ĐẠI HỌC CÔNG NGHỆ THÔNG TIN
ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH

BÁO CÁO CUỐI KỲ

CÁC VẤN ĐỀ CHỌN LỌC TRONG THỊ GIÁC MÁY TÍNH

Lớp CS420.P11

ĐỀ TÀI: **Phát hiện cảm xúc bằng CNN**

Giảng viên	TS. Mai Tiến Dũng ThS. Đỗ Văn Tiến
	Phạm Duy Long 20521573

Chương 1. Giới thiệu

Lý do chọn đề tài:

Cảm xúc đóng một vai trò quan trọng trong tương tác của con người và có thể ảnh hưởng lớn đến hành vi và việc ra quyết định của chúng ta. Hiểu và giải thích cảm xúc có thể là một thách thức nhưng cần thiết cho các lĩnh vực khác nhau như tâm lý học, nghiên cứu thị trường và tương tác giữa người và máy tính. Trong những năm gần đây, ngày càng có nhiều mối quan tâm đến việc sử dụng các kỹ thuật học máy để tự động phát hiện và phân loại cảm xúc. Một trong những kỹ thuật mạnh mẽ như vậy là Mạng nơ-ron tích chập (CNN). Trong bài viết này, chúng ta sẽ khám phá khái niệm phát hiện cảm xúc bằng CNN, các ứng dụng, thách thức và triển vọng trong tương lai của nó.

Giới thiệu:

Phát hiện cảm xúc bằng mạng nơ-ron tích chập (CNN) được chọn làm chủ đề nghiên cứu vì đây là một lĩnh vực quan trọng và đầy tiềm năng trong trí tuệ nhân tạo và thị giác máy tính. Cảm xúc là một yếu tố then chốt trong giao tiếp giữa con người, và khả năng nhận diện cảm xúc từ khuôn mặt không chỉ hỗ trợ trong các ứng dụng công nghệ mà còn mở ra cơ hội cải thiện nhiều lĩnh vực khác nhau.

Ứng dụng của phát hiện cảm xúc rất đa dạng, từ việc nâng cao trải nghiệm người dùng trong các hệ thống thông minh, cải thiện giáo dục qua nhận diện trạng thái học tập của học sinh, đến hỗ trợ y tế trong việc phát hiện các vấn đề về tâm lý. CNN là một phương pháp lý tưởng để xử lý bài toán này vì khả năng học tự động các đặc trưng từ hình ảnh, thay vì dựa trên các đặc trưng được thiết kế thủ công. Với khả năng học sâu từ dữ liệu lớn và hiệu quả vượt trội trong xử lý hình ảnh, CNN đã chứng minh là một công cụ mạnh mẽ để nhận diện cảm xúc từ hình ảnh khuôn mặt.

Việc áp dụng CNN vào phát hiện cảm xúc không chỉ giải quyết các bài toán khoa học mà còn đóng góp ý nghĩa thực tiễn lớn, đặc biệt trong kỷ nguyên công nghệ 4.0, nơi trí tuệ nhân tạo đóng vai trò ngày càng quan trọng trong đời sống.

Chương 2: Tiền Xử Lý Dữ Liệu

Thu thập dữ liệu:

Dữ liệu được thu thập từ kaggle có tên là "FER-2013" và được lưu trong file zip.

Chi tiết:

Dữ liệu bao gồm các hình ảnh khuôn mặt dạng thang độ xám kích thước 48x48 pixel. Các khuôn mặt đã được căn chỉnh tự động sao cho khuôn mặt nằm ở trung tâm và chiếm một lượng không gian gần như tương tự trong mỗi hình ảnh.

Nhiệm vụ là phân loại mỗi khuôn mặt dựa trên cảm xúc được thể hiện trong biểu cảm khuôn mặt thành một trong bảy loại sau:
(0 = Giận dữ, 1 = Ghê tởm, 2 = Sợ hãi, 3 = Vui vẻ, 4 = Buồn bã, 5 = Ngạc nhiên, 6 = Bình thản).

Tập dữ liệu huấn luyện bao gồm 28,709 mẫu và tập kiểm tra công khai bao gồm 3,589 mẫu.
Linhk dataset: FER-2013

Lý do chọn bộ dữ liệu này là vì Bộ dữ liệu có kích thước nhỏ chỉ 53.89 MB với hình ảnh grayscale 48x48 pixel. Điều này giúp dễ dàng tải xuống và sử dụng trong các dự án học máy.

Dữ liệu được sử dụng để huấn luyện mô hình phát hiện cảm xúc và sau khi huấn luyện, mô hình sẽ được áp dụng cho các dự đoán trong thời gian thực. Mô hình này cho phép nhận diện cảm xúc từ video và nguồn dữ liệu trực tiếp.

Các thư viện cần thiết cho dự án được xác định trong tệp requirements.txt. Điều này giúp người dùng biết được những thư viện nào cần cài đặt để chạy mã.

Tiền xử lý dữ liệu:

```
# Initialize image data generator with rescaling
train_data_gen = ImageDataGenerator(rescale=1./255)
validation_data_gen = ImageDataGenerator(rescale=1./255)

# Preprocess all test images
train_generator = train_data_gen.flow_from_directory(
        'data/train',
        target_size=(48, 48),
        batch_size=64,
        color_mode="grayscale",
        class_mode='categorical')

# Preprocess all train images
validation_generator = validation_data_gen.flow_from_directory(
        'data/test',
        target_size=(48, 48),
        batch_size=64,
        color_mode="grayscale",
        class_mode='categorical')
```

Rescaling (rescale=1./255): Mọi giá trị pixel được chuẩn hóa về khoảng [0, 1] (vốn có giá trị từ 0-255). Điều này giúp mô hình học tốt hơn.

flow_from_directory():

- Đọc ảnh từ các thư mục data/train và data/test.
- **target_size=(48, 48)**: Resize tất cả ảnh về kích thước 48x48 (đã biết dữ liệu đầu vào là 48x48).
- **color_mode="grayscale"**: Chuyển đổi ảnh thành dạng thang độ xám.
- **batch_size=64**: Xử lý từng lô 64 ảnh trong mỗi lần huấn luyện.
- **class_mode='categorical'**: Chuyển nhãn của các ảnh thành dạng one-hot vector (do bài toán có 7 nhãn cảm xúc).

đọc ảnh, chuẩn hóa và chuẩn bị batch dữ liệu để huấn luyện.

Chương 3. Xây dựng cấu trúc mô hình CNN.

Mô hình CNN này bắt đầu với lớp tích chập 2D có 32 bộ lọc và kích thước hạt nhân 3x3. Hình ảnh đầu vào được chuẩn bị với kích thước 48x48 và mã màu xám.

```
emotion_model = Sequential()

emotion_model.add(Conv2D(32, kernel_size=(3, 3), activation='relu', input_shape=(48, 48, 1)))
```

emotion_model.add(Conv2D(...)):

- **32**: Số lượng bộ lọc trong lớp tích chập đầu tiên. Mỗi bộ lọc sẽ học một đặc trưng khác nhau từ ảnh.
- **kernel_size=(3, 3)**: Kích thước hạt nhân (kernel) là 3x3, tức mỗi bộ lọc sẽ trượt trên ảnh theo các vùng 3x3 để tính toán đặc trưng.
- **activation='relu'**: Hàm kích hoạt ReLU, giúp tăng tính phi tuyến và giảm vấn đề gradient biến mất.
- **input_shape=(48, 48, 1)**: Kích thước đầu vào của ảnh:
  - 48x48: Kích thước không gian của ảnh (rộng x cao).
  - 1: Kênh màu, biểu thị ảnh thang độ xám (grayscale).

Ý nghĩa:

- **Ảnh đầu vào** đã được chuẩn bị ở giai đoạn tiền xử lý (chuẩn hóa và chuyển đổi ảnh sang thang độ xám), đảm bảo kích thước 48x48 và chỉ có một kênh màu.
- Lớp tích chập đầu tiên này đóng vai trò học các đặc trưng cơ bản như cạnh, góc, hoặc hoa văn đơn giản từ ảnh.

Để giảm thiểu hiện tượng overfitting, các lớp dropout được thêm vào sau các lớp tích chập. Điều này giúp cải thiện độ chính xác của mô hình trong quá trình huấn luyện.

1. Dropout sau lớp tích chập và pooling đầu tiên:

```
emotion_model.add(Dropout(0.25))
```

Vị trí: Sau các lớp:
Conv2D(32): Lớp tích chập đầu tiên.
Conv2D(64): Lớp tích chập tiếp theo.
MaxPooling2D(pool_size=(2, 2)): Lớp gộp tối đa (max pooling).
Ý nghĩa:
Giảm ngẫu nhiên 25% (tỷ lệ 0.25) số lượng kết nối giữa các tầng trong mạng.
Giúp mô hình tránh phụ thuộc quá mức vào một tập hợp trọng số cụ thể, cải thiện khả năng khái quát hóa (generalization).

2. Dropout sau lớp tích chập và pooling thứ hai:

```
emotion_model.add(Dropout(0.25))
```

Vị trí: Sau các lớp:

- **Conv2D(128)**: Hai lớp tích chập kế tiếp.
- **MaxPooling2D(pool_size=(2, 2))**: Lớp gộp tối đa thứ hai.

Ý nghĩa: Tương tự, giảm **25%** kết nối giữa các tầng để cải thiện khả năng khái quát hóa.

3. Dropout sau lớp Dense (Fully Connected):

```
emotion_model.add(Dropout(0.5))
```

Ý nghĩa:

- Giảm ngẫu nhiên **50%** (tỷ lệ 0.5) số kết nối trong lớp fully connected.
- Lớp này chứa số lượng lớn tham số, dễ gây ra hiện tượng **overfitting**, nên dropout đặc biệt quan trọng.

Mô hình được biên dịch với hàm mất mát categorical crossentropy và bộ tối ưu hóa Adam. Các tham số huấn luyện được điều chỉnh để đảm bảo hiệu suất tối ưu cho mạng nơ-ron.

```
emotion_model.compile(loss='categorical_crossentropy', optimizer=Adam(lr=0.0001, decay=1e-6), metrics=['accuracy'])
```

loss='categorical_crossentropy':

- Đây là hàm mất mát (loss function) được sử dụng để huấn luyện mô hình.
- **Lý do chọn:**
  - Dữ liệu đầu ra có dạng **one-hot encoding** (với 7 nhãn cảm xúc), phù hợp với categorical_crossentropy.
  - Hàm này tính toán mức độ sai lệch giữa dự đoán của mô hình và nhãn thực tế.

optimizer=Adam(...):

- **Adam (Adaptive Moment Estimation):** Một bộ tối ưu hóa hiện đại được sử dụng rộng rãi vì:
  - Tự động điều chỉnh tốc độ học (learning rate) trong quá trình huấn luyện.
  - Kết hợp ưu điểm của SGD (Gradient Descent) và RMSProp.
- **Các tham số điều chỉnh:**
  - lr=0.0001: Tốc độ học ban đầu là 0.0001, giá trị nhỏ giúp đảm bảo mạng hội tụ dần dần.
  - decay=1e-6: Giảm tốc độ học theo thời gian, giúp tránh hiện tượng dao động khi tiến gần tới giá trị tối ưu.

metrics=['accuracy']:

- Đo lường **độ chính xác** (accuracy) trong quá trình huấn luyện và kiểm tra.
- Giúp theo dõi hiệu suất mô hình trên cả dữ liệu huấn luyện và kiểm tra.

Việc lưu trữ cấu trúc mô hình trong tệp json và trọng số trong tệp h5 là cần thiết để tái sử dụng mô hình đã huấn luyện. Điều này giúp tiết kiệm thời gian và tài nguyên cho các lần huấn luyện sau.

```
emotion_model.save_weights('emotion_model.h5')
```

Chương 4. Phát hiện cảm xúc qua video

Quá trình phát hiện cảm xúc từ video bao gồm hai bước chính: phát hiện khuôn mặt và sau đó phân tích cảm xúc từ khuôn mặt đó. Việc này cần sử dụng một số công cụ và cấu hình đặc biệt để thực hiện hiệu quả.

```
face_detector = cv2.CascadeClassifier('haarcascades/haarcascade_frontalface_default.xml')
```

Sau khi phát hiện khuôn mặt, chúng ta cần chuyển đổi hình ảnh thành grayscale. Điều này giúp mô hình phát hiện cảm xúc hoạt động chính xác hơn với dữ liệu đã được huấn luyện.

```
gray_frame = cv2.cvtColor(frame, cv2.COLOR_BGR2GRAY)
```

Giải thích:

- **cv2.cvtColor()**:
  - Đây là hàm của OpenCV dùng để chuyển đổi không gian màu của hình ảnh.
  - Tham số cv2.COLOR_BGR2GRAY chỉ định rằng chúng ta đang chuyển đổi từ không gian màu **BGR** (chuẩn mặc định của OpenCV) sang ảnh **grayscale**.
- **Tại sao cần chuyển đổi sang grayscale?**
  - Dữ liệu dùng để huấn luyện mô hình nhận diện cảm xúc là các hình ảnh grayscale (48x48).
  - Ảnh grayscale giảm bớt thông tin không cần thiết (như màu sắc), giúp mô hình tập trung vào đặc trưng về hình dạng và độ sáng, vốn quan trọng hơn cho việc phát hiện cảm xúc.
  - Điều này cũng giúp giảm kích thước dữ liệu đầu vào và tăng tốc độ xử lý.

Cuối cùng, kích thước hình ảnh khuôn mặt cần được điều chỉnh trước khi đưa vào mô hình. Hình ảnh được chuyển đổi thành kích thước 48x48 pixel để phù hợp với yêu cầu của mô hình phát hiện cảm xúc.

```
cropped_img = np.expand_dims(np.expand_dims(cv2.resize(roi_gray_frame, (48, 48)), -1), 0)
```

Giải thích:

- **cv2.resize()**:
  - Đây là hàm OpenCV dùng để thay đổi kích thước của hình ảnh.
  - Ở đây, **(48, 48)** chỉ định kích thước mong muốn của hình ảnh là 48x48 pixel, đúng với yêu cầu của mô hình đã huấn luyện.
- **np.expand_dims()**:
  - Sau khi thay đổi kích thước, hình ảnh có thể có dạng (48, 48). Tuy nhiên, mô hình yêu cầu hình ảnh đầu vào có ba chiều (batch size, chiều cao, chiều rộng, số kênh màu). Vì vậy, chúng ta cần thêm hai chiều phụ: một cho **batch size** (số lượng hình ảnh) và một cho **số kênh màu** (ở đây là 1 cho ảnh grayscale).

    - np.expand_dims(cropped_img, -1) thêm chiều kênh màu (1) vào hình ảnh.
    - np.expand_dims(..., 0) thêm chiều batch size (1) vào, làm cho ảnh có dạng (1, 48, 48, 1).

Lý do cần thay đổi kích thước:

- **Mô hình đã huấn luyện** yêu cầu ảnh đầu vào có kích thước cụ thể là **48x48 pixel** (vì trong quá trình huấn luyện, các ảnh cũng có kích thước này).
- **Ảnh gốc có thể có kích thước lớn hơn**, và cần phải thay đổi kích thước sao cho đồng nhất với kích thước mà mô hình mong đợi để đảm bảo tính nhất quán và khả năng dự đoán chính xác.

Chương 5. Kết luận

Các ứng dụng của phát hiện cảm xúc bằng CNN

Phát hiện cảm xúc bằng CNN đã tìm thấy các ứng dụng trong nhiều lĩnh vực khác nhau. Một ứng dụng nổi bật là phân tích biểu cảm khuôn mặt, nơi CNN có thể xác định chính xác cảm xúc từ hình ảnh khuôn mặt hoặc video. Công nghệ này có thể được sử dụng trong các lĩnh vực như theo dõi sức khỏe tâm thần, phát hiện nói dối và trải nghiệm người dùng được cá nhân hóa.

Một ứng dụng khác là tương tác giữa người và máy tính, nơi các hệ thống có thể điều chỉnh hành vi của họ dựa trên cảm xúc được phát hiện của người dùng. Ví dụ: trợ lý ảo có thể cung cấp phản hồi đồng cảm hơn nếu phát hiện sự thất vọng hoặc buồn bã trong giọng nói của người dùng.

Phát hiện cảm xúc bằng CNN cũng có tiềm năng đáng kể trong nghiên cứu thị trường và phân tích tâm lý. Bằng cách phân tích các bài đăng trên mạng xã hội, đánh giá của khách hàng và nội dung văn bản khác, các công ty có thể hiểu rõ hơn về sự hài lòng của khách hàng, xu hướng tình cảm và hiệu quả của các chiến dịch tiếp thị của họ.

Thách thức và hạn chế

Mặc dù hiệu quả, phát hiện cảm xúc bằng CNN phải đối mặt với một số thách thức và hạn chế. Một thách thức là sự thay đổi trong biểu hiện cảm xúc giữa các cá nhân. Mọi người có thể thể hiện cùng một cảm xúc khác nhau, gây khó khăn cho việc tạo ra một mô hình toàn diện có thể khái quát hóa trên các nhóm dân cư đa dạng.

Sự khác biệt giữa các nền văn hóa trong biểu hiện cũng đặt ra những thách thức. Cảm xúc có thể bị ảnh hưởng bởi các chuẩn mực văn hóa và các yếu tố xã hội, đòi hỏi các mô hình phải được đào tạo trên các bộ dữ liệu đa dạng để đảm bảo tính chính xác và toàn diện.

Mối quan tâm về quyền riêng tư và đạo đức là một cân nhắc quan trọng khác. Công nghệ phát hiện cảm xúc đặt ra câu hỏi về quyền riêng tư dữ liệu, sự đồng ý và khả năng lạm dụng thông tin cá nhân. Các biện pháp bảo vệ phải được áp dụng để bảo vệ quyền riêng tư của người dùng và đảm bảo sử dụng công nghệ này một cách có trách nhiệm.

Sự phát triển và xu hướng trong tương lai

Lĩnh vực phát hiện cảm xúc bằng CNN không ngừng phát triển, với những tiến bộ không ngừng và xu hướng mới nổi. Một xu hướng là sự phát triển của các thuật toán học sâu phức tạp hơn có thể trích xuất các tín hiệu cảm xúc tinh tế từ

các phương thức khác nhau, chẳng hạn như âm thanh và văn bản, ngoài biểu cảm khuôn mặt.
Tích hợp với các công nghệ khác là một lĩnh vực quan tâm khác. Kết hợp phát hiện cảm xúc với xử lý ngôn ngữ tự nhiên và nhận dạng giọng nói có thể nâng cao khả năng của trợ lý ảo, chatbot và các hệ thống tương tác khác, cho phép trải nghiệm người dùng liền mạch và được cá nhân hóa hơn.
Tác động của việc phát hiện cảm xúc bằng CNN vượt ra ngoài các ứng dụng riêng lẻ. Nó có tiềm năng chuyển đổi các ngành như chăm sóc sức khỏe, giáo dục và tiếp thị, cho phép các hệ thống chăm sóc sức khỏe đồng cảm hơn, môi trường học tập được cá nhân hóa và các chiến dịch tiếp thị được nhắm mục tiêu.

Kết thúc

Phát hiện cảm xúc bằng Mạng nơ-ron tích chập cung cấp một công cụ mạnh mẽ để hiểu và phân tích cảm xúc của con người. Bằng cách tận dụng khả năng của CNN, chúng ta có thể mở ra những hiểu biết mới về hành vi của con người và phát triển các ứng dụng sáng tạo trong các lĩnh vực khác nhau. Tuy nhiên, điều quan trọng là phải giải quyết những thách thức và cân nhắc đạo đức liên quan đến công nghệ này để đảm bảo việc sử dụng có trách nhiệm và có lợi.

Chương 6. Demo