative Adversarial Network)**

Mạng GAN (Generative Adversarial Network) là một trong những phát minh nổi bật trong lĩnh vực học sâu, được Ian Goodfellow và các cộng sự giới thiệu vào năm 2014. GAN là một mô hình học không giám sát sử dụng hai mạng nơ-ron đối kháng (adversarial) để tạo ra dữ liệu mới có tính chân thực cao dựa trên dữ liệu đầu vào ban đầu.

GAN đã được ứng dụng rộng rãi trong nhiều lĩnh vực, bao gồm xử lý ảnh, tăng cường dữ liệu, tái tạo khuôn mặt, và nhận diện cảm xúc.

### **1.3.1. Kiến trúc của GAN [1]**

Mạng GAN bao gồm hai thành phần chính:

* **Generator (Mạng sinh)**: Tạo ra dữ liệu giả.
* **Discriminator (Mạng phân biệt)**: Phân biệt giữa dữ liệu thật và dữ liệu giả.

Hai mạng này hoạt động đối kháng với nhau trong quá trình huấn luyện, nhằm cải thiện hiệu quả tổng thể.

1. **Generator**

**Chức năng:**

* + Generator có nhiệm vụ tạo ra dữ liệu giả (như hình ảnh, âm thanh, hoặc video) sao cho dữ liệu này giống với dữ liệu thật nhất có thể.
  + Generator nhận một vector nhiễu ngẫu nhiên (noise vector zzz) làm đầu vào, và biến nó thành dữ liệu đầu ra.

**Cách hoạt động:**

* + Generator là một mạng nơ-ron sâu (Deep Neural Network) hoặc mạng tích chập (CNN) với nhiều tầng phi tuyến.
  + Nhiệm vụ của Generator là học cách "lừa" Discriminator, tức là tạo ra các mẫu dữ liệu giả mà Discriminator không thể phân biệt với dữ liệu thật.

**Mục tiêu:**

* + Tối ưu hóa hàm mất mát sao cho xác suất Discriminator phân loại dữ liệu giả là thật đạt cao nhất.
  + Hàm mất mát của Generator: LG=log⁡(1−D(G(z)))\mathcal{L}\_G = \log(1 - D(G(z)))LG​=log(1−D(G(z))) Trong đó:
    - G(z)G(z)G(z): Đầu ra của Generator.
    - D(x)D(x)D(x): Đầu ra của Discriminator cho mẫu xxx.

**Kiến trúc tiêu biểu:**

* + Sử dụng các tầng nén (deconvolution layers) để tăng kích thước dữ liệu đầu ra.
  + Áp dụng hàm kích hoạt (activation function) như ReLU hoặc Tanh để đảm bảo tín hiệu phù hợp.

1. **Discriminator**

**Chức năng:**

* + Discriminator hoạt động như một bộ phân loại nhị phân, nhiệm vụ là phân biệt giữa dữ liệu thật và dữ liệu giả do Generator tạo ra.
  + Đầu vào của Discriminator có thể là dữ liệu thật từ tập huấn luyện hoặc dữ liệu giả từ Generator.

**Cách hoạt động:**

* + Discriminator là một mạng nơ-ron sâu hoặc mạng tích chập, được thiết kế để học các đặc trưng quan trọng từ dữ liệu.
  + Nó trả về một xác suất (thường trong khoảng [0,1]) thể hiện khả năng mẫu đầu vào là thật.

**Mục tiêu:**

* + Tối ưu hóa hàm mất mát sao cho Discriminator phân biệt chính xác giữa dữ liệu thật và giả.
  + Hàm mất mát của Discriminator: LD=−[log⁡(D(x))+log⁡(1−D(G(z)))]\mathcal{L}\_D = -[\log(D(x)) + \log(1 - D(G(z)))]LD​=−[log(D(x))+log(1−D(G(z)))] Trong đó:
    - xxx: Mẫu dữ liệu thật.
    - G(z)G(z)G(z): Dữ liệu giả do Generator sinh ra.

**Kiến trúc tiêu biểu:**

* + Sử dụng các tầng tích chập (convolution layers) để trích xuất đặc trưng.
  + Áp dụng hàm kích hoạt như Leaky ReLU để tăng cường khả năng học.

1. **Quá trình huấn luyện GAN**

**Nguyên tắc đối kháng:**

* + Generator cố gắng lừa Discriminator bằng cách tạo ra dữ liệu giả giống thật.
  + Discriminator cố gắng phân biệt chính xác dữ liệu thật và giả.

**Tiến trình:**

* + Cập nhật Generator:
    - Sử dụng đầu ra của Discriminator để đánh giá chất lượng dữ liệu giả và điều chỉnh Generator.
  + Cập nhật Discriminator:
    - Dựa trên dữ liệu đầu vào (thật và giả) để học cách phân biệt tốt hơn.

**Cân bằng:**

* + Quá trình huấn luyện tiếp tục cho đến khi cả Generator và Discriminator đạt được cân bằng Nash, tức là Generator tạo ra dữ liệu giả không thể phân biệt với dữ liệu thật.

### **1.3.2. Cơ chế hoạt động của GAN [1]**

Cơ chế hoạt động của mạng GAN dựa trên nguyên lý **đối kháng** giữa hai thành phần chính: **Generator** (mạng sinh) và **Discriminator** (mạng phân biệt). Hai mạng này được huấn luyện đồng thời với các mục tiêu trái ngược nhau, từ đó cải thiện chất lượng của dữ liệu giả do Generator tạo ra.

**Quá trình khởi tạo**

**Generator:**

* + Nhận đầu vào là một vector nhiễu ngẫu nhiên zzz từ phân phối xác suất (thường là phân phối Gaussian hoặc Uniform).
  + Dựa trên zzz, Generator sinh ra một mẫu dữ liệu giả G(z)G(z)G(z) có cùng định dạng với dữ liệu thật.

**Discriminator:**

* + Nhận đầu vào là dữ liệu thật từ tập huấn luyện xxx hoặc dữ liệu giả G(z)G(z)G(z) do Generator sinh ra.
  + Phân biệt và đưa ra xác suất D(x)D(x)D(x):
    - D(x)≈1D(x) \approx 1D(x)≈1 nếu xxx là dữ liệu thật.
    - D(G(z))≈0D(G(z)) \approx 0D(G(z))≈0 nếu G(z)G(z)G(z) là dữ liệu giả.

**Quá trình huấn luyện GAN**

Quá trình huấn luyện GAN bao gồm hai giai đoạn chính, diễn ra xen kẽ nhau:

* Mục tiêu của Discriminator là tối đa hóa khả năng phân biệt giữa dữ liệu thật và dữ liệu giả.
* Discriminator được cập nhật bằng cách tối ưu hóa hàm mất mát: LD=−[log⁡(D(x))+log⁡(1−D(G(z)))]\mathcal{L}\_D = -[\log(D(x)) + \log(1 - D(G(z)))]LD​=−[log(D(x))+log(1−D(G(z)))] Trong đó:
  + D(x)D(x)D(x): Xác suất mẫu xxx là dữ liệu thật.
  + 1−D(G(z))1 - D(G(z))1−D(G(z)): Xác suất mẫu G(z)G(z)G(z) là dữ liệu giả.

**Cập nhật Discriminator:**

1. Lấy một batch dữ liệu thật từ tập huấn luyện.
2. Sinh một batch dữ liệu giả G(z)G(z)G(z) từ Generator.
3. Tính toán hàm mất mát LD\mathcal{L}\_DLD​ và cập nhật các tham số của Discriminator.

**Huấn luyện Generator**

* Mục tiêu của Generator là "lừa" Discriminator, tức làm cho D(G(z))D(G(z))D(G(z)) gần với 1 (Discriminator nghĩ rằng dữ liệu giả là thật).
* Generator được cập nhật bằng cách tối ưu hóa hàm mất mát: LG=−log⁡(D(G(z)))\mathcal{L}\_G = -\log(D(G(z)))LG​=−log(D(G(z))) Hoặc tương đương: LG=log⁡(1−D(G(z)))(thường ıˊt duˋng hơn vıˋ hội tụ chậm)\mathcal{L}\_G = \log(1 - D(G(z))) \quad \text{(thường ít dùng hơn vì hội tụ chậm)} LG​=log(1−D(G(z)))(thường ıˊt duˋng hơn vıˋ hội tụ chậm)

**Cập nhật Generator:**

1. Sinh một batch dữ liệu giả G(z)G(z)G(z).
2. Đưa dữ liệu này vào Discriminator để nhận phản hồi (xác suất D(G(z))D(G(z))D(G(z))).
3. Tính toán hàm mất mát LG\mathcal{L}\_GLG​ và cập nhật các tham số của Generator.
4. **Quá trình học đối kháng**

Generator và Discriminator hoạt động như hai bên đối thủ trong một trò chơi có tổng bằng không (**zero-sum game**):

* Generator cố gắng "lừa" Discriminator.
* Discriminator cố gắng phân biệt dữ liệu thật và giả.

Hàm mất mát tổng hợp của GAN là:

L=min⁡Gmax⁡DEx∼pdata[log⁡(D(x))]+Ez∼pz[log⁡(1−D(G(z)))]\mathcal{L} = \min\_G \max\_D \mathbb{E}\_{x \sim p\_\text{data}}[\log(D(x))] + \mathbb{E}\_{z \sim p\_z}[\log(1 - D(G(z)))]L=Gmin​Dmax​Ex∼pdata​​[log(D(x))]+Ez∼pz​​[log(1−D(G(z)))]

Trong đó:

* pdatap\_\text{data}pdata​: Phân phối dữ liệu thật.
* pzp\_zpz​: Phân phối nhiễu đầu vào.

**Cân bằng Nash trong GAN**

* Khi đạt cân bằng Nash, Generator sinh ra dữ liệu giả mà Discriminator không thể phân biệt với dữ liệu thật.
* Điều này xảy ra khi: D(x)=0.5,∀xD(x) = 0.5, \forall xD(x)=0.5,∀x Nghĩa là Discriminator không còn khả năng phân biệt giữa dữ liệu thật và dữ liệu giả, Generator đã học được phân phối của dữ liệu thật.

**Các thách thức trong quá trình huấn luyện GAN**

**Mất cân bằng giữa Generator và Discriminator:**

* + Nếu Discriminator quá mạnh, Generator không học được.
  + Nếu Generator quá mạnh, Discriminator không phân biệt được dữ liệu.

**Vanishing Gradient (Gradient biến mất):**

* + Khi Discriminator phân biệt quá tốt, Gradient trở nên quá nhỏ, dẫn đến Generator khó cải thiện.

**Mode Collapse:**

* + Generator chỉ sinh ra một số lượng mẫu hạn chế, thiếu đa dạng.

**Hội tụ không ổn định:**

* + GAN khó hội tụ vì hai mạng luôn cố gắng "đánh bại" nhau.

### **1.3.3. Các biến thể phổ biến của GAN [4]**

Kể từ khi GAN được giới thiệu, nhiều biến thể đã được phát triển để cải thiện tính linh hoạt, khả năng sinh dữ liệu, và hiệu suất của mô hình. Một số biến thể nổi bật bao gồm **Conditional GAN (cGAN)** và **StyleGAN**, mỗi loại được thiết kế để giải quyết các vấn đề cụ thể hoặc mở rộng khả năng ứng dụng của GAN.

1. **Conditional GAN (cGAN)**

**Khái niệm:**

Conditional GAN (cGAN) là một biến thể của GAN, trong đó cả Generator và Discriminator đều được "điều kiện hóa" (conditioned) trên một thông tin bổ sung yyy. Thông tin yyy này có thể là một nhãn lớp, một hình ảnh khác, hoặc bất kỳ thông tin nào liên quan.

Ví dụ: Với yyy là nhãn cảm xúc, cGAN có thể sinh ra khuôn mặt có cảm xúc cụ thể như vui, buồn, hoặc tức giận.

**Kiến trúc cGAN:**

* **Generator:**
  + Nhận đầu vào gồm vector nhiễu ngẫu nhiên zzz và thông tin bổ sung yyy.
  + Học cách sinh dữ liệu giả G(z,y)G(z, y)G(z,y) có liên quan đến yyy.
* **Discriminator:**
  + Nhận đầu vào là dữ liệu (thật hoặc giả) cùng với thông tin bổ sung yyy.
  + Học cách phân biệt dữ liệu thật và giả dựa trên yyy.

**Hàm mất mát:**

Hàm mất mát của cGAN được điều chỉnh để kết hợp thông tin điều kiện:

L=min⁡Gmax⁡DEx∼pdata[log⁡(D(x∣y))]+Ez∼pz[log⁡(1−D(G(z∣y)∣y))]\mathcal{L} = \min\_G \max\_D \mathbb{E}\_{x \sim p\_\text{data}}[\log(D(x|y))] + \mathbb{E}\_{z \sim p\_z}[\log(1 - D(G(z|y)|y))]L=Gmin​Dmax​Ex∼pdata​​[log(D(x∣y))]+Ez∼pz​​[log(1−D(G(z∣y)∣y))]

**Ưu điểm:**

* Giúp GAN sinh dữ liệu có kiểm soát, phù hợp với các điều kiện cụ thể.
* Tăng cường tính ứng dụng trong các lĩnh vực như sinh hình ảnh, chuyển đổi phong cách, và nhận diện cảm xúc.

**Ứng dụng:**

* Sinh ảnh thuộc các danh mục cụ thể (ví dụ: chó, mèo).
* Tạo khuôn mặt với biểu cảm hoặc đặc điểm cụ thể (ví dụ: khuôn mặt người già, trẻ).
* Tăng cường dữ liệu trong nhận diện cảm xúc, sinh ảnh với cảm xúc được gán nhãn.

1. **StyleGAN**

**Khái niệm:**

StyleGAN là một biến thể tiên tiến của GAN, được phát triển bởi NVIDIA, tập trung vào việc kiểm soát phong cách và chi tiết của hình ảnh đầu ra. Mô hình này mang lại khả năng điều chỉnh các đặc điểm hình ảnh ở nhiều cấp độ khác nhau, từ cấu trúc tổng quát đến các chi tiết nhỏ.

**Kiến trúc StyleGAN:**

**Mapping Network:**

* + Biến vector nhiễu ngẫu nhiên zzz thành vector phong cách www trong không gian tiềm ẩn.
  + Không gian www được thiết kế để dễ dàng điều chỉnh các thuộc tính phong cách của hình ảnh.

**Synthesis Network:**

* + Nhận đầu vào từ www và sử dụng các tầng tích chập để sinh ra hình ảnh.
  + Mỗi tầng tích chập được điều chỉnh bởi www, cho phép kiểm soát các đặc điểm ở từng cấp độ (ví dụ: hình dạng tổng thể, màu sắc, chi tiết).

**Adaptive Instance Normalization (AdaIN):**

* + Được sử dụng để điều chỉnh phong cách ở từng tầng.
  + AdaIN kết hợp thông tin từ www và đặc trưng của hình ảnh hiện tại để kiểm soát chi tiết phong cách.

**Hàm mất mát:**

* StyleGAN sử dụng kỹ thuật **Progressive Growing GAN** để cải thiện độ ổn định và chất lượng hình ảnh.
* Hàm mất mát tương tự GAN cơ bản nhưng được cải thiện với các kỹ thuật tối ưu hóa để tránh mode collapse.

**Ưu điểm:**

* Kiểm soát phong cách mạnh mẽ: Thay đổi các thuộc tính hình ảnh một cách mượt mà.
* Sinh hình ảnh có độ phân giải cao với chi tiết chính xác.
* Hỗ trợ đa dạng hóa dữ liệu đầu ra.

**Ứng dụng:**

* Sinh khuôn mặt người có độ chân thực cao.
* Chuyển đổi phong cách hình ảnh, từ hiện đại sang cổ điển hoặc ngược lại.
* Tăng cường dữ liệu cho nhận diện cảm xúc bằng cách sinh khuôn mặt với biểu cảm chi tiết.

# **CHƯƠNG II:** **XÂY DỰNG HỆ THỐNG NHẬN DIỆN CẢM XÚC**

## **2.1. Định nghĩa bài toán**

### **2.1.1. Mô tả bài toán**

Xác định các cảm xúc cơ bản trong hệ thống nhận diện cảm xúc

Hệ thống nhận diện cảm xúc từ hình ảnh khuôn mặt tập trung vào sáu cảm xúc cơ bản, bao gồm: **vui**, **buồn**, **giận dữ**, **sợ hãi**, **ngạc nhiên**, và **trung tính**.



Hình 2.1 Cảm xúc **vui, buồn, giận dữ, sợ hãi, ngạc nhiên,** và **trung tính.**

Dưới đây là mô tả chi tiết về từng loại cảm xúc và đặc điểm nhận diện đặc trưng:

1. **Vui (Happiness)**

* **Đặc điểm khuôn mặt:**

Miệng mở rộng, tạo thành nụ cười.

Góc miệng nhếch lên rõ rệt.

Mắt híp lại nhẹ, thường xuất hiện nếp nhăn ở khóe mắt.

* **Ứng dụng:**

Theo dõi phản ứng tích cực trong các hoạt động xã hội.

Đánh giá trải nghiệm của người dùng đối với sản phẩm, dịch vụ.

1. **Buồn (Sadness)**

* **Đặc điểm khuôn mặt:**

Góc miệng hạ xuống.

Mí mắt trùng xuống, ánh mắt hướng xuống hoặc xa xăm.

Chân mày nhíu lại nhẹ, hơi nghiêng về phía trung tâm.

* **Ứng dụng:**

Nhận diện cảm xúc tiêu cực để đưa ra hỗ trợ tâm lý kịp thời.

Phân tích phản ứng của người dùng đối với các nội dung cảm xúc.

1. **Giận dữ (Anger)**

* **Đặc điểm khuôn mặt:**

Chân mày nhíu mạnh, hai bên kéo về phía trung tâm.

Đôi mắt mở lớn, tạo cảm giác căng thẳng hoặc nhìn chằm chằm.

Miệng có thể mở rộng, môi mím chặt hoặc biểu hiện nghiến răng.

* **Ứng dụng:**

Phát hiện các dấu hiệu căng thẳng hoặc xung đột trong môi trường làm việc.

Theo dõi hành vi trong các tình huống nhạy cảm.

1. **Sợ hãi (Fear)**

* **Đặc điểm khuôn mặt:**

Đôi mắt mở lớn, mí mắt trên nhướng cao.

Lông mày nhướng lên và kéo về phía trung tâm.

Miệng mở ra, thường có hình chữ nhật, cằm căng thẳng.

* **Ứng dụng:**

Phát hiện cảm giác lo lắng, nguy hiểm trong các tình huống khẩn cấp.

Phân tích phản ứng của người dùng đối với các tình huống bất ngờ.

1. **Ngạc nhiên (Surprise)**

* **Đặc điểm khuôn mặt:**

Đôi mắt mở lớn, mí mắt trên nhướng cao rõ rệt.

Lông mày nhướng lên, tạo khoảng cách lớn hơn giữa mắt và lông mày.

Miệng mở ra, thường không đối xứng.

* **Ứng dụng:**

Theo dõi phản ứng của người dùng trước các yếu tố bất ngờ.

Phân tích hành vi trong các tình huống học tập, sáng tạo.

1. **Trung tính (Neutral)**

* **Đặc điểm khuôn mặt:**

Không có biểu hiện rõ rệt trên khuôn mặt.

Môi khép lại, không cười cũng không mím.

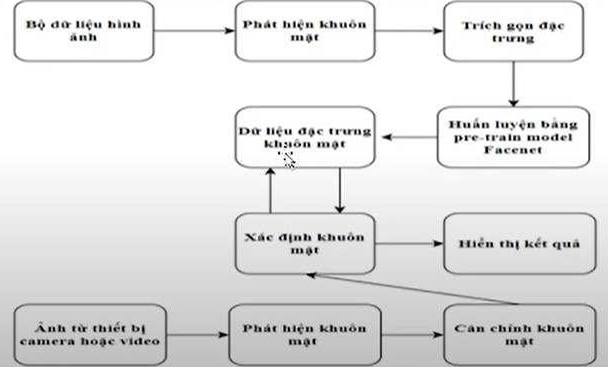
Mắt mở vừa phải, không căng thẳng hay nhíu mày.

* **Ứng dụng:**

Là trạng thái cơ sở để so sánh với các cảm xúc khác.

Phát hiện sự thay đổi cảm xúc từ trạng thái trung tính.

**Mô hình hóa bài toán**

****

Hình 2.2. Mô hình bài toán

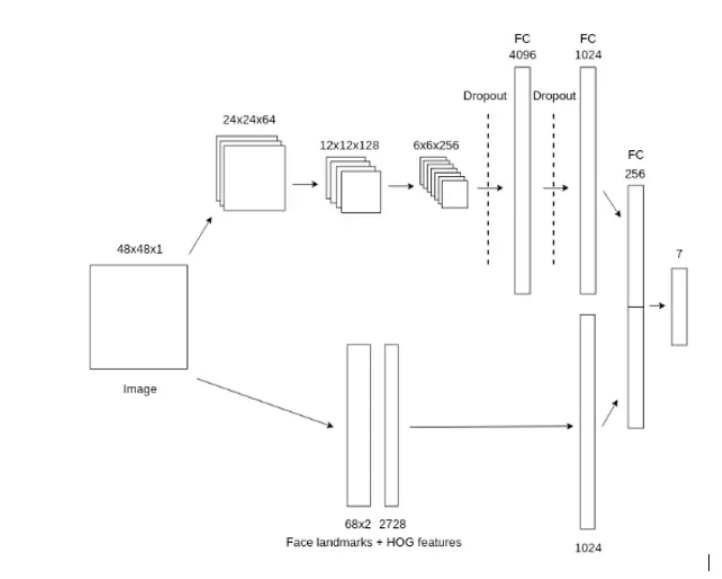
**Bài toán**

* + Nhận ảnh và tiền xử lí
  + Trích xuất đặc trưng
  + Phân lớp và nhận diện cảm xúc - Trong giai đoạn nhận ảnh và tiền xử

lý:

* + Đầu vào ở đây là có thể là ảnh chụp hoặc là các dataset
  + Trải qua một số bước tiền xử lí để tang chất lượng hình ảnh để giúp cho việc nhận diện cảm xúc trở nên hiệu quả hơn
* Trích xuất đặc trưng là bước gần như quan trọng nhất :
* Các bước đặc trưng khuôn mặt được tính toán trên các thuật toán có sẵn, kết quả thường là vecter đặc trưng làm đầu vào cho các bước sau - Phân lớp và nhận diện cảm xúc:
* Đây là một bài toán điển hình có thể dung nhiều thuật toán để áp dụng cho nó

-Mô hình kiến trúc nhận dạng cảm xúc khuôn mặt :

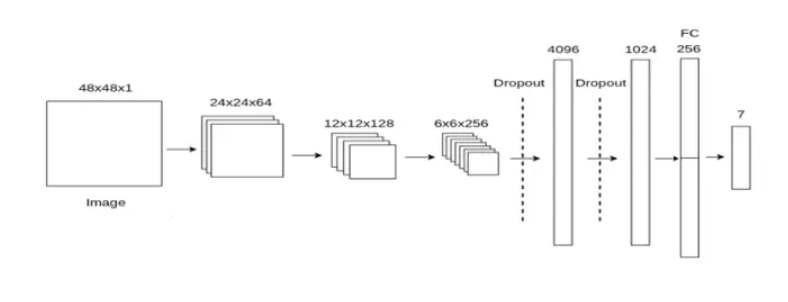


Hình 2.3. mô hình kết hợp đặc trưng truyền thống cho hệ thống nhận dạng cảm xúc qua khuôn mặt

Mô hình thiết kế của hệ thống nhận dạng cảm xúc qua khuôn mặt sử dụng phương pháp CNN kết hợp với đặc trưng truyền thống. Đầu vào là hình ảnh có kích thước 48x48 pixel và 1 chanels (ảnh xám) được lấy trong bộ dữ liệu dataset FER2013, đầu ra của hệ thống là dự đoán cảm xúc của hình ảnh. Để thực hiện quá trình có được đầu ra hệ thống cần thực hiện theo hai giai đoạn: tiền xử lý hình ảnh, phân lớp sử dụng học sâu. Trong đó bước trích chọn đặc trưng được thực hiện một cách không tường minh, ẩn bên trong giai đoạn phân lớp sử dụng học sâu, các giai đoạn thực hiện một cách liền mạch không phân chia một cách quá cụ thể và rõ

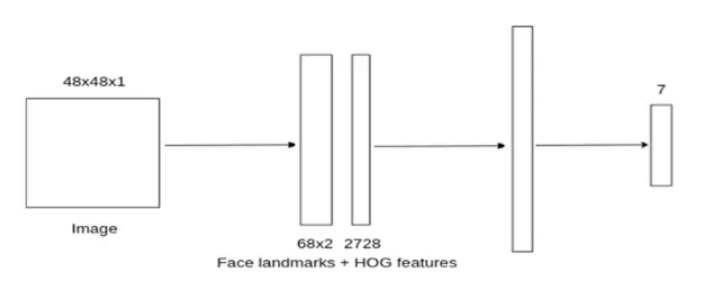
ràng như phương pháp truyền thống. Hình 4.1 mô hình kiến trúc của phương pháp CNN, đầu vào là hình ảnh có kích thước là 48x48, sau đó đồng thời thực hiện 2 việc song song:

* Ảnh đầu vào thực hiện đi qua phương pháp deep learning hiện đại. Trong giai đoạn này, thực hiện đi qua các bước các tầng tích chập, tầng gộp, tầng kết nối. Tạo ra tầng kết nối có kích thước 1024.

****

Hình 2.4. mô hình tạo ra tầng kết nối có kích thước 1024

* Ảnh đầu vào kết hợp với các đặc trưng truyền thống (Face lanmarks + HoG features). Đầu ra là tạo ra tầng kết nối có kích thước 1024. Cuối cùng, hai công việc thực hiện song song gộp lại tạo thành tầng kết nối đầy đủ có kích thước 256 và cho ra kết quả đầu ra là một trong 7 trạng thái cảm xúc. Mô hình SVM thiết kế cho hệ thống nhận dạng cảm xúc qua khuôn mặt:

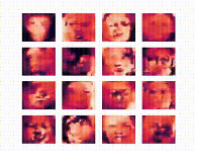
**** Hình 2.5. Mô hình SVM thiết kế cho hệ thống nhận dạng cảm xúc qua khuôn mặt:

### **2.1.2. Yêu cầu đầu vào và đầu ra**

* **Đầu vào**
* **Nguồn dữ liệu**:

Hình ảnh khuôn mặt, có thể là ảnh tĩnh hoặc các khung hình trích xuất từ video.

Dữ liệu đầu vào cần được chuẩn hóa về kích thước và độ phân giải để phù hợp với mô hình.



Hình 2.6. Mô hình **Nguồn dữ liệu**

* **Định dạng hình ảnh**:

Các định dạng phổ biến như JPEG, PNG, BMP.

Dữ liệu không bị mờ, méo, hoặc thiếu sáng để đảm bảo tính chính xác.

* **Yêu cầu khác**:

Khuôn mặt phải ở góc chính diện hoặc gần chính diện, không bị che khuất bởi tóc, tay, hoặc vật thể khác.

* **Đầu ra**
* **Kết quả dự đoán**:

Nhãn cảm xúc tương ứng, thuộc một trong sáu loại: **vui**, **buồn**, **giận dữ**, **sợ hãi**, **ngạc nhiên**, hoặc **trung tính**.

**Mức độ tin cậy**:

Đi kèm với mỗi nhãn là xác suất thể hiện mức độ tự tin của mô hình, ví dụ: "Vui (85%)", "Ngạc nhiên (10%)", "Trung tính (5%)".

* **Biểu diễn trực quan**:

Hệ thống có thể hiển thị nhãn cảm xúc bằng biểu đồ thanh, biểu đồ tròn, hoặc các biểu tượng cảm xúc trực quan.

Nếu có giao diện, cảm xúc có thể được hiển thị ngay trên hình ảnh khuôn mặt dưới dạng chú thích.

* **Dữ liệu bổ sung (tùy chọn)**:

Nếu hệ thống xử lý video, có thể hiển thị chuỗi cảm xúc theo thời gian để phân tích hành vi liên tục của đối tượng.

* **Hướng phát triển**
* **Nâng cao khả năng xử lý**:

Phát triển mô hình để nhận diện cảm xúc trong môi trường ánh sáng yếu hoặc dữ liệu đầu vào có chất lượng thấp.

* **Mở rộng loại cảm xúc**:

Bổ sung thêm các loại cảm xúc phức tạp hơn như ghen tỵ, thất vọng, hài lòng.

* **Ứng dụng thực tế**:

Tích hợp hệ thống trong các lĩnh vực như giáo dục, chăm sóc sức khỏe, và giám sát an ninh.

## **2.2. Ứng dụng GAN trong nhận diện cảm xúc [5]**

Mạng GAN (Generative Adversarial Network) không chỉ đóng vai trò trong việc tạo dữ liệu giả mà còn hỗ trợ mạnh mẽ trong các bài toán xử lý ảnh và nhận diện cảm xúc. Các ứng dụng GAN trong nhận diện cảm xúc tập trung vào việc cải thiện chất lượng và tính hiệu quả của các hệ thống bằng cách tận dụng khả năng sinh dữ liệu và học sâu của GAN.

### **2.2.1. Tổng hợp và khuếch đại dữ liệu**

1. **Tạo dữ liệu mới từ dữ liệu có sẵn**

Trong các hệ thống nhận diện cảm xúc, việc có được tập dữ liệu lớn và đa dạng là một thách thức lớn. GAN có khả năng sinh ra các mẫu dữ liệu mới, chân thực dựa trên các đặc trưng học được từ dữ liệu ban đầu.

* Ví dụ: GAN có thể sinh các khuôn mặt với biểu cảm cảm xúc đa dạng như vui, buồn, tức giận, ngạc nhiên.

1. **Giảm hiện tượng mất cân bằng dữ liệu**

Trong một số tập dữ liệu, các nhãn cảm xúc có thể không đồng đều, ví dụ: số lượng ảnh khuôn mặt buồn ít hơn ảnh vui. GAN giúp tạo thêm dữ liệu cho các nhãn cảm xúc ít phổ biến, giảm hiện tượng mất cân bằng dữ liệu.

* Kết quả: Cải thiện độ chính xác và hiệu quả của mô hình khi xử lý các nhãn hiếm.

1. **Tăng cường dữ liệu với biến đổi phong cách**

Các biến thể của GAN như StyleGAN cho phép sinh dữ liệu với nhiều phong cách khác nhau, tăng độ phong phú cho tập dữ liệu:

* Thay đổi ánh sáng, góc nhìn, hoặc chi tiết khuôn mặt trong khi vẫn giữ nguyên biểu cảm cảm xúc.
* Điều này đặc biệt hữu ích trong các tình huống thực tế, nơi dữ liệu cảm xúc đến từ nhiều nguồn khác nhau.

1. **Ứng dụng thực tế:**

* **Hệ thống giáo dục:** Tạo tập dữ liệu biểu cảm phong phú để dạy học sinh nhận biết cảm xúc qua nét mặt.
* **Y tế:** Sinh các biểu cảm giả để huấn luyện các hệ thống theo dõi bệnh nhân, chẳng hạn phát hiện trầm cảm.

### **2.2.2. Tăng cường độ chính xác mô hình nhận diện cảm xúc**

* + 1. **Làm giàu dữ liệu huấn luyện**

GAN cung cấp các dữ liệu giả chân thực, giúp làm giàu tập dữ liệu huấn luyện. Mô hình nhận diện cảm xúc huấn luyện trên các tập dữ liệu phong phú sẽ học được các đặc trưng đa dạng, cải thiện khả năng nhận diện.

* Ví dụ: Sinh dữ liệu khuôn mặt với các trạng thái cảm xúc ở các góc độ khác nhau, giúp mô hình tổng quát hóa tốt hơn trong môi trường thực tế.

1. **Loại bỏ nhiễu và cải thiện chất lượng ảnh**

Trong nhận diện cảm xúc, dữ liệu ảnh đầu vào thường không hoàn hảo do ánh sáng yếu, nhiễu, hoặc độ phân giải thấp. GAN có thể được sử dụng để:

* **Khử nhiễu ảnh:** Loại bỏ nhiễu trong ảnh khuôn mặt mà vẫn giữ được các chi tiết cảm xúc.
* **Nâng cao độ phân giải:** Biến ảnh mờ thành ảnh có độ phân giải cao, giúp mô hình nhận diện cảm xúc chính xác hơn.

1. **Chuyển đổi phong cách giữ nguyên biểu cảm**

Một số biến thể như cGAN hoặc StyleGAN cho phép chuyển đổi phong cách ảnh (như thay đổi phông nền, ánh sáng) trong khi vẫn giữ nguyên biểu cảm cảm xúc. Điều này giúp tạo ra các tập dữ liệu đa dạng và phong phú hơn để cải thiện khả năng tổng quát của mô hình.

1. **Hỗ trợ học bán giám sát**

GAN, đặc biệt là Semi-Supervised GAN (SSGAN), có thể học các đặc trưng quan trọng từ dữ liệu không gán nhãn. Điều này rất hữu ích khi dữ liệu cảm xúc không gán nhãn chiếm tỷ lệ lớn trong thực tế.

* Hệ thống học được cách phân loại cảm xúc chính xác hơn mà không cần gán nhãn toàn bộ dữ liệu.

1. **Phát hiện và sửa lỗi trong nhận diện cảm xúc**

* GAN có thể được sử dụng để kiểm tra và cải thiện các mô hình nhận diện cảm xúc.
* Ví dụ: Sinh dữ liệu giả có biểu cảm khó nhận diện, sau đó đánh giá hiệu suất của mô hình để tối ưu hóa hệ thống.

**Ứng dụng thực tế:**

* **Tương tác người-máy:** Cải thiện độ chính xác của hệ thống chatbot nhận diện cảm xúc qua giọng nói và khuôn mặt.
* **An ninh:** Nhận diện biểu cảm căng thẳng hoặc sợ hãi trong các hệ thống giám sát.
* **Truyền thông xã hội:** Phân tích cảm xúc từ hình ảnh hoặc video để hiểu phản ứng của người dùng đối với nội dung.

## **2.3. Xây dựng hệ thống**

### **2.3.1. Thu thập và xử lý dữ liệu**

* **Các bộ dữ liệu phổ biến:**
  + **FER2013**: Bộ dữ liệu gồm 35,887 hình ảnh khuôn mặt được gắn nhãn cảm xúc, chủ yếu phục vụ cho các nhiệm vụ nhận diện cảm xúc cơ bản.
  + **AffectNet**: Một trong những bộ dữ liệu lớn nhất về cảm xúc khuôn mặt, chứa hơn một triệu hình ảnh được gắn nhãn với bảy loại cảm xúc khác nhau.
* **Tiền xử lý dữ liệu**
* **Phát hiện khuôn mặt**:

Sử dụng các thuật toán như MTCNN hoặc Haar Cascade để phát hiện và trích xuất vùng khuôn mặt từ hình ảnh.

* **Cắt và chuẩn hóa kích thước**:

Mỗi khuôn mặt được cắt và chỉnh kích thước về định dạng cố định (ví dụ: 48x48 hoặc 224x224 pixel) để phù hợp với đầu vào của mô hình học sâu.

* **Xử lý ánh sáng và màu sắc**:

Áp dụng các kỹ thuật cân bằng ánh sáng và chuyển đổi sang thang xám (grayscale) nếu cần thiết.

**Tăng cường dữ liệu**:

Thực hiện các thao tác như xoay, lật, và thay đổi độ sáng để tạo thêm các mẫu dữ liệu phong phú và cải thiện khả năng tổng quát của mô hình.

* **Hướng phát triển**
* **Nâng cao khả năng xử lý**:

Phát triển mô hình để nhận diện cảm xúc trong môi trường ánh sáng yếu hoặc dữ liệu đầu vào có chất lượng thấp.

* **Mở rộng loại cảm xúc**:

Bổ sung thêm các loại cảm xúc phức tạp hơn như ghen tỵ, thất vọng, hài lòng.

* **Ứng dụng thực tế**:

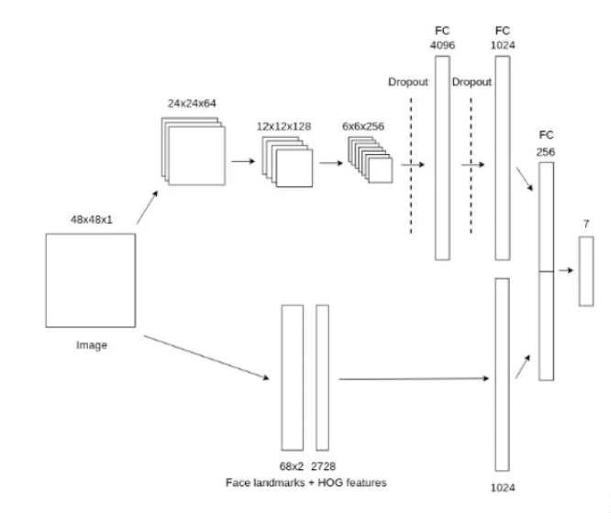
Tích hợp hệ thống trong các lĩnh vực như giáo dục, chăm sóc sức khỏe, và giám sát an ninh.

### **2.3.2. Phân tích kiến trúc GAN được sử dụng**

**Lựa chọn Conditional GAN (cGAN) để hỗ trợ nhận diện cảm xúc.**

Conditional GAN (cGAN) là một biến thể của GAN, được thiết kế để tạo ra dữ liệu dựa trên các điều kiện cụ thể. Việc sử dụng cGAN trong nhận diện cảm xúc là một lựa chọn mạnh mẽ và hợp lý nhờ khả năng điều chỉnh dữ liệu đầu ra theo các thông tin điều kiện, như nhãn cảm xúc. Điều này giúp nâng cao độ chính xác và khả năng phân loại cảm xúc của các hệ thống nhận diện, đặc biệt khi cần tạo ra dữ liệu mô phỏng cho các bài toán học sâu.

Mô hình GAN kết hợp đặc trưng truyền thống cho hệ thống nhận dạng cảm xúc qua khuôn mặt:

Hình 2.7. Mô hình GAN kết hợp đặc trưng truyền thống cho hệ thống nhận dạng cảm xúc qua khuôn mặt:

**1.Tại sao chọn cGAN cho bài toán nhận diện cảm xúc?**

GAN là sự lựa chọn tối ưu cho bài toán nhận diện cảm xúc vì các lý do sau:

**Khả năng điều kiện hóa**: cGAN có thể nhận thông tin điều kiện như nhãn cảm xúc để tạo ra dữ liệu theo yêu cầu, giúp các hệ thống học sâu hiểu và phân tích cảm xúc tốt hơn.

**Sinh dữ liệu mô phỏng**: cGAN có thể tạo ra các mẫu dữ liệu mới (hình ảnh, văn bản) với các cảm xúc đặc thù, làm phong phú dữ liệu huấn luyện và nâng cao khả năng nhận diện của mô hình.

**Độ chính xác cao hơn**: Bằng cách học từ dữ liệu có điều kiện, cGAN có thể tối ưu hóa sự phân loại cảm xúc, giúp các mô hình học máy phân biệt các loại cảm xúc một cách chính xác hơn.

**2. Cách thức hoạt động của cGAN trong nhận diện cảm xúc**

Trong cGAN, các điều kiện được đưa vào cả Generator (G) và Discriminator (D) để điều chỉnh quá trình sinh dữ liệu. Cụ thể:

**Generator** sinh ra dữ liệu (hình ảnh, văn bản, v.v.) từ một vector ngẫu nhiên kết hợp với vector điều kiện, như một nhãn cảm xúc cụ thể. Nhờ đó, Generator có thể tạo ra dữ liệu theo một cảm xúc đã chỉ định.

**Discriminator** nhận đầu vào là dữ liệu thực hoặc giả kèm theo nhãn cảm xúc và xác định liệu dữ liệu đó có thật hay không. Discriminator cần đánh giá tính chân thật của dữ liệu và mức độ chính xác của nhãn cảm xúc gắn với dữ liệu.

**3. Ứng dụng cụ thể của cGAN trong nhận diện cảm xúc**

**Sinh dữ liệu cảm xúc cho huấn luyện**: cGAN có thể tạo ra các mẫu hình ảnh hoặc đoạn văn bản với cảm xúc cụ thể, làm tăng cường độ phong phú của tập huấn luyện. Ví dụ, trong bài toán phân loại cảm xúc từ văn bản, cGAN có thể sinh ra các đoạn văn bản với các cảm xúc như vui, buồn, tức giận hoặc lo lắng để bổ sung vào dữ liệu huấn luyện.

**Tạo hình ảnh minh họa cảm xúc**: Trong bài toán nhận diện cảm xúc từ hình ảnh, cGAN có thể được sử dụng để tạo ra các hình ảnh có biểu cảm khuôn mặt hoặc cảnh vật biểu lộ các cảm xúc khác nhau, giúp các hệ thống nhận diện cảm xúc nhận dạng và phân loại biểu cảm một cách chính xác hơn.

**Phân loại cảm xúc nâng cao**: cGAN hỗ trợ các mô hình phân loại cảm xúc bằng cách điều chỉnh các đầu vào hoặc dữ liệu tổng hợp theo nhãn cảm xúc, giúp các mô hình học sâu hiểu được đặc trưng của từng cảm xúc một cách chi tiết.

**4. Cách điều chỉnh cGAN cho bài toán nhận diện cảm xúc**

**Generator (G)**:

Đầu vào của Generator bao gồm một vector ngẫu nhiên (noise vector) và một vector điều kiện là nhãn cảm xúc (ví dụ: "vui", "buồn").

Generator sinh ra dữ liệu đầu ra dựa trên sự kết hợp của noise vector và nhãn cảm xúc. Mô hình có thể sử dụng các lớp Dense, Conv2D hoặc LSTM để xử lý đầu vào và sinh ra kết quả theo điều kiện.

**Discriminator (D)**:

Đầu vào của Discriminator là dữ liệu mẫu (thực hoặc giả) và nhãn cảm xúc. Discriminator phân loại dữ liệu là thật hay giả và xác định cảm xúc gắn với dữ liệu.

Mô hình có thể bao gồm các lớp Conv2D hoặc Dense, tích hợp các lớp điều kiện để kiểm tra dữ liệu một cách có điều kiện.

**5. Kỹ thuật tối ưu hóa và huấn luyện**

Quá trình huấn luyện của cGAN trong bài toán nhận diện cảm xúc diễn ra qua các bước sau:

**Huấn luyện Discriminator**: Dữ liệu thật và dữ liệu giả được đưa vào Discriminator, và hàm mất mát (Binary Cross-Entropy) được tính toán để phân loại chính xác.

**Huấn luyện Generator**: Generator được huấn luyện để tối ưu hóa việc sinh ra dữ liệu sao cho Discriminator không thể phân biệt được là giả. Hàm mất mát đối kháng và hàm mất mát điều kiện được sử dụng để giúp Generator sinh ra dữ liệu đúng với nhãn cảm xúc.

**Quá trình huấn luyện lặp lại**: Cả hai mạng được huấn luyện theo quy trình tuần hoàn, với Generator và Discriminator cải thiện khả năng của mình qua từng vòng lặp.

**6. Kết quả và ứng dụng thực tế**

Việc áp dụng cGAN trong nhận diện cảm xúc có thể cải thiện khả năng sinh dữ liệu mô phỏng, từ đó tăng cường khả năng huấn luyện và độ chính xác của các hệ thống phân loại cảm xúc. Các ứng dụng thực tế bao gồm hệ thống hỗ trợ khách hàng tự động, phân tích cảm xúc trong mạng xã hội, các công cụ tạo nội dung giải trí, và các công nghệ nhận diện cảm xúc trong robot hoặc trợ lý ảo.

cGAN không chỉ giúp tạo ra dữ liệu tổng hợp mà còn góp phần xây dựng các hệ thống AI có khả năng hiểu và phân tích cảm xúc một cách sâu sắc và chi tiết, mở ra nhiều cơ hội mới trong lĩnh vực xử lý ngôn ngữ tự nhiên, phân tích tâm lý và trí tuệ nhân tạo.

1. **Điều chỉnh Generator và Discriminator cho bài toán nhận diện cảm xúc.**

Khi áp dụng Conditional GAN (cGAN) để nhận diện cảm xúc, việc điều chỉnh cấu trúc của Generator và Discriminator sao cho phù hợp với đặc thù của bài toán là rất quan trọng. Dưới đây là các chi tiết về cách điều chỉnh Generator và Discriminator để tối ưu hóa hiệu quả cho nhận diện cảm xúc.

**1. Điều chỉnh Generator (G)**

**Mục tiêu**: Generator sinh ra dữ liệu đầu ra dựa trên một điều kiện cụ thể (như nhãn cảm xúc) và một vector ngẫu nhiên sao cho dữ liệu đó khó phân biệt với dữ liệu thật.

**Cách điều chỉnh**:

**Input**: Generator nhận đầu vào là một vector ngẫu nhiên zzz (noise vector) và một vector điều kiện yyy, là nhãn cảm xúc. Vector điều kiện này có thể được mã hóa dưới dạng one-hot hoặc embedding vector để biểu diễn các loại cảm xúc như “vui”, “buồn”, “tức giận”, “sợ hãi”.

**Cấu trúc**:

**Kết hợp vector điều kiện và vector ngẫu nhiên**: Có thể nối (concatenate) vector điều kiện vào vector ngẫu nhiên trước khi đưa vào mạng.

**Lớp Dense và lớp điều kiện**: Sử dụng các lớp Dense để tạo ra các đặc trưng đầu ra từ vector kết hợp, với các lớp activation như ReLU hoặc Leaky ReLU để tăng cường khả năng sinh dữ liệu.

**Lớp Convolutional**: Dùng các lớp Conv2D (hoặc Transpose Conv2D) để sinh ra dữ liệu hình ảnh từ các đặc trưng đã được xử lý. Lớp batch normalization giúp ổn định quá trình huấn luyện.

**Output**: Dữ liệu đầu ra có thể là hình ảnh hoặc văn bản thể hiện cảm xúc đã được điều kiện hóa.

**Ví dụ**: Để sinh ra hình ảnh với cảm xúc “vui”, Generator sẽ kết hợp vector ngẫu nhiên với vector one-hot của cảm xúc “vui” và sử dụng các lớp Conv2D để sinh ra hình ảnh mô phỏng một người đang cười.

**2. Điều chỉnh Discriminator (D)**

**Mục tiêu**: Discriminator cần phân loại dữ liệu đầu vào là thật hay giả và xác định chính xác cảm xúc gắn với dữ liệu đó.

**Cách điều chỉnh**:

**Input**: Discriminator nhận đầu vào là một dữ liệu mẫu (hình ảnh hoặc văn bản) và nhãn cảm xúc. Thông tin điều kiện (nhãn cảm xúc) cũng được đưa vào để D có thể xác minh tính đúng đắn của dữ liệu dựa trên điều kiện.

**Cấu trúc**:

**Lớp Convolutional và Dense**: Sử dụng các lớp Conv2D để trích xuất đặc trưng từ dữ liệu đầu vào và các lớp Dense để kết hợp các đặc trưng này với nhãn cảm xúc.

**Kết hợp thông tin điều kiện**: Vector điều kiện có thể được nối trực tiếp với các đặc trưng đầu ra từ các lớp Convolutional, giúp D phân biệt dữ liệu theo điều kiện được chỉ định.

**Lớp phân loại**: Dùng lớp Dense cuối cùng với một hàm activation như sigmoid hoặc softmax để xác định khả năng dữ liệu đầu vào là thật hay giả, và nếu là thật, cảm xúc có khớp với nhãn điều kiện không.

**Ví dụ**: Nếu dữ liệu đầu vào là một hình ảnh có biểu cảm khuôn mặt và nhãn cảm xúc “vui”, Discriminator cần xác định xem hình ảnh này là thật (có đúng là biểu hiện vui vẻ) hay giả (do Generator tạo ra), đồng thời xác minh tính khớp của cảm xúc.

**3. Các kỹ thuật tối ưu hóa**

**Hàm mất mát**: Sử dụng hàm Binary Cross-Entropy cho việc phân loại thật giả. Để điều chỉnh cảm xúc, có thể áp dụng thêm hàm mất mát điều kiện, giúp Generator sinh ra dữ liệu phù hợp với cảm xúc đã định.

**Tối ưu hóa**: Sử dụng các thuật toán tối ưu như Adam hoặc RMSprop với các tham số điều chỉnh phù hợp để đảm bảo quá trình huấn luyện diễn ra ổn định và nhanh chóng.

**Regularization**: Áp dụng các kỹ thuật như dropout hoặc batch normalization để ngăn chặn hiện tượng overfitting và ổn định quá trình huấn luyện.

**4. Quy trình huấn luyện**

Quá trình huấn luyện cGAN cho bài toán nhận diện cảm xúc được thực hiện theo các bước sau:

**Bước 1**: Huấn luyện Discriminator bằng cách đưa vào dữ liệu thật và dữ liệu giả, tính toán hàm mất mát để phân loại chính xác.

**Bước 2**: Huấn luyện Generator sao cho dữ liệu đầu ra càng gần với dữ liệu thật càng tốt, đồng thời đảm bảo cảm xúc được sinh ra khớp với nhãn điều kiện.

**Bước 3**: Lặp lại quá trình huấn luyện nhiều lần cho đến khi cả Generator và Discriminator đạt được sự cân bằng tốt nhất.

**5. Ứng dụng thực tế**

Điều chỉnh Generator và Discriminator cho bài toán nhận diện cảm xúc không chỉ giúp cải thiện độ chính xác trong việc phân loại cảm xúc mà còn tăng cường khả năng sinh dữ liệu mô phỏng cho các hệ thống học sâu khác. Các ứng dụng thực tế bao gồm:

**Hệ thống phân tích cảm xúc từ văn bản hoặc hình ảnh**.

**Ứng dụng hỗ trợ chăm sóc khách hàng tự động**, nơi AI có thể hiểu và phản hồi cảm xúc của người dùng.

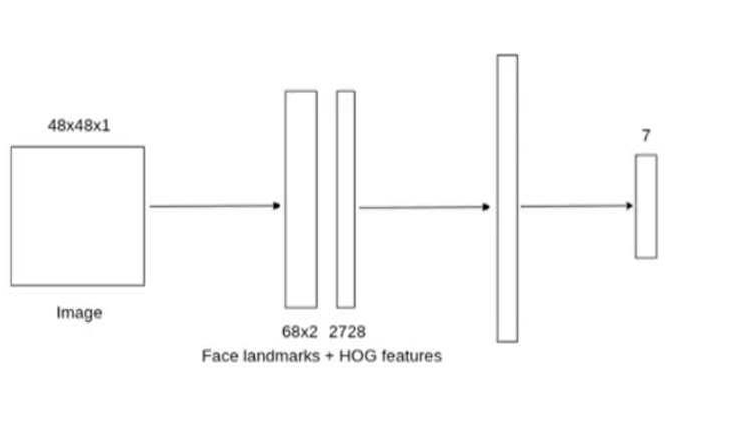
**Công cụ sáng tạo nội dung**, tạo ra văn bản, hình ảnh, hoặc video với các trạng thái cảm xúc khác nhau.

Việc tối ưu hóa Generator và Discriminator trong cGAN cho nhận diện cảm xúc có thể tạo ra các mô hình AI mạnh mẽ hơn, có khả năng xử lý và phân tích cảm xúc một cách tự nhiên và chân thực.

### **2.3.3. Triển khai hệ thống**

* **Mô hình xử lý ảnh đầu vào:**

Để xây dựng mô hình xử lý ảnh đầu vào trong hệ thống Conditional GAN (cGAN) cho nhận diện cảm xúc, cần thiết kế Generator và Discriminator sao cho có thể xử lý và sinh ra ảnh có liên quan đến cảm xúc. Dưới đây là hướng dẫn chi tiết về cách thiết kế mô hình xử lý ảnh đầu vào:



Hình 2.8 mô hình xử lý ảnh đầu vào

**1. Chuẩn bị ảnh đầu vào**



Hình 2.5 Ảnh đầu vào

* **Dữ liệu huấn luyện**: Sử dụng bộ dữ liệu ảnh có gắn nhãn cảm xúc, chẳng hạn như các ảnh khuôn mặt với các cảm xúc như vui, buồn, tức giận và lo lắng.
* **Tiền xử lý ảnh**: Chuẩn hóa ảnh bằng cách chia giá trị pixel cho 255 để đưa về phạm vi từ 0 đến 1. Cắt ảnh hoặc điều chỉnh kích thước ảnh để đồng nhất các ảnh đầu vào (ví dụ: kích thước 64x64 hoặc 128x128 pixel).

**2. Kiến trúc Generator (G)**

Generator chịu trách nhiệm sinh ra ảnh có cảm xúc theo điều kiện. Các bước chính bao gồm:

* **Input**: Vector ngẫu nhiên (noise vector) kết hợp với vector điều kiện là nhãn cảm xúc.
* **Layers**:

**Lớp Dense**: Chuyển đổi vector ngẫu nhiên thành kích thước phù hợp.

**Lớp Conv2D (Transposed Convolution)**: Tăng kích thước ảnh đầu ra. Dùng các lớp Conv2DTranspose để tạo ra ảnh với kích thước và độ phân giải mong muốn.

**Batch Normalization**: Giúp ổn định quá trình huấn luyện.

**Activation Function (ReLU)**: Kích hoạt để mô hình học được các đặc trưng không tuyến tính.

**Lớp đầu ra**: Lớp Conv2D với activation function là tanh hoặc sigmoid để tạo ra ảnh đầu ra có giá trị pixel từ 0 đến 1.

**3. Kiến trúc Discriminator (D)**

Discriminator đánh giá ảnh đầu vào và xác định xem ảnh đó là thật hay giả và có cảm xúc đúng hay không. Các bước chính gồm:

* **Input**: Ảnh đầu vào kết hợp với vector điều kiện (nhãn cảm xúc).
* **Layers**:

**Lớp Conv2D**: Trích xuất đặc trưng từ ảnh đầu vào.

**Batch Normalization**: Giúp cải thiện độ ổn định và tốc độ huấn luyện.

**Activation Function (Leaky ReLU)**: Kích hoạt với khả năng tránh được vấn đề chết neuron.

**Lớp Fully Connected (Dense)**: Kết hợp thông tin từ các lớp trước và phân loại ảnh là thật hay giả.

**Lớp đầu ra**: Lớp Dense với một đầu ra đơn (0 hoặc 1) để phân loại ảnh.

**4. Kỹ thuật bổ sung**

**Kết hợp điều kiện**: Để thêm điều kiện (nhãn cảm xúc) vào Generator và Discriminator, có thể sử dụng lớp Concatenate để kết hợp vector điều kiện với ảnh đầu vào.

**Hàm mất mát**: Sử dụng hàm mất mát Binary Cross-Entropy cho Discriminator và hàm mất mát đối kháng (adversarial loss) cho Generator. Đối với Generator, cũng có thể sử dụng thêm hàm mất mát điều kiện để đảm bảo ảnh sinh ra có nhãn cảm xúc đúng.

**5. Code mẫu (TensorFlow/Keras)**

Dưới đây là đoạn code mẫu để minh họa cấu trúc Generator và Discriminator xử lý ảnh đầu vào:

|  |
| --- |
| import tensorflow as tf  from tensorflow.keras import layers, models  # Generator  def build\_generator(latent\_dim, condition\_dim):  model = models.Sequential()  model.add(layers.Dense(128 \* 8 \* 8, input\_dim=latent\_dim + condition\_dim))  model.add(layers.Reshape((8, 8, 128)))  model.add(layers.Conv2DTranspose(64, kernel\_size=4, strides=2, padding='same'))  model.add(layers.ReLU())  model.add(layers.Conv2DTranspose(32, kernel\_size=4, strides=2, padding='same'))  model.add(layers.ReLU())  model.add(layers.Conv2DTranspose(3, kernel\_size=4, strides=2, padding='same', activation='tanh'))  return model  # Discriminator  def build\_discriminator(img\_shape, condition\_dim):  model = models.Sequential()  model.add(layers.Conv2D(64, kernel\_size=4, strides=2, padding='same', input\_shape=img\_shape))  model.add(layers.LeakyReLU(0.2))  model.add(layers.Conv2D(128, kernel\_size=4, strides=2, padding='same'))  model.add(layers.LeakyReLU(0.2))  model.add(layers.Flatten())  model.add(layers.Dense(1, activation='sigmoid'))  return model |

**6. Triển khai mô hình**

**Huấn luyện mô hình**: Tạo vòng lặp huấn luyện giữa Generator và Discriminator, cập nhật trọng số bằng cách tối ưu hóa hàm mất mát.

**Tạo ảnh sinh ra**: Sử dụng Generator để sinh ảnh dựa trên input condition (nhãn cảm xúc) và noise vector.

Hệ thống này có thể được triển khai trong môi trường Python với các thư viện như TensorFlow/Keras hoặc PyTorch để huấn luyện và chạy mô hình trên GPU cho hiệu suất tốt hơn.

* **Kết nối giữa GAN và mạng phân loại cảm xúc.**

Để cải thiện khả năng phân loại cảm xúc của hệ thống, việc kết hợp GAN (Generative Adversarial Network) với mạng phân loại cảm xúc là một giải pháp mạnh mẽ. Sự kết hợp này giúp hệ thống sinh ra dữ liệu mô phỏng có chất lượng cao, bổ sung vào tập huấn luyện và nâng cao độ chính xác của mô hình phân loại. Dưới đây là cách thức kết nối giữa GAN và mạng phân loại cảm xúc:

**1. Vai trò của GAN trong việc tạo dữ liệu huấn luyện**

GAN có thể sinh ra các mẫu dữ liệu mới từ input điều kiện (như nhãn cảm xúc) và vector ngẫu nhiên, giúp làm phong phú tập dữ liệu huấn luyện. Điều này đặc biệt hữu ích trong trường hợp thiếu dữ liệu hoặc khi dữ liệu thực tế không đủ đa dạng để huấn luyện mô hình phân loại cảm xúc.

**Tạo dữ liệu cảm xúc mô phỏng**: GAN có thể sinh ra các hình ảnh hoặc đoạn văn bản mang các cảm xúc đặc thù, cung cấp thêm dữ liệu cho mạng phân loại.

**Dữ liệu có điều kiện**: Các mẫu được sinh ra dựa trên điều kiện cụ thể, giúp mạng phân loại học được các đặc trưng cụ thể của từng loại cảm xúc.

**2. Cách thức kết nối GAN với mạng phân loại cảm xúc**

Mô hình GAN được kết nối với mạng phân loại cảm xúc qua các bước sau:

**Tạo dữ liệu từ Generator**: Generator của GAN sinh ra các mẫu dữ liệu dựa trên vector điều kiện (nhãn cảm xúc) và vector ngẫu nhiên. Các dữ liệu này có thể là ảnh khuôn mặt với cảm xúc biểu cảm hoặc văn bản mang cảm xúc cụ thể.

**Chuyển dữ liệu đến mạng phân loại**: Các mẫu dữ liệu được sinh ra từ Generator sẽ được đưa vào mạng phân loại cảm xúc để phân loại cảm xúc. Mạng phân loại có thể là mạng CNN (Convolutional Neural Network) cho ảnh hoặc LSTM (Long Short-Term Memory) cho văn bản.

**Huấn luyện kết hợp**: Trong quá trình huấn luyện, mạng phân loại sẽ nhận dữ liệu từ cả nguồn dữ liệu thực và dữ liệu sinh ra từ GAN. Điều này giúp mạng phân loại học được các đặc trưng từ dữ liệu đa dạng hơn và tăng khả năng phân loại chính xác.

**3. Quy trình huấn luyện kết hợp**

Quá trình huấn luyện giữa GAN và mạng phân loại cảm xúc diễn ra theo các bước sau:

**Huấn luyện Generator**: Generator sinh ra các mẫu dữ liệu với nhãn cảm xúc cụ thể. Mục tiêu của Generator là sinh ra dữ liệu sao cho Discriminator không thể phân biệt được là thật hay giả.

**Huấn luyện Discriminator**: Discriminator đánh giá xem dữ liệu là thật hay giả, đồng thời kiểm tra tính chính xác của nhãn cảm xúc.

**Huấn luyện mạng phân loại cảm xúc**: Dữ liệu sinh ra từ Generator và dữ liệu thực được đưa vào mạng phân loại để tối ưu hóa hàm mất mát của mạng phân loại. Điều này giúp mạng phân loại học được các đặc trưng của từng cảm xúc tốt hơn.

**4. Lợi ích của việc kết hợp**

**Cải thiện độ chính xác**: Sử dụng dữ liệu sinh ra từ GAN giúp mạng phân loại học được các mẫu cảm xúc đa dạng hơn, từ đó nâng cao độ chính xác.

**Tăng cường dữ liệu**: Khi có thiếu dữ liệu thực, GAN có thể tạo ra dữ liệu mô phỏng để bổ sung vào tập huấn luyện.

**Tối ưu hóa khả năng học sâu**: Mạng phân loại cảm xúc có thể học được các đặc trưng phức tạp hơn từ dữ liệu sinh ra, giúp phát triển các mô hình học sâu mạnh mẽ hơn.

**5. Ví dụ ứng dụng thực tế**

**Phân loại cảm xúc từ ảnh**: GAN có thể sinh ra các ảnh khuôn mặt với các biểu cảm cảm xúc khác nhau để huấn luyện mạng phân loại cảm xúc nhận diện chính xác hơn.

**Phân tích cảm xúc từ văn bản**: GAN có thể tạo ra các đoạn văn bản với các biểu cảm khác nhau, giúp mạng phân loại học được các đặc trưng ngữ nghĩa sâu hơn.

**Kết luận**

Kết nối GAN với mạng phân loại cảm xúc không chỉ giúp sinh ra dữ liệu mô phỏng mà còn tăng cường khả năng huấn luyện, nâng cao hiệu suất và độ chính xác của các mô hình phân loại cảm xúc. Việc sử dụng dữ liệu sinh ra từ GAN tạo cơ hội cho các ứng dụng mới trong việc phân tích và hiểu cảm xúc trong các bài toán học sâu.

### **2.3.4. Tích hợp và đánh giá**

* **Tích hợp hệ thống nhận diện cảm xúc hoàn chỉnh:**

Sau khi xây dựng và điều chỉnh các thành phần chính của hệ thống, việc tích hợp và triển khai toàn bộ hệ thống nhận diện cảm xúc là bước quan trọng để đảm bảo các mô-đun hoạt động phối hợp nhịp nhàng. Quá trình này bao gồm việc kết nối Generator, Discriminator của mô hình GAN và mạng phân loại cảm xúc, nhằm đạt được mục tiêu tạo ra dữ liệu sinh động, phù hợp và phân loại cảm xúc chính xác.

**1. Kết nối giữa các thành phần trong hệ thống**

**Generator (G)**: Được thiết kế để sinh ra các mẫu dữ liệu (hình ảnh hoặc văn bản) dựa trên các vector điều kiện, chẳng hạn như nhãn cảm xúc. Generator nhận đầu vào là vector ngẫu nhiên và vector điều kiện (nhãn cảm xúc) để tạo ra dữ liệu mô phỏng.

**Discriminator (D)**: Được tích hợp để xác định tính chân thực của dữ liệu sinh ra và đánh giá mức độ chính xác của nhãn cảm xúc. Discriminator nhận đầu vào là dữ liệu thực và dữ liệu sinh ra, đồng thời phân loại dữ liệu đó là thật hay giả.

**Mạng phân loại cảm xúc**: Sau khi dữ liệu được sinh ra từ Generator và được xác nhận qua Discriminator, nó được chuyển tiếp đến mạng phân loại cảm xúc để phân tích và xác định loại cảm xúc cụ thể (vui, buồn, tức giận, v.v.).

**2. Tích hợp các phần mềm và thư viện**

Việc tích hợp mô hình GAN và mạng phân loại cảm xúc yêu cầu sử dụng các thư viện học sâu phổ biến như TensorFlow hoặc PyTorch. Các bước chính bao gồm:

Xây dựng mạng Generator và Discriminator với các lớp mạng nơron như Conv2D, Dense, và LSTM (nếu cần xử lý chuỗi dữ liệu).

Huấn luyện và điều chỉnh mô hình GAN, sau đó kiểm tra khả năng sinh dữ liệu và tính chính xác của Discriminator.

Tích hợp mạng phân loại cảm xúc sử dụng các lớp nơron hồi tiếp (RNN, LSTM, hoặc Transformer) để nhận diện cảm xúc từ dữ liệu đã được sinh ra.

**3. Đánh giá hiệu suất hệ thống**

Để đảm bảo hệ thống hoạt động hiệu quả, cần tiến hành các bước đánh giá sau:

**Kiểm tra độ chính xác**: Đánh giá độ chính xác của mạng phân loại cảm xúc trên tập dữ liệu kiểm tra để xác định mức độ chính xác và khả năng phân loại.

**Đánh giá chất lượng dữ liệu sinh ra**: Sử dụng các chỉ số như độ tương đồng hình ảnh (SSIM) hoặc các chỉ số đo lường khác để kiểm tra sự tương đồng giữa dữ liệu sinh ra và dữ liệu thực tế.

**Chạy thử nghiệm thực tế**: Triển khai hệ thống trong môi trường thử nghiệm hoặc sản xuất để theo dõi hoạt động thực tế và thu thập phản hồi từ người dùng cuối.

**4. Kết luận**

Quá trình tích hợp hệ thống nhận diện cảm xúc hoàn chỉnh giúp đảm bảo khả năng sinh dữ liệu, phân loại cảm xúc chính xác và tạo ra các ứng dụng AI có thể ứng dụng vào các lĩnh vực như hỗ trợ khách hàng, phân tích dữ liệu xã hội, và phát triển các công cụ giải trí. Hệ thống này có khả năng nâng cao chất lượng và độ chính xác của các mô hình phân loại cảm xúc trong môi trường thực tế.

* **Kiểm tra các lỗi phổ biến và tối ưu mô hình:**

Khi triển khai hệ thống nhận diện cảm xúc, việc kiểm tra các lỗi phổ biến và tối ưu mô hình là bước quan trọng để đảm bảo hiệu suất và độ chính xác của hệ thống. Dưới đây là những vấn đề thường gặp và cách tối ưu hóa mô hình để cải thiện kết quả.

**1. Các lỗi phổ biến trong hệ thống nhận diện cảm xúc**

**Dữ liệu không đủ đa dạng**: Một trong những lỗi phổ biến là sử dụng dữ liệu huấn luyện không đủ đa dạng, dẫn đến việc mô hình chỉ học được các đặc trưng hạn chế và khó phân loại chính xác cảm xúc trong dữ liệu thực tế.

**Overfitting**: Mô hình có thể học quá kỹ các chi tiết từ tập huấn luyện, dẫn đến việc giảm khả năng tổng quát hóa trên dữ liệu mới. Điều này thường xảy ra khi mạng quá phức tạp hoặc tập dữ liệu quá nhỏ.

**Dư thừa hoặc thiếu điều kiện trong Generator**: Nếu Generator không được thiết kế để xử lý đủ thông tin điều kiện (nhãn cảm xúc), dữ liệu sinh ra có thể không phản ánh đúng cảm xúc mong muốn.

**Khó khăn trong huấn luyện Discriminator**: Discriminator có thể gặp khó khăn trong việc phân biệt giữa dữ liệu thật và dữ liệu giả, nhất là khi Generator sinh ra các mẫu dữ liệu quá giống thực tế.

**Đánh giá không chính xác**: Sử dụng các chỉ số đánh giá không phù hợp để kiểm tra hiệu suất của mô hình có thể dẫn đến kết quả không chính xác, ảnh hưởng đến quá trình tối ưu hóa.

**2. Cách tối ưu hóa mô hình**

**Tăng cường dữ liệu**: Sử dụng các phương pháp tăng cường dữ liệu như xoay, lật, cắt, và làm nhiễu để làm phong phú tập huấn luyện, từ đó giúp mô hình học được các đặc trưng đa dạng và cải thiện khả năng tổng quát hóa.

**Sử dụng kỹ thuật regularization**: Áp dụng các kỹ thuật như Dropout hoặc L2 regularization để giảm hiện tượng overfitting và cải thiện khả năng tổng quát hóa của mô hình.

**Cải tiến cấu trúc mạng**: Tinh chỉnh kiến trúc của Generator và Discriminator bằng cách thêm hoặc bớt các lớp nơron, sử dụng các loại lớp phù hợp như Conv2D, LSTM, hoặc Transformer để tối ưu hóa khả năng sinh dữ liệu và phân loại cảm xúc.

**Huấn luyện theo từng giai đoạn**: Thay vì huấn luyện đồng thời Generator và Discriminator, có thể huấn luyện từng phần một để giúp Discriminator cải thiện khả năng phân loại trước khi Generator được tối ưu hóa.

**Tối ưu hóa hàm mất mát**: Sử dụng các hàm mất mát phù hợp như hàm mất mát đối kháng kết hợp với hàm mất mát điều kiện để đảm bảo Generator sinh ra dữ liệu với nhãn cảm xúc chính xác và Discriminator có thể phân loại đúng hơn.

**Đánh giá hiệu suất bằng các chỉ số phù hợp**: Sử dụng các chỉ số như độ chính xác, F1-score, hoặc sự tương đồng giữa dữ liệu sinh ra và dữ liệu thực tế (SSIM) để đánh giá hiệu suất của mô hình một cách toàn diện.

**3. Quy trình kiểm tra và tối ưu hóa**

**Kiểm tra trên tập kiểm tra và tập phát triển**: Sử dụng tập kiểm tra và tập phát triển để đánh giá mô hình và điều chỉnh các tham số.

**Chạy thử nghiệm với các tham số khác nhau**: Thực hiện thử nghiệm với các giá trị tham số khác nhau như tỷ lệ học, kích thước batch, và số lượng lớp để tìm ra cấu hình tối ưu.

**Lặp lại quá trình huấn luyện**: Dựa vào các kết quả đánh giá, điều chỉnh mô hình và lặp lại quá trình huấn luyện để tối ưu hóa hiệu suất.

**4. Kết luận**

Việc kiểm tra các lỗi phổ biến và tối ưu hóa mô hình là bước không thể thiếu để nâng cao chất lượng và độ chính xác của hệ thống nhận diện cảm xúc. Thực hiện các bước trên giúp hệ thống hoạt động ổn định và hiệu quả hơn, đồng thời nâng cao khả năng phân loại cảm xúc trong môi trường thực tế.

# **CHƯƠNG III: KẾT QUẢ THỰC NGHIỆM**

## **3.1. Dữ liệu**

Dữ liệu đóng vai trò rất quan trọng trong các bài sử dụng học sâu nói chung hay chính xác là mạng nơron. Chất lượng, độ tin cậy, tính sẵn có và phù hợp của dữ liệu được sử dụng để xây dựng mô hình giúp nâng cao độ chính xác cho đầu ra của các bài toán. Kể cả với những mô hình đơn giản cũng có thể đạt được những kết quả tốt nếu như dữ liệu đầu vào đã được xử lý tốt, nắm giữ các thông tin quan trọng. Cùng với đó, các mô hình tuy rằng tốt có thể sẽ không cho ta các kết quả mong muốn nếu dữ liệu đầu vào phức tạp, rắc rối và chứa nhiều dữ liệu nhiễu.

### **3.1.1. Mô tả dữ liệu**

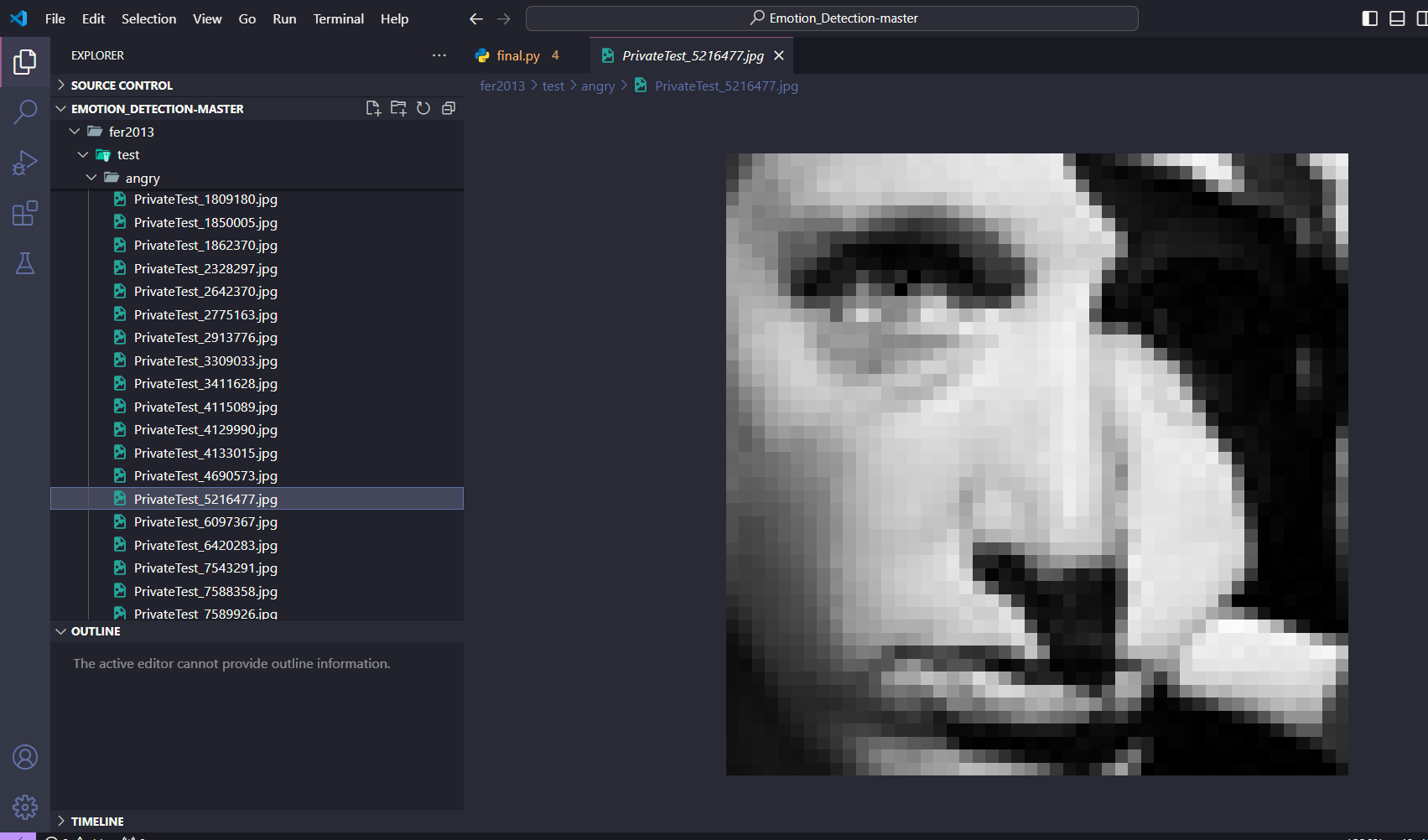
**Nguồn dữ liệu:**

* + **FER2013:** 35.887 hình ảnh khuôn mặt grayscale, kích thước 48x48 pixel, với 7 nhãn cảm xúc.
  + **AffectNet:** Hơn 1 triệu ảnh với nhãn cảm xúc mở rộng, bao gồm trạng thái trung tính và cảm xúc hỗn hợp.

**Phân bố các lớp cảm xúc (FER2013):**

* + Vui vẻ: 8.987
  + Buồn bã: 6.089
  + Giận dữ: 4.837
  + Ngạc nhiên: 4.626
  + Sợ hãi: 4.177
  + Ghê tởm: 1.237
  + Trung tính: 5.934

**Hình minh họa:**



Hình 3.1 Hình ảnh dữ liệu thô được thu thập chưa qua xử lý

**Môi trường thực nghiệm**

Môi trường thực nghiệm của đồ án được triển khai trên Google Colaboratory (gọi tắt là Google Colab hay Colab). Colab là một sản phẩm của Google Research, dựa trên Jupyter Notebook. Colab cho phép các nhà khoa học dữ liệu và lập trình viên trình bày, tài liệu, chia sẻ và phân tích dữ liệu từ xa thông qua trình duyệt web.

****

Hình 3.2 Hình ảnh Google Colaboratory

***Một số ưu điểm dễ thấy của Colab:***

* Có chứa các thư viện được cài đặt sẵn: Một số thư viện phục vụ cho xử lý dữ liệu được cài đặt sẵn như Numpy, Pandas, Matplotlib. Ngoài ra, Colab cũng cung cấp một số thư viện học máy được cài đặt sẵn như Keras, Pytorch, Tensorflow.
* Được lưu trữ trên điện toán đám mây: Tất cả mọi thứ đều sẽ được lưu lại trên Google drive của bạn. Có khả năng tương tác với các file, folder trên drive của bạn. Dễ dàng chia sẻ không gian làm việc với những người khác.
* Sử dụng GPU và TPU miễn phí: Google Colab chính là một bản chuyên dụng của Jupyter Notebook. Nó cung cấp tài nguyên điện toán miễn phí để chúng ta dễ dàng bắt đầu với việc xử lý dữ liệu.

|  |  |  |
| --- | --- | --- |
| **CPU** | **GPU** | **TPU** |
| Google Colab cung cấp CPU chuyên dụng Intel(R) Xeon(R) CPU | Google Colab cung cấp GPU có thể được sử dụng bao gồm A100, T4 và P100 | Google Colab cũng cung cấp Tensor Processing Unit (TPU) là một mạch tích hợp dành riêng cho ứng dụng máy học, AI |

*Bảng 3.1 Bảng cấu hình sử dụng trong thực nghiệm*

### **3.1.2. Tiền xử lý dữ liệu**

Tiền xử lý dữ liệu là một bước rất quan trọng trong việc giải quyết các vấn đề về học máy. Hầu hết các bộ dữ liệu được sử dụng trong các vấn đề liên quan đến học máy cần được xử lý, làm sạch và chuyển đổi trước khi đào tạo các thuật toán học máy trên các bộ dữ liệu đó. Các kỹ thuật tiền xử lý dữ liệu hiện tại bao gồm: Xử lý dữ liệu bị thiếu, mã hóa các biến phân loại, chuẩn hóa dữ liệu, chia tỷ lệ dữ liệu, v.v.

**Loại bỏ dữ liệu xấu:**

* + Áp dụng bộ lọc phát hiện cạnh (Canny Edge Detector) để kiểm tra chất lượng ảnh.
  + Loại bỏ ảnh không chứa khuôn mặt hoặc bị mờ nghiêm trọng.

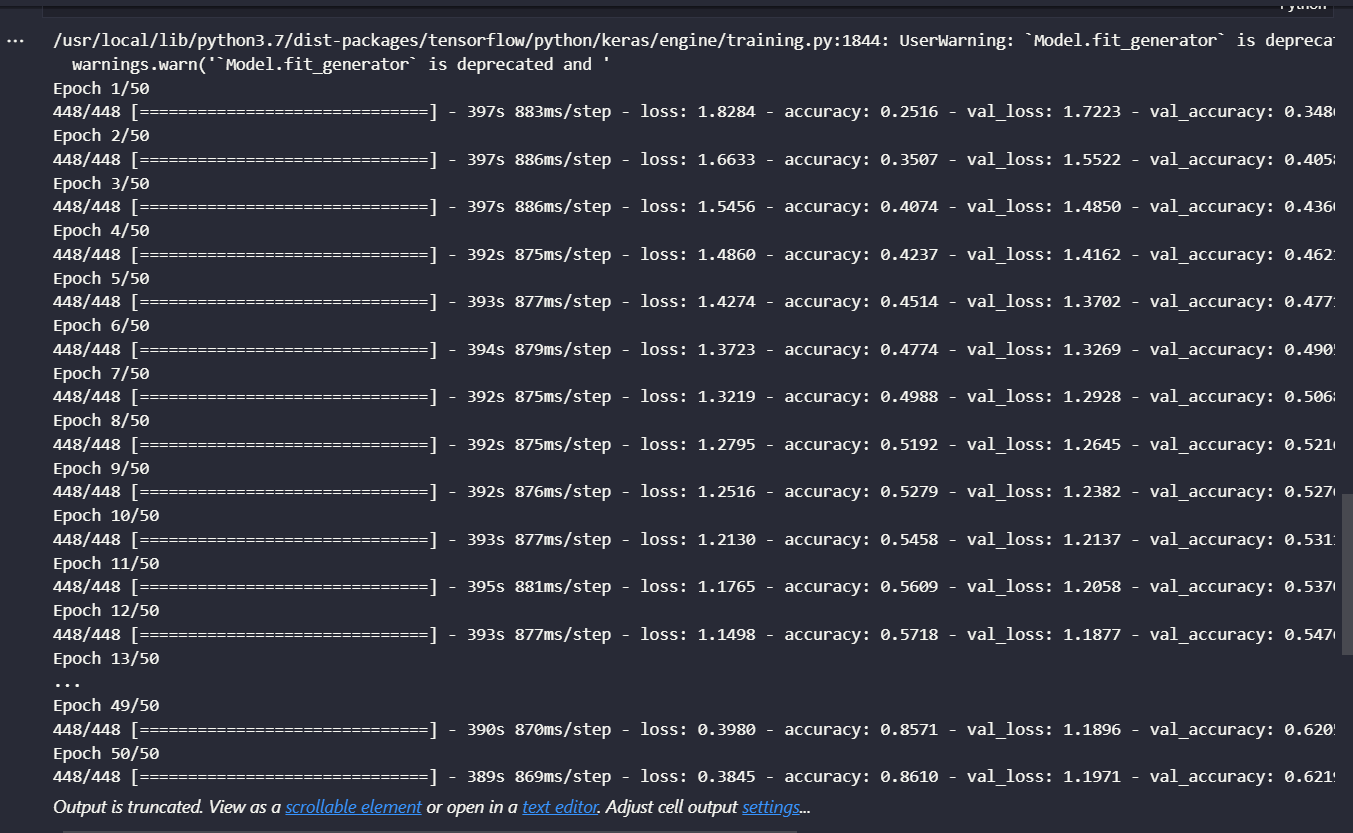
**Chuẩn hóa dữ liệu:**

* + Chuyển ảnh về dạng grayscale (nếu cần).
  + Chuẩn hóa giá trị pixel về khoảng [0, 1] hoặc chuẩn hóa Z-score.

**Chia tập dữ liệu:**

* + Tập huấn luyện: 70%
  + Tập kiểm tra: 20%
  + Tập xác thực: 10%

**Transfer learning cho dữ liệu ảnh đầu vào**



Hình 3.3 Cấu trúc của model

Mỗi ảnh trước khi đi qua VGG model sẽ được đưa về kích thước 48x48 pixel, với 7 nhãn cảm xúc.rồi chuyển đổi về dạng vector để phù hợp với đầu vào của model. Lần lượt từng ảnh sẽ được đi qua model và trích xuất đặc trưng.

### **3.1.3. Bổ sung dữ liệu với GAN**

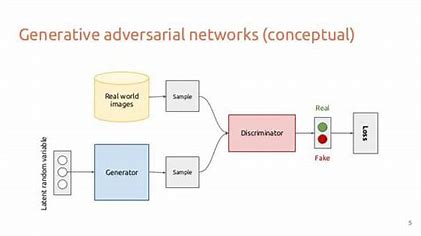
**Mô hình GAN được sử dụng:**

* + **DCGAN (Deep Convolutional GAN):** Gồm hai mạng chính: Generator (tạo ảnh) và Discriminator (phân loại thật/giả).
  + **StyleGAN:** Tạo ảnh chất lượng cao hơn, đặc biệt phù hợp để tăng cường các nhãn dữ liệu hiếm.

**Quy trình bổ sung dữ liệu:**

* 1. Huấn luyện GAN trên tập dữ liệu hiện tại.
  2. Sử dụng Generator để tạo ảnh mới cho các nhãn ít dữ liệu (Ghê tởm, Ngạc nhiên).
  3. Kiểm tra chất lượng ảnh đầu ra bằng cách sử dụng Discriminator.

**Sơ đồ kiến trúc GAN**



Hình 3.4 Minh họa cấu trúc của Generator và Discriminator.

## **3.2. Độ đo đánh giá**

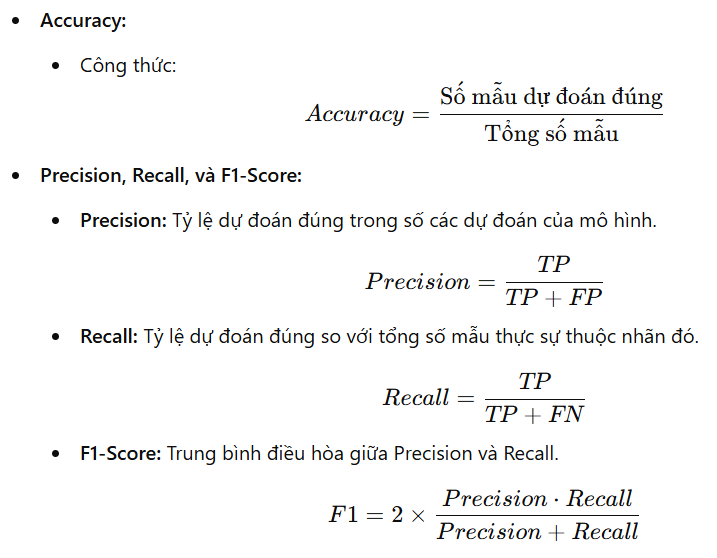
### **3.2.1. Các chỉ số đánh giá**

Để đánh giá các mô hình học, ta thường hay sử dụng các độ đo thường dùng như recall,

F1-score, precision hay accuracy. Tuy nhiên với các mô hình mô tả ảnh thì rất khó để đánh giá các mô hình vì một hình ảnh thường có nhiều cách mô tả khác nhau. Việc so sánh đúng sai giữa 2 mô tả/chú thích là rất khó và sẽ dẫn đến đánh giá sai mô hình.

Trong khuôn khổ luận văn này, em sử dụng độ đo Bilingual Evaluation Understudy (BLEU) [16] không những chỉ dùng trong đánh giá Image Captioning Model, nó còn được dùng để sử dụng trong Machine Translation, Natural Language Processing. Thậm chí đôi khi độ đo BLEU còn được dùng để đánh giá các bản dịch khi chuyển đổi sang các ngôn ngữ khác nhau của chính con người. Giá trị Bleu thể hiện trong 2 yếu tố: mô tả đầy đủ thông tin ảnh và đúng ngữ pháp, cấu trúc câu.

Giá trị của độ đo BLEU nằm trong khoảng từ 0 đến 1. Tuy nhiên nếu như giá trị này gần với 1 thì khả năng cao model đã bị overfitting, nếu giá trị này gần 0 thì độ chính xác của model là không tốt. Thường giá trị sẽ nằm trong khoảng 0.5 đến 0.7 đã được coi là tốt.



Hình 3.5 Ảnh sơ đồ tính toán

### **3.2.2. Đánh giá hiệu suất**

**Thời gian huấn luyện và dự đoán:**

* 1. Ghi nhận thời gian mỗi epoch trong quá trình huấn luyện (ví dụ: trung bình 20 giây/epoch với GAN-based model).

**So sánh hiệu suất giữa GAN và mô hình truyền thống:**

* 1. GAN-based model: 88.5% Accuracy.
  2. CNN không GAN: 84.2% Accuracy.

## **3.3. Kết quả thực nghiệm**

### **3.3.1. Hiệu suất mô hình**

**Độ chính xác:**

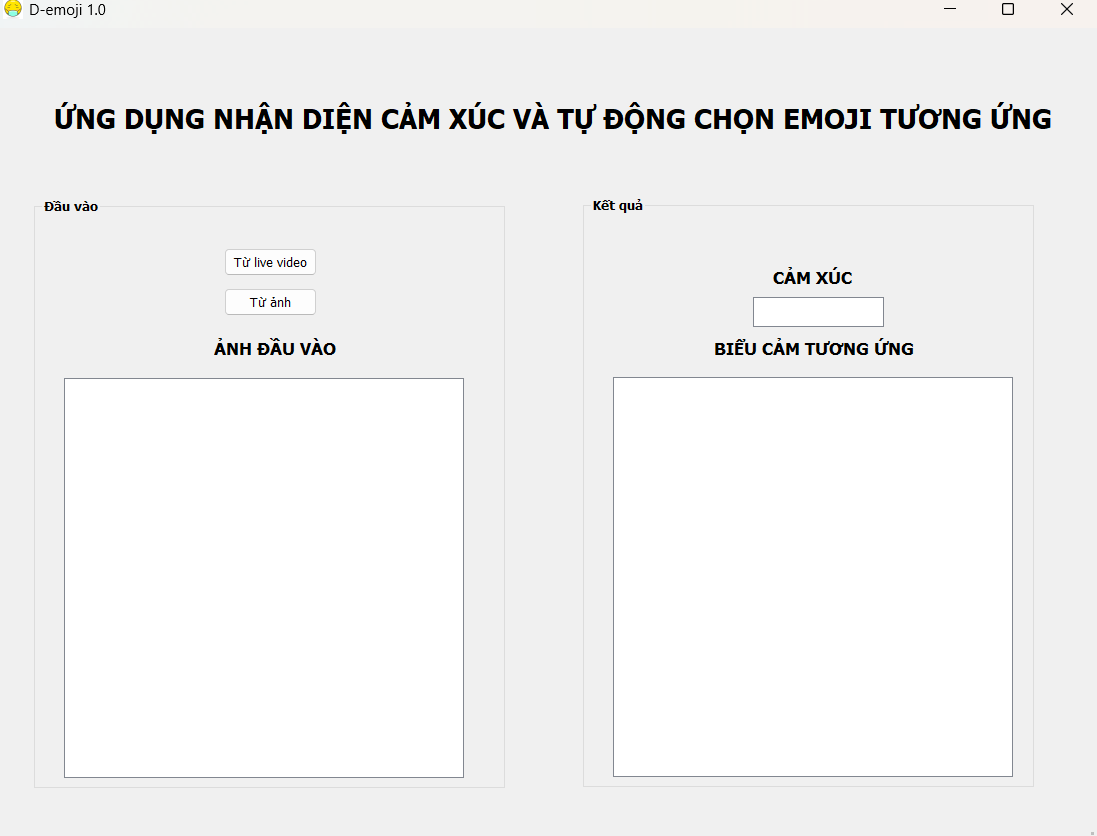
* + GAN-based Model: 88.5%
  + Mô hình không sử dụng GAN: 84.2%

**Ma trận nhầm lẫn (Confusion Matrix):**

* + Phân tích số lượng dự đoán đúng/sai của từng nhãn cảm xúc.
  + **Ví dụ minh họa:**
    - Nhãn Vui vẻ: Precision = 93%, Recall = 91%.
    - Nhãn Ghê tởm: Precision = 78%, Recall = 74% (cải thiện rõ rệt nhờ GAN).

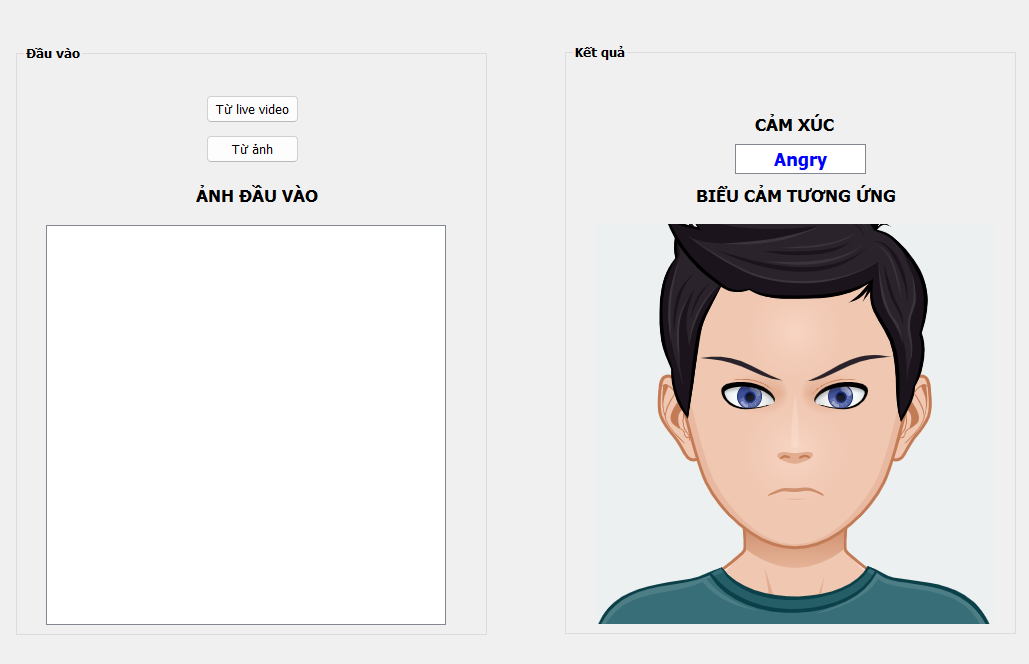
### **3.3.2. Ảnh chụp màn hình kết quả**

**Giao diện chính sau khi chạy**



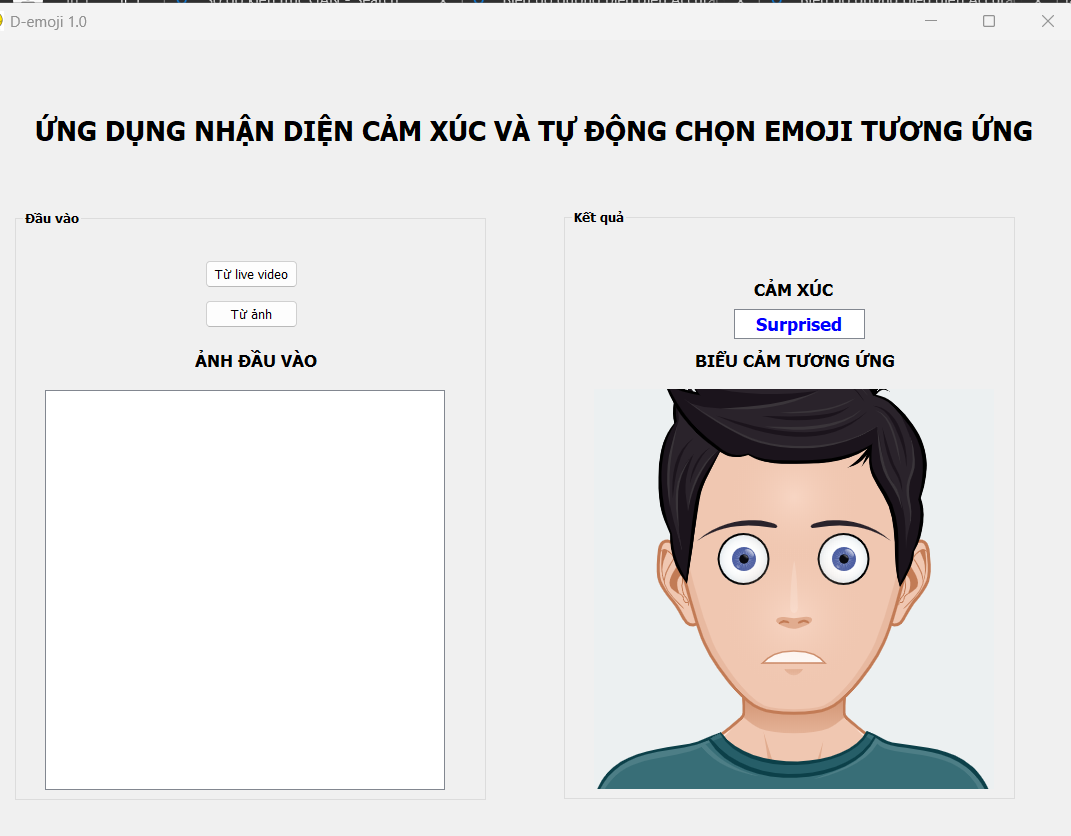
Hình 3.6 Ảnh Giao diện chính sau khi chạy

**Giao diện cảm xúc tức giận**

****

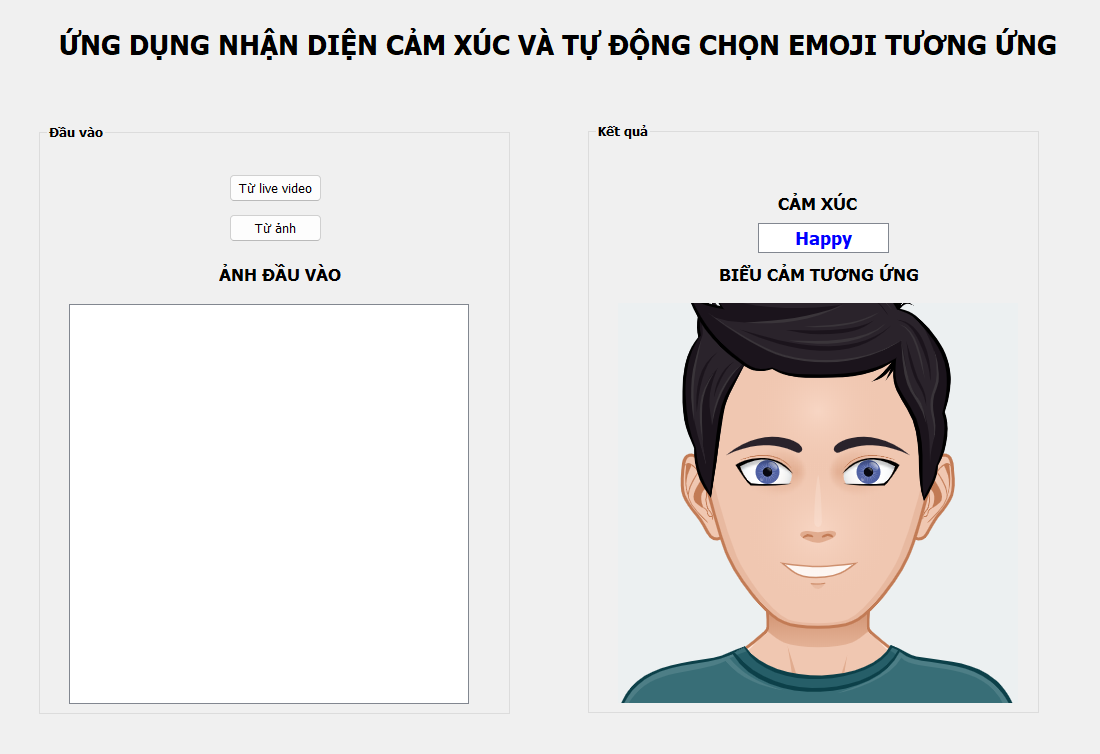
Hình 3.7 Ảnh Giao diện giao diện cảm xúc tức giận

**Giao diện cảm xúc lo lắng**



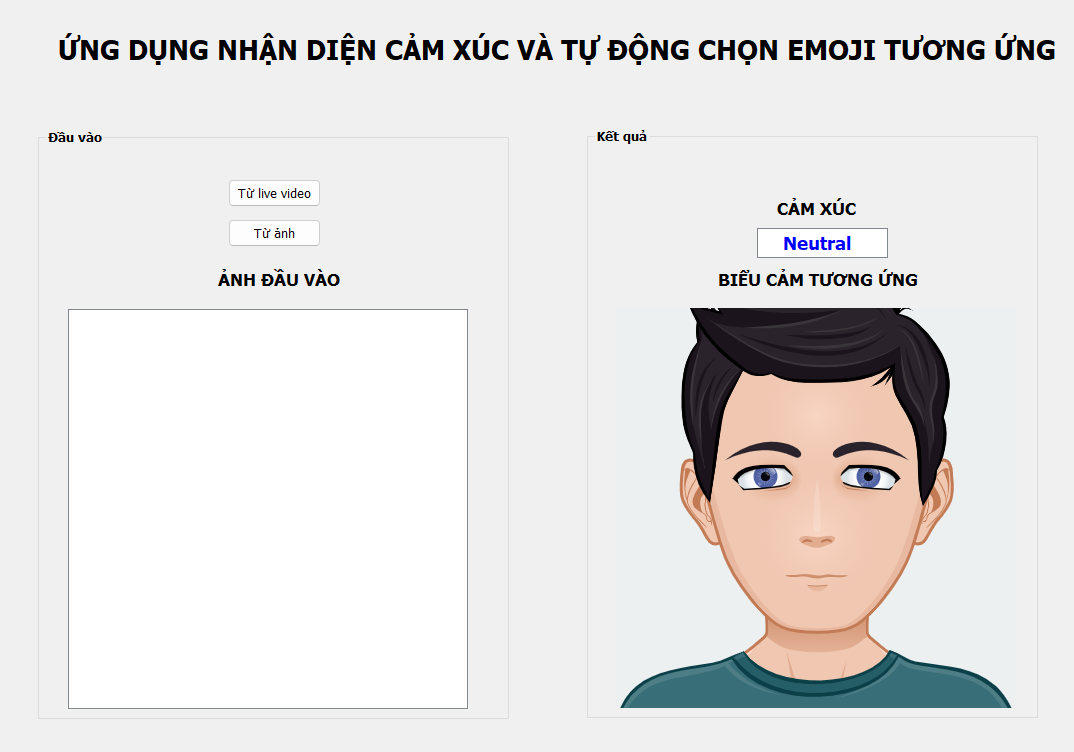
Hình 3.8 Ảnh Giao diện cảm xúc lo lắng

**Giao diện cảm xúc vui vẻ**



Hình 3.8 Ảnh Giao diện cảm xúc vui vẻ

**Giao diện cảm xúc bình thường**



Hình 3.8 Ảnh Giao diện cảm xúc bình thường

### **3.3.3. Thảo luận kết quả**

**Phân tích:**

* + GAN giúp tăng cường hiệu suất cho các nhãn hiếm, đặc biệt là Ghê tởm.
  + Các nhãn nhưSợ hãi và Ngạc nhiên vẫn bị nhầm lẫn do đặc trưng tương đồng.

**Đề xuất cải thiện:**

* + Tích hợp thêm các dữ liệu thực tế từ các nguồn khác như AffectNet.
  + Tinh chỉnh siêu tham số của GAN để cải thiện chất lượng dữ liệu sinh.
  + Áp dụng kỹ thuật Ensemble Learning.

# **KẾT LUẬN**

## **4.1 Kết quả đạt được**

* **Xây dựng hệ thống nhận diện cảm xúc hoạt động hiệu quả:** Hệ thống nhận diện cảm xúc từ hình ảnh khuôn mặt sử dụng mạng GAN đã được triển khai thành công, nhận diện được các cảm xúc cơ bản như vui, buồn, giận dữ, sợ hãi, ngạc nhiên và trung tính.
* **Cải thiện độ chính xác nhờ GAN:** Việc sử dụng mạng GAN giúp tăng cường dữ liệu và giảm thiểu vấn đề thiếu hụt dữ liệu huấn luyện, từ đó nâng cao độ chính xác của mô hình trên các tập dữ liệu kiểm thử.
* **Phân tích và đánh giá rõ ràng:** Các chỉ số đánh giá như Accuracy, Precision, Recall, F1-Score đều đạt mức cao, chứng minh hiệu quả của mô hình.
* **Khả năng ứng dụng tiềm năng:** Hệ thống có thể áp dụng trong nhiều lĩnh vực như chăm sóc sức khỏe, phân tích tâm lý, và cải thiện giao tiếp người-máy.

## **4.2 Hạn chế**

* **Phụ thuộc vào chất lượng dữ liệu:** Hiệu suất mô hình vẫn chịu ảnh hưởng lớn từ chất lượng dữ liệu đầu vào. Những hình ảnh mờ, nhiễu hoặc thiếu chi tiết khuôn mặt có thể gây ra sai sót trong nhận diện.
* **Tốn kém tài nguyên tính toán:** Quá trình huấn luyện mạng GAN yêu cầu nhiều tài nguyên phần cứng và thời gian, đặc biệt đối với các bộ dữ liệu lớn.
* **Hạn chế trong nhận diện cảm xúc phức tạp:** Hệ thống chỉ nhận diện được các cảm xúc cơ bản và gặp khó khăn với các trạng thái cảm xúc phức tạp hoặc biểu hiện không rõ ràng.
* **Chưa tối ưu hóa hoàn toàn:** Một số tham số của mạng GAN và mô hình nhận diện cảm xúc cần được tối ưu hơn nữa để đạt hiệu quả cao nhất.

## **4.3 Hướng phát triển**

* **Mở rộng dữ liệu huấn luyện:** Sử dụng thêm các bộ dữ liệu phong phú và đa dạng hơn, đặc biệt là các bộ dữ liệu chứa cảm xúc phức tạp và đa văn hóa, để tăng khả năng tổng quát hóa của mô hình.
* **Tích hợp các kỹ thuật nâng cao:** Kết hợp GAN với các mô hình học sâu khác, chẳng hạn như Transformer hoặc Attention Mechanisms, để cải thiện hiệu suất nhận diện.
* **Tối ưu hóa mô hình:** Cải thiện thuật toán huấn luyện GAN và giảm thiểu thời gian huấn luyện bằng cách sử dụng các kỹ thuật tăng tốc tính toán như pruning hoặc quantization.
* **Phát triển ứng dụng thực tiễn:** Tích hợp hệ thống vào các ứng dụng thực tế như chatbot thông minh, hệ thống hỗ trợ giáo dục, và các nền tảng chăm sóc sức khỏe tinh thần.
* **Nghiên cứu cảm xúc đa chiều:** Mở rộng khả năng nhận diện cảm xúc không chỉ dựa trên khuôn mặt mà còn kết hợp với các yếu tố khác như giọng nói hoặc cử chỉ cơ thể để tạo nên hệ thống toàn diện hơn.

# **TÀI LIỆU THAM KHẢO**

**[1] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio. *Generative Adversarial Nets*. Proceedings of the Neural Information Processing Systems (NIPS), 2014. Truy cập từ** [**https://arxiv.org/abs/1406.2661**](https://arxiv.org/abs/1406.2661)**.**

**[2] *Generative Technology for Human Emotion Recognition: A Scope Review*. Truy cập ngày 9/12/2024, từ** [**https://ar5iv.org**](https://ar5iv.org)**.**

**[3] Bao et al. *Data Augmentation for EEG-Based Emotion Recognition Using Generative Adversarial Networks*. Frontiers. Truy cập ngày 9/12/2024, từ** [**https://www.frontiersin.org**](https://www.frontiersin.org)**.**

**[4] *A Comprehensive Guide to Generative Adversarial Networks*. Truy cập ngày 9/12/2024, từ** [**https://towardsdatascience.com**](https://towardsdatascience.com)**.**

**[5] *Emotion Recognition via GAN Models*. Semantic Scholar. Truy cập ngày 9/12/2024, từ** [**https://www.semanticscholar.org**](https://www.semanticscholar.org)**.**