**f**

**BỘ CÔNG THƯƠNG**

**TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP HÀ NỘI**

------\*\*\*------



**BÁO CÁO THỰC NGHIỆM**

**MÔN:TRÍ TUỆ NHÂN TẠO**

**ĐỀ TÀI**

**XÂY DỰNG ỨNG DỤNG PHÁT HIỆN VIDEO GIẢ MẠO SỬ DỤNG MÔ HÌNH HỌC SÂU**

**Giảng viên hướng dẫn:** Ths.Trần Thanh Huân

**Các thành viên:NHÓM 6**

1.Lê Mạnh Cường MSV:2022601093

2.Nguyễn Hùng Cường MSV:2022600246

3.Hà Thị Trinh MSV:

4.Phạm Kiểu Trang MSV:

**HÀ NỘI:…/…/2025**

**LỜI CẢM ƠN**

Đầu tiên, chúng em xin gửi lời cảm ơn chân thành đến Trường Đại học Công nghiệp Hà Nội đã đưa môn học Trí tuệ nhân tạo vào trương trình giảng dạy. Đặc biệt, chúng em xin gửi lời cảm ơn sâu sắc đến giảng viên bộ môn – ThS. Trần Thanh Huân đã dạy dỗ, truyền đạt những kiến thức quý báu cho chúng em trong suốt thời gian học tập vừa qua. Trong thời gian tham gia lớp học của thầy, chúng em đã có thêm cho mình nhiều kiến thức bổ ích, tinh thần học tập hiệu quả, nghiêm túc. Đây chắc chắn sẽ là những kiến thức quý báu, là hành trang để chúng em có thể vững bước sau này.

Trí tuệ nhân tạo là một môn học thú vị, vô cùng bổ ích và có tính thực tế cao. Đảm bảo cung cấp đủ kiến thức, gắn liền với nhu cầu thực tiễn của sinh viên. Tuy nhiên, do vốn kiến thức còn nhiều hạn chế và khả năng tiếp thu thực tế còn nhiều bỡ ngỡ. Mặc dù chúng em đã cố gắng hết sức nhưng chắc chắn bài báo cáo khó có thể tránh khỏi những thiếu sót và nhiều chỗ còn chưa chính xác, kính mong thầy xem xét và góp ý để bài báo cáo của chúng em được hoàn thiện hơn.

Chúng em xin chân thành cảm ơn!

**MỤC LỤC**

[MỞ ĐẦU 1](#_Toc200313371)

[1. Lý do chọn đề tài 1](#_Toc200313372)

[2. Mục tiêu nghiên cứu 2](#_Toc200313373)

[3. Phương pháp nghiên cứu 2](#_Toc200313374)

[4. Đối tượng và phạm vi nghiên cứu 3](#_Toc200313375)

[5. Cấu trúc của báo cáo 3](#_Toc200313376)

[CHƯƠNG 1: TỔNG QUAN VỀ ĐỀ TÀI 4](#_Toc200313377)

[1.1. Trí tuệ nhân tạo và học sâu 4](#_Toc200313378)

[1.1.1. Khái niệm 4](#_Toc200313379)

[1.1.2. Vai trò trong nhận dạng và an ninh 4](#_Toc200313380)

[1.2. Video giả mạo và vấn đề deepfake 5](#_Toc200313381)

[1.2.1. Khái niệm và phân loại video giả mạo 5](#_Toc200313382)

[1.2.3. Tác động và nguy cơ 7](#_Toc200313383)

[1.3. Ứng dụng học sâu trong phát hiện video giả mạo 7](#_Toc200313384)

[1.3.1. Lịch sử và quá trình phát triển 7](#_Toc200313385)

[1.3.2. Hiện trạng nghiên cứu và ứng dụng 8](#_Toc200313386)

[1.3.3. Các lĩnh vực áp dụng 8](#_Toc200313387)

[CHƯƠNG 2: TÌM HIỂU VỀ DEEP LEARNING TRONG PHÁT HIỆN VIDEO GIẢ MẠO 10](#_Toc200313388)

[1. Giới thiệu chung về Deep Learning 10](#_Toc200313389)

[2. Sử dụng học sâu để phát hiện video giả 10](#_Toc200313390)

[2.1 Mô hình CNN (Convolutional Neural Network) để phân tích ảnh tĩnh trong video 10](#_Toc200313391)

[2.1.1. Giới thiệu về CNN (Convolutional Neural Network) 10](#_Toc200313392)

[2.1.2. Kiến trúc của CNN trong phân tích ảnh tĩnh từ video 11](#_Toc200313393)

[2.1.3. Cách sử dụng CNN để phát hiện ảnh Deepfake từ video 12](#_Toc200313394)

[2.2.RNN (Recurrent Neural Network) để phân tích chuỗi video 14](#_Toc200313395)

[2.2.1. Giới thiệu về RNN trong phân tích chuỗi video 14](#_Toc200313396)

[2.2.2. Kiến trúc RNN để phân tích video Deepfake 15](#_Toc200313397)

[2.3.Transformer Models (như Vision Transformer) giúp nâng cao hiệu suất phát hiện 16](#_Toc200313398)

[2.3.1. Giới thiệu về mô hình Transformer trong phát hiện Deepfake 16](#_Toc200313399)

[2.3.2. Kiến trúc của Vision Transformer (ViT) trong phát hiện Deepfake 17](#_Toc200313400)

[CHƯƠNG 3. ỨNG DỤNG THUẬT TOÁN CNN ĐỂ XÂY DỰNG CHƯƠNG TRÌNH PHÁT HIỆN VIDEO GIẢ MẠO 19](#_Toc200313401)

[3.1.Giới thiệu bài toán 19](#_Toc200313403)

[3.2Xây dựng mô hình bài toán 20](#_Toc200313404)

[3.3.Ngôn ngữ và công cụ sử dụng 20](#_Toc200313405)

[Ngôn ngữ và các thư viện 20](#_Toc200313406)

[IDE: Google Colab 21](#_Toc200313407)

[Các bước thực hiện: 22](#_Toc200313408)

[Thu thập dữ liệu 22](#_Toc200313409)

[Tiền xử lý dữ liệu 23](#_Toc200313410)

[3.4.Các bước thực hiện 23](#_Toc200313411)

[THỬ NGHIỆM VÀ ĐÁNH GIÁ HỆ THỐNG 31](#_Toc200313412)

[KẾT LUẬN 32](#_Toc200313413)

[TÀI LIỆU THAM KHẢO 33](#_Toc200313414)

DANH MỤC HÌNH ẢNH

[Hình 1: Kiến trúc CNN và cách hoạt động của từng lớp trong mạng nơ-ron. 11](#_Toc200313173)

[Hình 2: Cách một bộ lọc (filter) trích xuất đặc trưng từ ảnh khuôn mặt trong video. 12](#_Toc200313174)

[Hình 3: Quá trình lan truyền dữ liệu qua các lớp của CNN khi phân tích ảnh từ video. 12](#_Toc200313175)

[Hình ảnh 4: Minh họa một số ảnh thật và ảnh Deepfake trong bộ dữ liệu FaceForensics++. 13](#_Toc200313176)

[Hình ảnh 5: Kiến trúc RNN 15](#_Toc200313177)

[Hình ảnh 6: Quá trình trích xuất đặc trưng từ video trước khi đưa vào RNN. 15](#_Toc200313178)

[Hình 7: Kiến trúc cơ bản của Vision Transformer 17](#_Toc200313179)

[Hình 8: Quá trình chia ảnh thành các patch nhỏ trong Vision Transformer. 18](#_Toc200313180)

# MỞ ĐẦU

## 1. Lý do chọn đề tài

Trong thời đại công nghệ số phát triển mạnh mẽ, thông tin được lan truyền một cách nhanh chóng và rộng khắp thông qua các nền tảng truyền thông số như mạng xã hội, báo chí điện tử, video trực tuyến,... Tuy nhiên, cùng với sự phát triển đó là sự xuất hiện ngày càng phổ biến của các nội dung giả mạo, trong đó đáng lo ngại nhất là sự phát triển của công nghệ deepfake – kỹ thuật sử dụng trí tuệ nhân tạo để tạo ra video giả mạo với độ chân thực rất cao.

Deepfake có thể thay đổi khuôn mặt, giọng nói hoặc hành động của một người trong video một cách tinh vi, khiến người xem khó nhận ra đó là sản phẩm nhân tạo. Những video này có thể bị lợi dụng với nhiều mục đích tiêu cực như bôi nhọ danh dự, phát tán thông tin sai lệch, can thiệp vào chính trị, lừa đảo tài chính hoặc gây tổn hại đến an ninh cá nhân và xã hội. Vấn đề trở nên nghiêm trọng hơn khi các công cụ tạo deepfake ngày càng dễ tiếp cận, không đòi hỏi kiến thức chuyên sâu, và có thể được sử dụng bởi bất kỳ ai.

Chính vì vậy, nhu cầu nghiên cứu và phát triển các hệ thống phát hiện video giả mạo ngày càng trở nên cấp thiết. Trong số các giải pháp được đề xuất, mô hình học sâu (deep learning), đặc biệt là mạng nơ-ron tích chập (CNN – Convolutional Neural Network), đã chứng minh hiệu quả cao trong việc trích xuất đặc trưng hình ảnh và nhận diện các dấu hiệu bất thường trong video. CNN có khả năng tự động học và phát hiện các mẫu đặc trưng phức tạp trong dữ liệu hình ảnh, giúp hệ thống phát hiện các điểm sai lệch mà mắt thường khó nhận ra – chẳng hạn như sự khác biệt nhỏ trong cấu trúc khuôn mặt, ánh sáng, hoặc chuyển động không tự nhiên giữa các khung hình.

Ngoài ra, các phương pháp hỗ trợ như so sánh tương quan chéo chuẩn hóa (Normalized Cross-Correlation – NCC) cũng có thể được tích hợp để phân tích mức độ giống nhau giữa các vùng ảnh, phục vụ cho việc nhận diện các thao tác chỉnh sửa trong video. Khi kết hợp NCC với CNN, hệ thống có thể không chỉ học đặc trưng mà còn đánh giá được mức độ bất thường tại từng vùng ảnh cụ thể.

Với mong muốn tìm hiểu sâu hơn về các kỹ thuật trí tuệ nhân tạo hiện đại, đồng thời nâng cao kỹ năng xử lý ảnh và xây dựng mô hình học sâu, nhóm chúng em quyết định lựa chọn đề tài: “Xây dựng chương trình phát hiện video giả mạo sử dụng mô hình học sâu”. Đây không chỉ là cơ hội để áp dụng các kiến thức đã học vào một bài toán thực tiễn có tính thời sự cao, mà còn giúp chúng em rèn luyện các kỹ năng phân tích, thiết kế thuật toán và triển khai phần mềm trong lĩnh vực an toàn thông tin số.

## 2. Mục tiêu nghiên cứu

* Tìm hiểu tổng quan về video giả mạo và các kỹ thuật tạo deepfake phổ biến hiện nay.
* Nghiên cứu các mô hình học sâu được sử dụng trong bài toán phát hiện video giả mạo, đặc biệt là các mạng nơ-ron tích chập (CNN), mạng LSTM hoặc mô hình kết hợp.
* Xây dựng chương trình phát hiện video giả mạo trên cơ sở ứng dụng mô hình học sâu
* Đánh giá độ chính xác của mô hình thông qua một tập dữ liệu video giả mạo và thật.
* Đề xuất hướng cải tiến để tăng hiệu quả nhận diện trong tương lai.

## 3. Phương pháp nghiên cứu

* Phương pháp lý thuyết: Nghiên cứu tài liệu, báo cáo khoa học, các bài báo và tài nguyên học thuật liên quan đến lĩnh vực deepfake, phát hiện video giả mạo và học sâu.
* Phương pháp thực nghiệm: Xây dựng mô hình học sâu, huấn luyện trên bộ dữ liệu thật – giả, đánh giá kết quả thực nghiệm dựa trên các chỉ số độ chính xác, độ nhạy, F1-score...
* Phương pháp phân tích – tổng hợp: So sánh hiệu quả các mô hình, phân tích nguyên nhân ảnh hưởng đến kết quả và đề xuất cải tiến.

## 4. Đối tượng và phạm vi nghiên cứu

* Đối tượng nghiên cứu:
* Thuật toán CNN (Convolutional Neural Network).
* Hệ thống nhận diện khuôn mặt.
* Ngôn ngữ: Python.
* Công cụ: Google Colab.
* Phạm vi nghiên cứu:
* Không gian: Thực hiện trong môi trường mô phỏng với các tập dữ liệu tự thu thập.
* Thời gian: Từ khi nhận đề tài tới khi hoàn thành.
* Nội dung: Thuật toán CNN, tính năng nhận diện khuôn mặt.

## 5. Cấu trúc của báo cáo

Ngoài phần mở đầu và kết luận, quyển báo cáo còn bao gồm hai nội dung tương ứng với hai chương:

* + Chương 1: Tổng quan về đề tài
  + Chương 2: Tìm hiểu về deep learning trong phát hiện video giả mạo
* Chương 3: Ứng dụng thuật toán CNN để xây dựng chương trình phát hiện video giả mạo

# CHƯƠNG 1: TỔNG QUAN VỀ ĐỀ TÀI

## 1.1. Trí tuệ nhân tạo và học sâu

### 1.1.1. Khái niệm

Trí tuệ nhân tạo (AI - Artificial Intelligence) là một lĩnh vực nghiên cứu thuộc khoa học máy tính, với mục tiêu xây dựng các hệ thống có khả năng mô phỏng trí tuệ và hành vi của con người. AI cho phép máy tính thực hiện các tác vụ đòi hỏi trí thông minh như học tập từ dữ liệu, suy luận logic, xử lý ngôn ngữ tự nhiên, nhận diện hình ảnh, đưa ra quyết định, và tương tác với môi trường.

Trong số các nhánh của AI, học sâu (Deep Learning) nổi lên như một phương pháp đặc biệt hiệu quả trong việc xử lý các dữ liệu phi cấu trúc như ảnh, video và âm thanh. Học sâu sử dụng các mạng nơ-ron sâu (deep neural networks) với nhiều lớp ẩn, có khả năng tự động trích xuất và học các đặc trưng phức tạp từ dữ liệu. Chính sự phát triển của Deep Learning đã thúc đẩy mạnh mẽ các ứng dụng AI trong nhận diện khuôn mặt, dịch máy, xe tự lái, và đặc biệt là các ứng dụng liên quan đến phát hiện nội dung giả mạo.

### 1.1.2. Vai trò trong nhận dạng và an ninh

AI và học sâu không chỉ giúp nâng cao hiệu quả trong các lĩnh vực sản xuất, tài chính, y tế mà còn đóng vai trò then chốt trong lĩnh vực an ninh và xác thực thông tin. Trong bối cảnh thông tin số lan truyền mạnh mẽ như hiện nay, việc đảm bảo tính xác thực của hình ảnh và video trở thành vấn đề cấp bách.

Các thuật toán học sâu có khả năng nhận biết những biểu hiện rất nhỏ và khó phát hiện bằng mắt thường trong các video giả mạo, từ đó giúp phát hiện các sản phẩm deepfake một cách chính xác. Việc ứng dụng AI vào lĩnh vực này đang dần trở thành một công cụ quan trọng nhằm ngăn chặn thông tin sai lệch, bảo vệ người dùng và giữ gìn sự minh bạch trong truyền thông số.

## 1.2. Video giả mạo và vấn đề deepfake

### 1.2.1. Khái niệm và phân loại video giả mạo

#### 1.2.1.1. Khái niệm về video giả mạo

Video giả mạo là những đoạn phim đã bị can thiệp, chỉnh sửa hoặc tạo ra nhằm làm sai lệch thông tin thực tế, khiến người xem tin rằng những gì được thể hiện trong video là thật dù chúng đã bị thay đổi hoặc dàn dựng. Đây là một dạng thông tin sai lệch tinh vi trong kỷ nguyên số, đặc biệt nguy hiểm vì hình ảnh và âm thanh thường mang tính thuyết phục cao hơn so với văn bản.

Video giả mạo có thể được tạo ra bởi công nghệ trí tuệ nhân tạo (AI) hoặc bằng các kỹ thuật chỉnh sửa video truyền thống. Sự phát triển mạnh mẽ của AI, đặc biệt là học sâu (Deep Learning), đã khiến cho video giả mạo trở nên ngày càng tinh vi, khó phát hiện và gây nhiều lo ngại về đạo đức, an ninh và pháp lý.

#### 1.2.1.2. Phân loại video giả mạo

Dựa trên phương pháp tạo ra, video giả mạo có thể được chia thành hai loại chính:

**a. Deepfake Video (Video giả mạo bằng AI)**

Đây là các video được tạo hoặc chỉnh sửa bằng thuật toán trí tuệ nhân tạo, đặc biệt là các mô hình học sâu như Generative Adversarial Networks (GANs), Autoencoders, hoặc các mạng nơ-ron tích chập (CNN). Chúng có khả năng tổng hợp hình ảnh khuôn mặt, cử chỉ, giọng nói của con người một cách tự nhiên, chân thực đến mức khó phân biệt bằng mắt thường.

Một số hình thức phổ biến:

* Face Swapping (Hoán đổi khuôn mặt): Thay khuôn mặt của một người bằng người khác.
* Face Reenactment (Tái hiện biểu cảm khuôn mặt): Điều khiển biểu cảm khuôn mặt của người trong video bằng dữ liệu từ người khác.
* Lip Syncing: Đồng bộ môi và lời nói sai lệch với nội dung gốc.
* Voice Cloning: Giả giọng nói nhờ huấn luyện AI.
* Full-body Deepfake: Làm giả cả tư thế, cử động toàn thân.

Ví dụ: Một video phát biểu của chính trị gia được tạo ra bằng cách ghép khuôn mặt và giọng nói của họ vào hình ảnh và lời nói không có thật.

**b. Manipulated Video (Video bị chỉnh sửa thủ công)**

Đây là các video được can thiệp bằng các công cụ chỉnh sửa truyền thống, không sử dụng AI, thường bằng các phần mềm như Adobe Premiere, After Effects, hoặc các phần mềm dựng phim khác. Các chỉnh sửa có thể rất tinh vi, nhưng thường mất nhiều thời gian và công sức thủ công.

Một số dạng chỉnh sửa thủ công:

* Cắt ghép video: Loại bỏ một phần nội dung hoặc ghép nối các phần rời rạc để thay đổi ngữ cảnh.
* Thêm hiệu ứng giả: Chèn âm thanh, phụ đề sai lệch, hiệu ứng giả mạo để đánh lừa người xem.
* Tua nhanh/chậm, đảo ngược video: Thay đổi tốc độ hoặc trật tự của các cảnh để tạo ra hiểu nhầm.
* Dừng hình tại điểm mấu chốt: Dừng video tại khoảnh khắc gây hiểu lầm (thường thấy trên mạng xã hội).

Ví dụ: Một video người nổi tiếng bị cắt bỏ phần giải thích, chỉ giữ lại đoạn phát ngôn gây tranh cãi nhằm bôi nhọ họ.

### 1.2.3. Tác động và nguy cơ

Deepfake mang đến những thách thức to lớn về mặt đạo đức, xã hội và an ninh thông tin. Một mặt, công nghệ này thể hiện khả năng sáng tạo vô hạn của con người trong lĩnh vực trí tuệ nhân tạo. Nhưng mặt khác, nếu không được kiểm soát, deepfake có thể trở thành công cụ nguy hiểm, gây ra hậu quả nghiêm trọng:

* Trong chính trị: Deepfake có thể tạo ra các phát biểu hoặc hành động giả mạo của chính trị gia, ảnh hưởng đến uy tín và gây rối loạn xã hội.
* Trong đời sống cá nhân: Việc tạo video giả mạo một người trong các tình huống nhạy cảm có thể phá hoại cuộc sống cá nhân, làm mất danh dự và gây tổn thương tâm lý.
* Trong kinh tế và pháp lý: Deepfake có thể được sử dụng để giả mạo các giao dịch tài chính, hợp đồng, hoặc làm giả bằng chứng trong các vụ kiện.

Chính vì vậy, nhu cầu phát hiện và ngăn chặn các video giả mạo ngày càng trở nên cấp thiết.

## 1.3. Ứng dụng học sâu trong phát hiện video giả mạo

### 1.3.1. Lịch sử và quá trình phát triển

Trong giai đoạn đầu, việc phát hiện video giả chủ yếu dựa vào các kỹ thuật xử lý ảnh cổ điển như phân tích chuyển động, ánh sáng, hoặc dấu hiệu cắt ghép. Tuy nhiên, các phương pháp này thường kém hiệu quả trước những video được tạo bằng kỹ thuật cao.

Sự phát triển của học sâu đã mở ra một hướng đi mới. Các mạng nơ-ron tích chập (CNN) có khả năng học đặc trưng trực tiếp từ hình ảnh và video, phát hiện các dấu hiệu tinh vi mà con người không thể nhận ra. Các mô hình như XceptionNet, MesoNet, EfficientNet, và gần đây là Vision Transformer (ViT) đã được nghiên cứu và chứng minh tính hiệu quả trong việc phát hiện deepfake. Một số mô hình kết hợp cả thông tin về khuôn mặt, chuyển động môi, ánh sáng, và thời gian để đưa ra đánh giá tổng thể về độ tin cậy của video.

### 1.3.2. Hiện trạng nghiên cứu và ứng dụng

Hiện nay, rất nhiều nghiên cứu trên thế giới tập trung vào phát hiện deepfake. Một số tổ chức lớn như Facebook AI, Google Research, hay Microsoft đã công bố các công cụ hoặc cuộc thi liên quan, nhằm nâng cao nhận thức và tìm ra giải pháp hiệu quả. Một số bộ dữ liệu nổi bật được dùng trong huấn luyện và đánh giá mô hình gồm:

* FaceForensics**++**: Bộ dữ liệu chứa hơn 1.000 video gốc và giả mạo.
* DFDC (Deepfake Detection Challenge): Cuộc thi do Facebook tổ chức với bộ dữ liệu lên đến hàng trăm ngàn video deepfake.
* Celeb-DF: Dữ liệu gồm các video của người nổi tiếng bị giả mạo khuôn mặt.

Ngoài giới nghiên cứu, nhiều nền tảng mạng xã hội và truyền thông cũng đã triển khai các công cụ kiểm tra video, nhằm bảo vệ người dùng và ngăn chặn thông tin sai lệch.

### 1.3.3. Các lĩnh vực áp dụng

Các hệ thống phát hiện deepfake được ứng dụng rộng rãi trong nhiều lĩnh vực:

* Truyền thông: Kiểm chứng tính xác thực của nội dung trước khi phát sóng hoặc chia sẻ.
* Pháp luật và hành chính: Hỗ trợ điều tra, loại bỏ bằng chứng giả, ngăn ngừa các hành vi lừa đảo.
* An ninh quốc gia: Phát hiện các chiến dịch thông tin sai lệch có chủ đích.
* Mạng xã hội: Phát hiện và cảnh báo người dùng về nội dung bị chỉnh sửa hoặc giả mạo.

# CHƯƠNG 2: TÌM HIỂU VỀ DEEP LEARNING TRONG PHÁT HIỆN VIDEO GIẢ MẠO

## 1. Giới thiệu chung về Deep Learning

Deep Learning là một lĩnh vực của học máy (Machine Learning) sử dụng các mạng nơ-ron nhân tạo nhiều lớp để học và biểu diễn các mẫu dữ liệu phức tạp. Với khả năng xử lý dữ liệu hình ảnh, âm thanh và video hiệu quả, Deep Learning là nền tảng chính trong việc phát hiện các video giả mạo, đặc biệt là Deepfake.

## 2. Sử dụng học sâu để phát hiện video giả

## 2.1 Mô hình CNN (Convolutional Neural Network) để phân tích ảnh tĩnh trong video

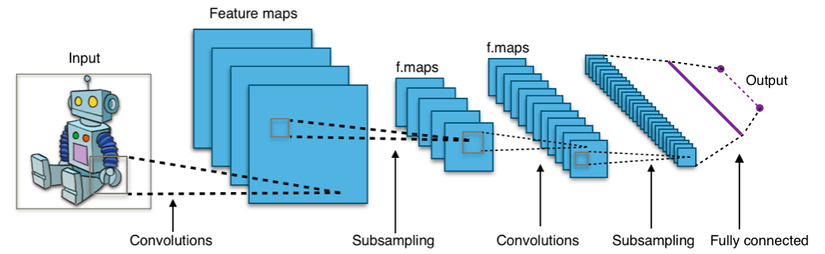
### 2.1.1. Giới thiệu về CNN (Convolutional Neural Network)

CNN (Convolutional Neural Network) là một mô hình Deep Learning chuyên biệt trong việc xử lý dữ liệu hình ảnh và video. Nhờ vào khả năng tự động trích xuất đặc trưng từ hình ảnh, CNN trở thành một công cụ mạnh mẽ trong việc phát hiện video Deepfake, đặc biệt là khi phân tích ảnh tĩnh trích xuất từ video.

Trong bối cảnh phát hiện video giả mạo, CNN có thể được sử dụng để:

* Phân tích từng khung hình (frame) của video để xác định dấu hiệu giả mạo**.**
* Phát hiện các bất thường trong khuôn mặt, ánh sáng, kết cấu da, bóng đổ do AI tạo ra.
* Tích hợp với các mô hình phân tích chuyển động (RNN, LSTM) để nâng cao độ chính xác.

**Hình ảnh minh họa:**



##### Hình 1: Kiến trúc CNN và cách hoạt động của từng lớp trong mạng nơ-ron.

### 2.1.2. Kiến trúc của CNN trong phân tích ảnh tĩnh từ video

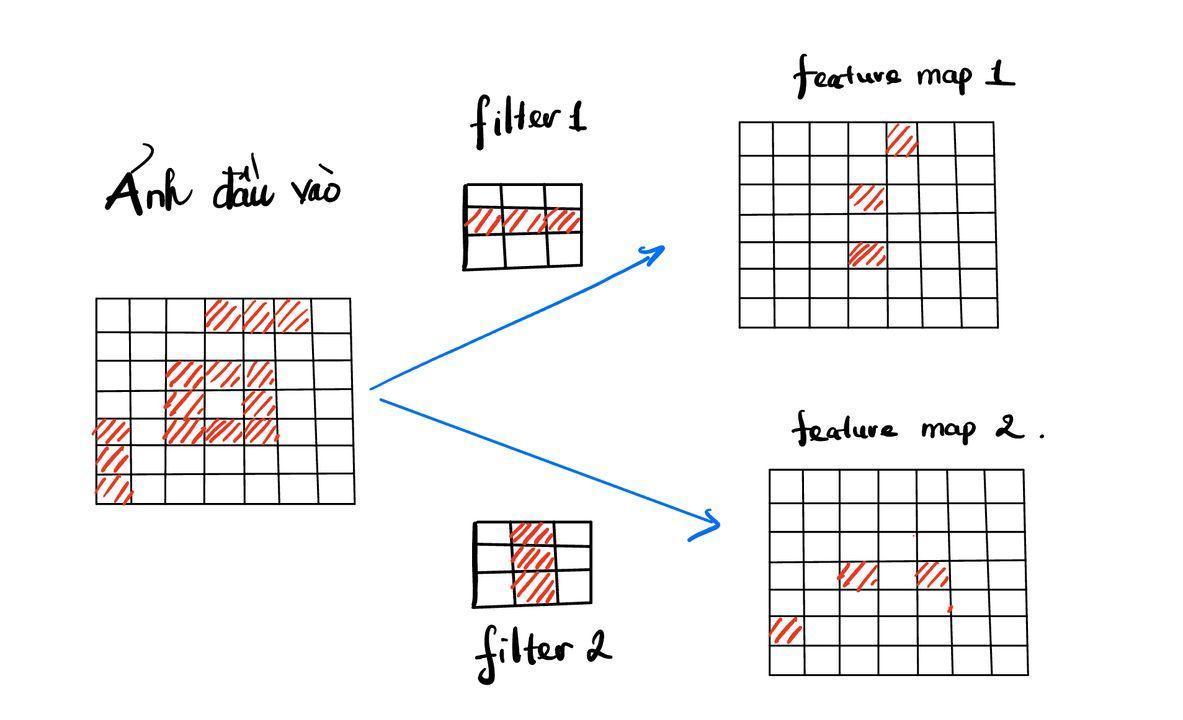
Mạng CNN có một kiến trúc đặc biệt bao gồm nhiều lớp giúp trích xuất đặc trưng của hình ảnh một cách hiệu quả. Dưới đây là các thành phần chính:

#### 2.1.2.1. Lớp Convolution (Lớp tích chập)

Lớp Convolution là lớp chính của CNN, có nhiệm vụ trích xuất các đặc trưng quan trọng từ ảnh, như:

* Các cạnh, góc, đường nét ở các cấp độ khác nhau.
* Chi tiết nhỏ trên khuôn mặt như nếp nhăn, lỗ chân lông (mà Deepfake thường tái tạo không chính xác).
* Bất thường trong ánh sáng và kết cấu da, giúp phát hiện những lỗi do AI tạo ra.

**Hình ảnh minh họa:**



##### Hình 2: Cách một bộ lọc (filter) trích xuất đặc trưng từ ảnh khuôn mặt trong video.

#### 2.1.2.2. Lớp Pooling (Lớp giảm chiều)

Lớp này giúp giảm kích thước của dữ liệu nhưng vẫn giữ được thông tin quan trọng, giúp mô hình hoạt động hiệu quả hơn.

**Max Pooling**: Giữ lại các giá trị lớn nhất trong một vùng nhất định.

**Average Pooling**: Tính trung bình các giá trị trong vùng.

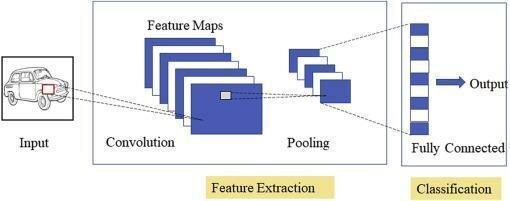
#### 2.1.2.3. Lớp Fully Connected (FC Layer - Lớp kết nối đầy đủ)

Sau khi trích xuất đặc trưng, các lớp Fully Connected sẽ xử lý và phân loại xem ảnh có phải là Deepfake hay không.

#### 2.1.2.4. Lớp Softmax / Sigmoid (Lớp đầu ra)

Lớp này đưa ra xác suất dự đoán xem ảnh là thật hay giả.

**Hình ảnh minh họa:**



##### Hình 3: Quá trình lan truyền dữ liệu qua các lớp của CNN khi phân tích ảnh từ video.

### 2.1.3. Cách sử dụng CNN để phát hiện ảnh Deepfake từ video

#### **Bước 1: Tiền xử lý dữ liệu**

Trước khi đưa ảnh vào CNN, cần thực hiện các bước tiền xử lý:

Trích xuất các khung hình từ video (mỗi giây lấy một số lượng ảnh nhất định).

Chuyển đổi ảnh về cùng kích thước (ví dụ: 224x224 pixels).

Chuẩn hóa dữ liệu (normalize pixel values về khoảng [0,1] hoặc [-1,1]).

#### **Bước 2: Huấn luyện mô hình CNN trên tập dữ liệu ảnh thật và ảnh giả**

Mô hình CNN được huấn luyện trên các bộ dữ liệu lớn chứa ảnh khuôn mặt thật và giả mạo để nhận diện các đặc trưng của Deepfake.  
Một số bộ dữ liệu phổ biến:

FaceForensics++: Bộ dữ liệu video Deepfake có độ phân giải cao.

DFDC (Deepfake Detection Challenge Dataset): Dữ liệu từ Facebook phục vụ phát hiện Deepfake.

Celeb-DF: Bộ dữ liệu Deepfake của các nhân vật nổi tiếng.

**Hình ảnh minh họa:**



##### Hình ảnh 4: Minh họa một số ảnh thật và ảnh Deepfake trong bộ dữ liệu FaceForensics++.

#### **Bước 3: Sử dụng CNN để dự đoán ảnh trong video có phải là Deepfake không**

Với mỗi ảnh đầu vào, CNN sẽ phân tích và trích xuất các đặc trưng.

Mô hình sẽ dự đoán xác suất ảnh là thật hay giả.

Nếu một số lượng lớn ảnh từ video có xác suất là giả, thì video có thể là Deepfake.

Mô hình CNN đóng vai trò quan trọng trong việc phân tích ảnh tĩnh từ video để phát hiện Deepfake. Nhờ khả năng trích xuất đặc trưng mạnh mẽ, CNN có thể xác định những bất thường trong khuôn mặt, ánh sáng, kết cấu da, giúp nhận diện video giả mạo với độ chính xác cao.

Tuy nhiên, để tăng cường hiệu quả, CNN có thể được kết hợp với các phương pháp khác như phân tích chuyển động (RNN, LSTM) và phân tích âm thanh để nâng cao khả năng phát hiện Deepfake một cách toàn diện.

## 2.2.RNN (Recurrent Neural Network) để phân tích chuỗi video

### 2.2.1. Giới thiệu về RNN trong phân tích chuỗi video

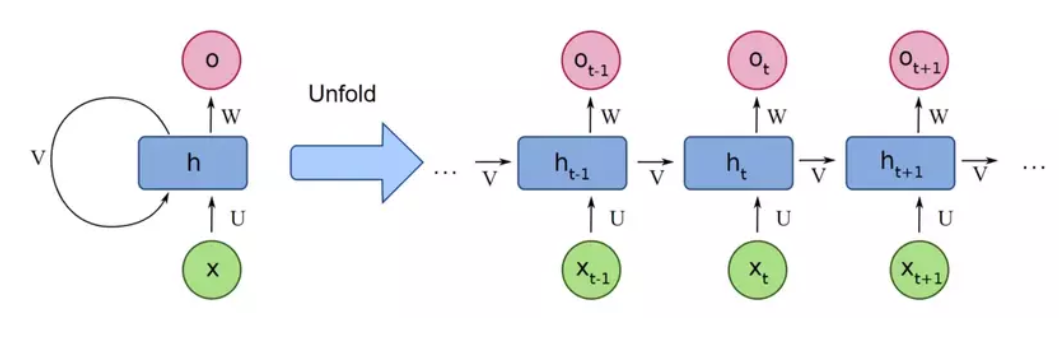
Recurrent Neural Network (RNN) là một loại mạng nơ-ron nhân tạo được thiết kế để xử lý dữ liệu chuỗi, chẳng hạn như văn bản, âm thanh và đặc biệt là video, nơi thông tin không chỉ phụ thuộc vào một khung hình (frame) đơn lẻ mà còn dựa trên sự liên kết giữa các khung hình theo thời gian.

Trong bối cảnh phát hiện video Deepfake, RNN có thể giúp:

Phân tích sự thay đổi của khuôn mặt, cử động môi và nét mặt qua từng khung hình của video.

Nhận diện sự không đồng bộ giữa hình ảnh và âm thanh.

Xác định các điểm bất thường trong chuyển động hoặc biểu cảm, điều mà CNN khó phát hiện khi chỉ phân tích ảnh tĩnh.

**Hình ảnh minh họa:**

##### Hình ảnh 5: Kiến trúc RNN

### 2.2.2. Kiến trúc RNN để phân tích video Deepfake

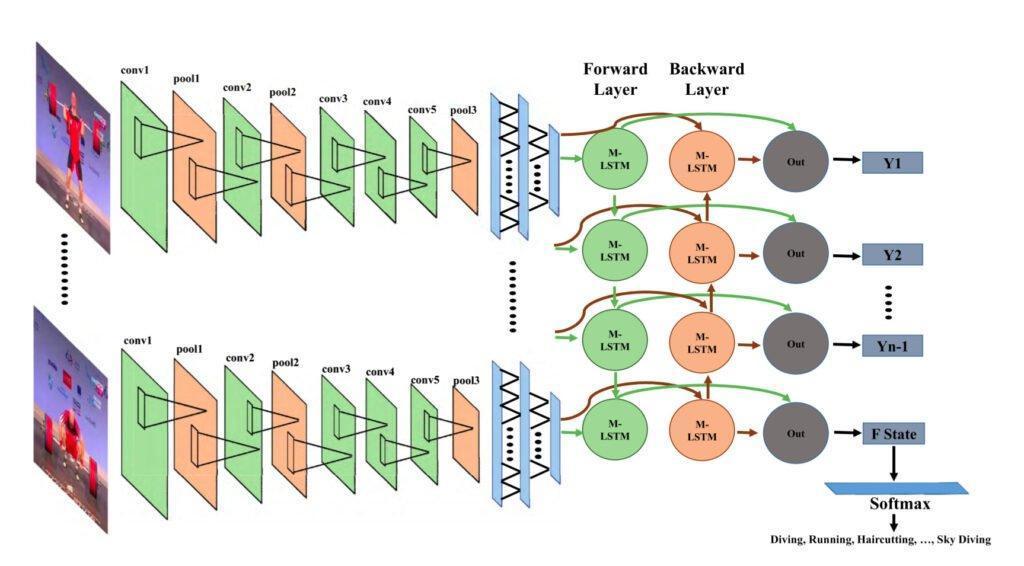
#### 2.2.2.1. Lớp đầu vào: Trích xuất đặc trưng từ video

Mỗi video được chia thành một chuỗi khung hình (frames).

Mô hình CNN có thể được sử dụng trước để trích xuất đặc trưng từ từng khung hình.

Các đặc trưng trích xuất từ mỗi frame sẽ được đưa vào RNN để phân tích theo thời gian.

**Hình ảnh minh họa:**



##### Hình ảnh 6: Quá trình trích xuất đặc trưng từ video trước khi đưa vào RNN.

#### 2.2.2.2. Lớp RNN: Xử lý chuỗi dữ liệu theo thời gian

RNN có khả năng nhớ trạng thái trước đó trong chuỗi dữ liệu, giúp nó hiểu được sự thay đổi của hình ảnh theo thời gian. Tuy nhiên, mô hình RNN truyền thống thường gặp vấn đề vanishing gradient (độ dốc biến mất) khi xử lý các chuỗi dài, vì vậy ta thường sử dụng LSTM (Long Short-Term Memory) hoặc GRU (Gated Recurrent Unit) để tối ưu hóa:

**LSTM:** Giữ lại thông tin quan trọng từ các bước trước đó để dự đoán chính xác hơn.

**GRU:** Giống LSTM nhưng có ít tham số hơn, giúp tăng tốc độ xử lý.

#### 2.2.2.3. Lớp Fully Connected và đầu ra

Sau khi chuỗi video được phân tích bởi RNN/LSTM, nó được đưa vào một lớp Fully Connected (FC) để tổng hợp thông tin.

Lớp đầu ra sử dụng hàm Softmax hoặc Sigmoid để dự đoán video có phải là Deepfake hay không.

RNN (đặc biệt là LSTM và GRU) là một công cụ mạnh mẽ giúp phân tích chuỗi hình ảnh từ video để phát hiện Deepfake.

**Ưu điểm của RNN trong phát hiện Deepfake:**  
- Có thể phát hiện sự bất thường trong chuyển động khuôn mặt theo thời gian.  
- Nhận diện sự không đồng bộ giữa âm thanh và cử động môi.  
- Kết hợp tốt với CNN để tạo ra mô hình mạnh mẽ hơn.

**Nhược điểm:**

- RNN/LSTM tốn nhiều tài nguyên tính toán hơn CNN.

- Cần dữ liệu huấn luyện lớn để đạt độ chính xác cao.

Nhìn chung, khi kết hợp CNN để trích xuất đặc trưng từ từng khung hình và RNN để phân tích chuỗi thời gian, ta có thể phát hiện Deepfake chính xác hơn, góp phần ngăn chặn các video giả mạo trên mạng xã hội và truyền thông.

## 2.3.Transformer Models (như Vision Transformer) giúp nâng cao hiệu suất phát hiện

### 2.3.1. Giới thiệu về mô hình Transformer trong phát hiện Deepfake

Transformer là một kiến trúc mạng nơ-ron mạnh mẽ được phát triển ban đầu cho xử lý ngôn ngữ tự nhiên (NLP), nhưng gần đây đã được áp dụng thành công vào lĩnh vực xử lý hình ảnh và video. Một trong những mô hình nổi bật là Vision Transformer (ViT), sử dụng cơ chế self-attention để phân tích hình ảnh thay vì các lớp tích chập (Convolutional Layers) như CNN.

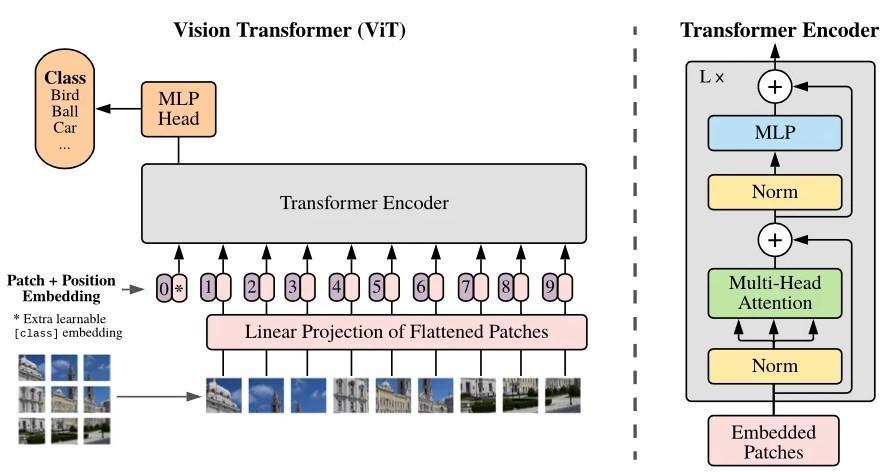
Trong phát hiện video Deepfake, Vision Transformer có thể giúp:

Trích xuất đặc trưng từ từng khung hình trong video với độ chính xác cao hơn CNN.

Xác định các bất thường trong kết cấu da, ánh sáng, và bóng đổ trên khuôn mặt.

Kết hợp với mô hình RNN hoặc LSTM để phân tích sự thay đổi theo thời gian, nâng cao hiệu quả nhận diện.

**Hình ảnh minh họa:**



Hình 7: Kiến trúc cơ bản của Vision Transformer*.*

### 2.3.2. Kiến trúc của Vision Transformer (ViT) trong phát hiện Deepfake

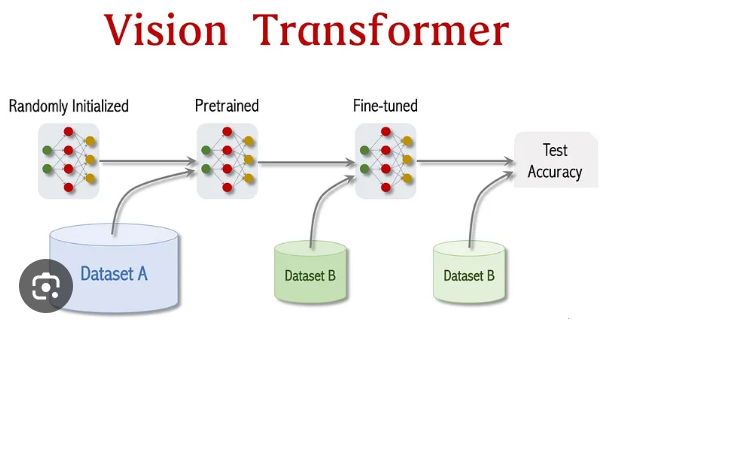
Vision Transformer hoạt động dựa trên cơ chế **self-attention**, giúp mô hình học được mối quan hệ giữa các điểm ảnh trong toàn bộ khung hình, thay vì chỉ tập trung vào từng vùng nhỏ như CNN. Dưới đây là các bước hoạt động chính:

#### 2.3.2.1. Chia ảnh thành các patch nhỏ

Một hình ảnh đầu vào (ví dụ: 224x224 pixels) sẽ được chia thành nhiều patch nhỏ (ví dụ: 16x16 pixels).

Mỗi patch này sẽ được biến đổi thành vector và được xử lý như một token trong mô hình NLP.

**Hình ảnh minh họa:**



##### Hình 8: Quá trình chia ảnh thành các patch nhỏ trong Vision Transformer.

#### 2.3.2.2. Ánh xạ (Embedding) và thêm vị trí (Positional Encoding)

Mỗi patch sau khi được vector hóa sẽ được **mã hóa thông tin vị trí** để giúp mô hình nhận diện mối quan hệ giữa các vùng khác nhau trong ảnh.

#### 2.3.2.3. Cơ chế Self-Attention trong Transformer

Self-Attention giúp mô hình học được các mối quan hệ giữa các vùng khác nhau của ảnh.

Điều này đặc biệt hữu ích trong phát hiện Deepfake, vì nó có thể phát hiện các bất thường trong kết cấu da, chi tiết mắt, miệng, và tóc.

#### 2.3.2.4. Lớp MLP Head và đầu ra

Sau khi được xử lý qua nhiều lớp Transformer, ảnh sẽ được đưa vào lớp MLP (Multilayer Perceptron) để phân loại xem ảnh có phải là Deepfake hay không.

Hàm Softmax hoặc Sigmoid được sử dụng để đưa ra xác suất dự đoán.

# CHƯƠNG 3. ỨNG DỤNG THUẬT TOÁN CNN ĐỂ XÂY DỰNG CHƯƠNG TRÌNH PHÁT HIỆN VIDEO GIẢ MẠO



## 3.1.Giới thiệu bài toán

Trong thời đại bùng nổ của công nghệ số và trí tuệ nhân tạo, các công cụ chỉnh sửa và tạo dựng nội dung số ngày càng trở nên tinh vi. Một trong những thách thức lớn hiện nay là sự xuất hiện của các video giả mạo (deepfake), trong đó hình ảnh khuôn mặt hoặc giọng nói của con người bị thay thế hoặc làm giả một cách tinh vi bằng công nghệ học sâu. Những video này có thể gây ra hậu quả nghiêm trọng về mặt đạo đức, pháp lý và xã hội nếu bị lợi dụng vào mục đích xấu.

Trước vấn đề đó, việc phát triển các phương pháp tự động và chính xác để phát hiện video giả mạo là vô cùng cần thiết. Trong nghiên cứu này, chúng tôi đề xuất sử dụng mạng nơ-ron tích chập (Convolutional Neural Network - CNN), cụ thể là kiến trúc **Xception**, để xây dựng mô hình phát hiện video giả mạo. Xception là một mô hình mạnh mẽ đã chứng minh hiệu quả vượt trội trong các bài toán phân loại và nhận dạng hình ảnh nhờ vào cơ chế tách biệt không gian - chiều sâu trong phép tích chập, giúp mô hình học được các đặc trưng phức tạp từ dữ liệu.

Bài toán đặt ra là: **Làm thế nào để phân biệt một video là thật hay giả mạo dựa trên các đặc trưng hình ảnh trích xuất từ từng khung hình (frame)?** Để giải quyết bài toán này, chúng em sử dụng mô hình Xception để học và phân tích các đặc trưng tiềm ẩn trong khuôn mặt người, từ đó phát hiện các dấu hiệu giả mạo không thể nhìn thấy bằng mắt thường.

## 3.2.Xây dựng mô hình bài toán

Bài toán phát hiện video giả mạo được chia thành 9 bước thực hiện chính, mỗi bước đóng vai trò quan trọng trong việc đảm bảo tính chính xác và hiệu quả của mô hình:

* Bước 1:Import thư viện cần thiết
* Bước 2:Trích xuất khung hình từ video và lưu vào thư mục ảnh
* Bước 3:Khởi tạo mô hình Xception và xây dựng tầng phân loại
* Bước 4:Tiền xử lý và nạp dữ liệu ảnh huấn luyện mô hình
* Bước 5:Tải mô hình đã huấn luyện để sử dụng
* Bước 6:Hàm trích xuất khung hình từ video đầu vào để dự đoán
* Bước 7:Hàm phát hiện Deepfake từ video đầu vào
* Bước 8:Thử nghiệm mô hình

## 3.3.Ngôn ngữ và công cụ sử dụng

### Ngôn ngữ và các thư viện

#### Ngôn ngữ Python

Python là một ngôn ngữ lập trình bậc cao, mã nguồn mở, đa năng được sử dụng rộng rãi trong nhiều lĩnh vực khác nhau. Ngôn ngữ này được tạo ra bởi Guido Van Rossum vào năm 1991 và được phát triển và duy trì bởi một cộng đồng lớn các lập trình viên trên toàn thế giới.

Python được ưa chuộng trong lĩnh vực học máy và trí tuệ nhân tạo nhờ vào những ưu điểm sau:

* + Cú pháp dễ đọc, dễ học: Python có cú pháp rất rõ ràng, dễ hiểu, gần giống với ngôn ngữ tự nhiên, giúp người mới bắt đầu dễ dàng tiếp cận.
  + Đa năng: Python có thể được sử dụng cho nhiều mục đích khác nhau, bao gồm phát triển ứng dụng web, phát triển phần mềm, khoa học dữ liệu, học máy, tự động hóa, …
  + Hiệu suất cao: Python có khả năng xử lý dữ liệu ổn định và hiệu quả ngay cả với các dự án lớn.
  + Môi trường phát triển tích hợp (IDE) mạnh mẽ: Python có nhiều IDE mạnh mẽ như PyCharm, Jupyter Notebook, Google Colab, hỗ trợ lập trình viên phát triển và gỡ lỗi mã.

#### Các thư viện

OpenCV là một thư viện mã nguồn mở cho xử lý hình ảnh và thị giác máy tính. Thư viện này cung cấp một loạt các chức năng để xử lý hình ảnh, video và dữ liệu nhận dạng. OpenCV được sử dụng rộng rãi trong nhiều lĩnh vực khác nhau, bao gồm xử lý hình ảnh, xử lý video, thị giác máy tính, …

TensorFlow là một thư viện mã nguồn mở cho học máy. Thư viện này cung cấp một nền tảng cho việc xây dựng và đào tạo các mô hình học máy. TensorFlow được sử dụng rộng rãi trong nhiều lĩnh vực khác nhau, bao gồm học máy, trí tuệ nhân tạo, khoa học dữ liệu, …

Keras là một thư viện mã nguồn mở cho học máy. Thư viện này cung cấp một giao diện lập trình ứng dụng (API) đơn giản và dễ sử dụng để xây dựng các mô hình học máy. Keras được xây dựng trên nền tảng TensorFlow, vì vậy nó có thể tận dụng khả năng tính toán mạnh mẽ của TensorFlow

### IDE: Google Colab

Google Colab hay Google Colaboratory, là một nền tảng trực tuyến miễn phí cho phép người dùng viết và thực thi mã Python trong môi trường Jupyter Notebook. Nó được phát triển bởi Google và cho phép truy cập vào tài nguyên tính toán mạnh mẽ như GPU và TPU.

Các tính năng nổi bật của Google Colab:

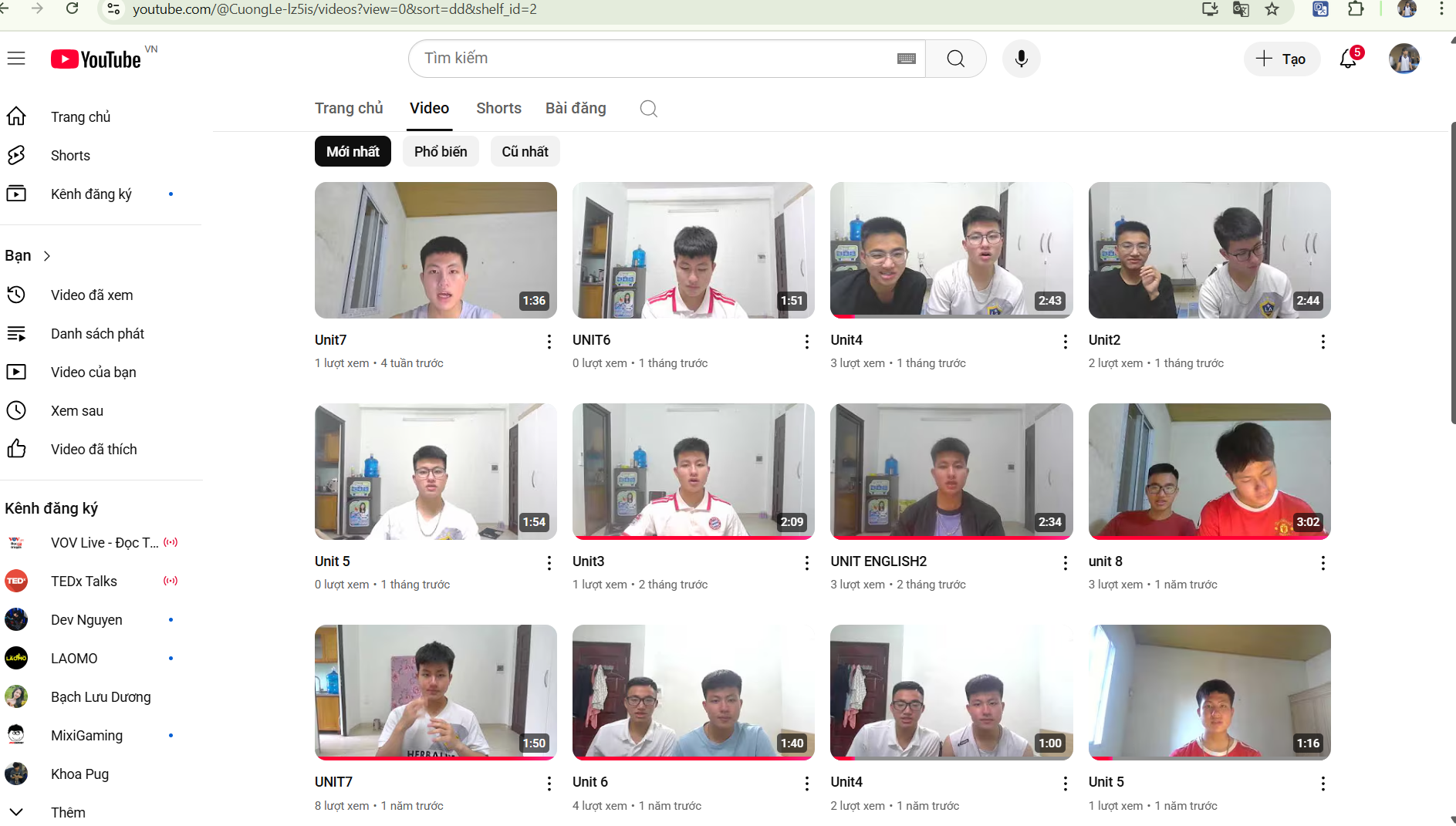
* Môi trường trực tuyến: Người dùng có thể truy cập và làm việc từ bất kỳ đâu chỉ cần có kết nối Internet mà không cần cài đặt phần mềm.
* Hỗ trợ GPU và TPU: Google Colab cung cấp quyền truy cập miễn phí vào GPU và TPU, giúp tăng tốc độ tính toán cho các tác vụ như học máy và phân tích dữ liệu.
* Chia sẻ dễ dàng: Người dùng có thể dễ dàng chia sẻ sổ ghi chép của mình với người khác thông qua liên kết, tương tự như Google Docs.
* Tích hợp với Google Drive: Tài liệu và mã nguồn có thể được lưu trữ trực tiếp trên Google Drive, giúp dễ dàng quản lý và truy cập.
* Thư viện phong phú: Hỗ trợ nhiều thư viện Python phổ biến như TensorFlow, PyTorch, NumPy và Pandas, giúp người dùng thực hiện các dự án phức tạp một cách dễ dàng.

## Các bước thực hiện:

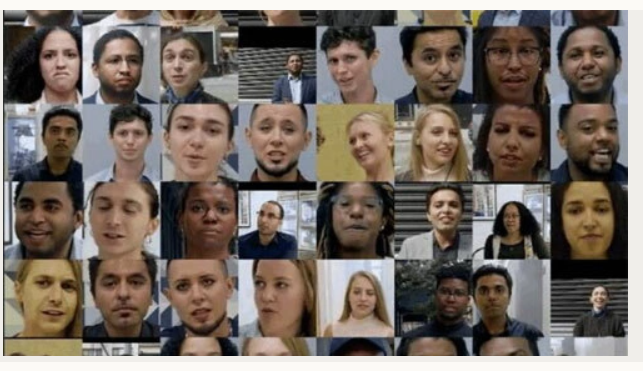
### Thu thập dữ liệu

Dữ liệu đầu vào là các video thật và video Deepfake. Cụ thể:

* **Video thật**: Được thu thập từ các nguồn như video cá nhân



* **Video giả mạo (Deepfake)**: Sử dụng từ các bộ dữ liệu công khai chuyên biệt như [FaceForensics++](https://github.com/ondyari/FaceForensics)



Hai tập dữ liệu này được lưu trữ riêng biệt trong các thư mục khác nhau (VideoReal, VideoFake) để phục vụ cho quá trình trích xuất ảnh và gán nhãn.

### Tiền xử lý dữ liệu

Trước khi đưa ảnh vào CNN, cần thực hiện các bước tiền xử lý:

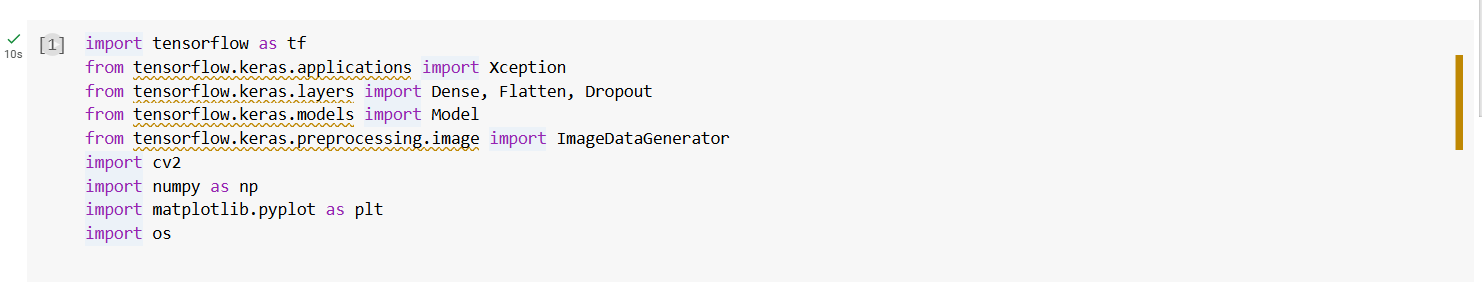
Trích xuất các khung hình từ video (mỗi giây lấy một số lượng ảnh nhất định).

Chuyển đổi ảnh về cùng kích thước (ví dụ: 224x224 pixels).

Chuẩn hóa dữ liệu (normalize pixel values về khoảng [0,1] hoặc [-1,1]).

## 3.4.Các bước thực hiện

* Bước 1:Import thư viện cần thiết



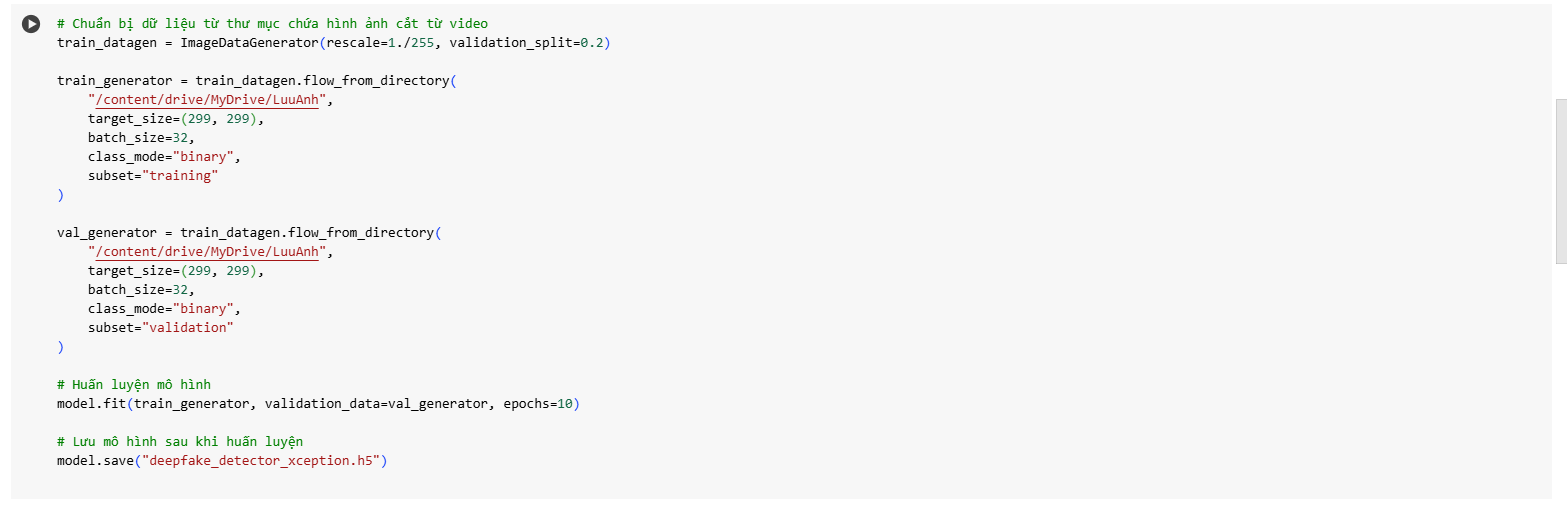
* Bước 2:Trích xuất khung hình từ video và lưu vào thư mục ảnh

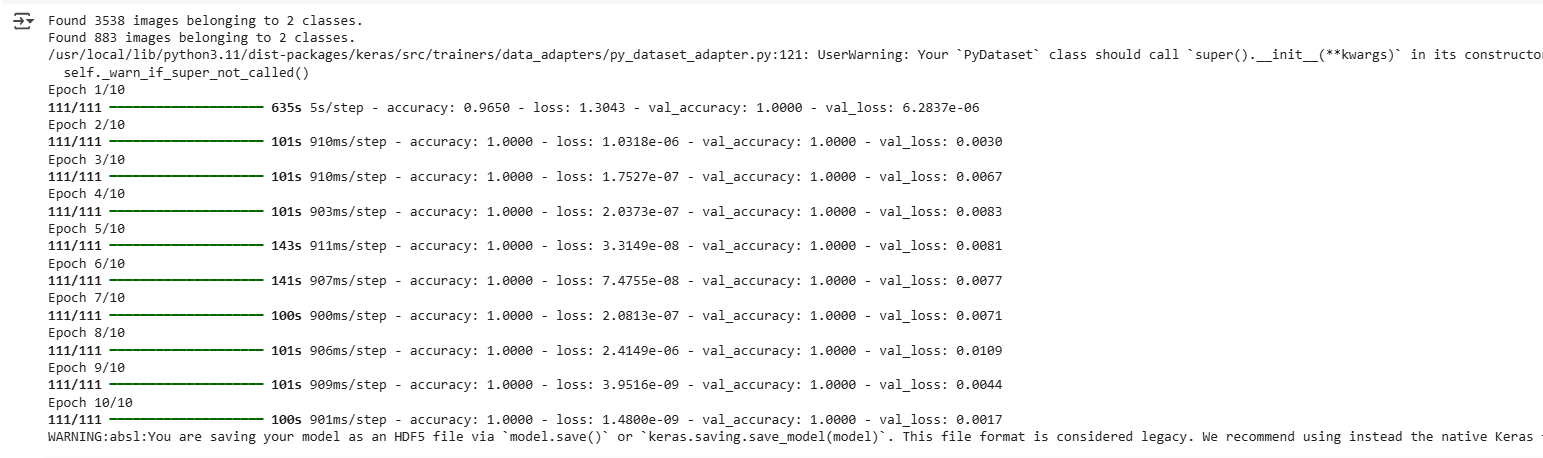


* Bước 3:Khởi tạo mô hình Xception và xây dựng tầng phân loại



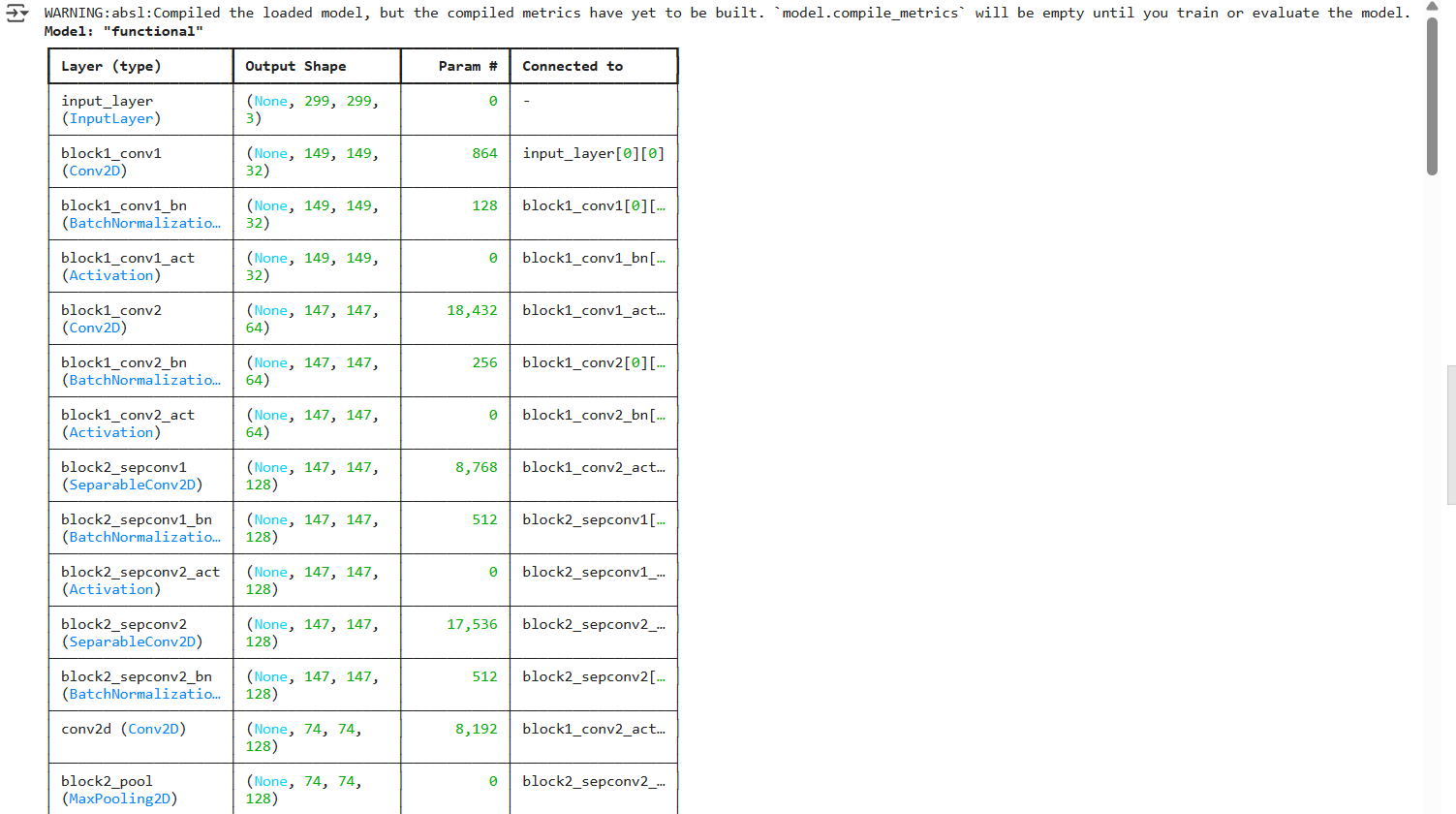
* Bước 4:Tiền xử lý và nạp dữ liệu ảnh huấn luyện mô hình

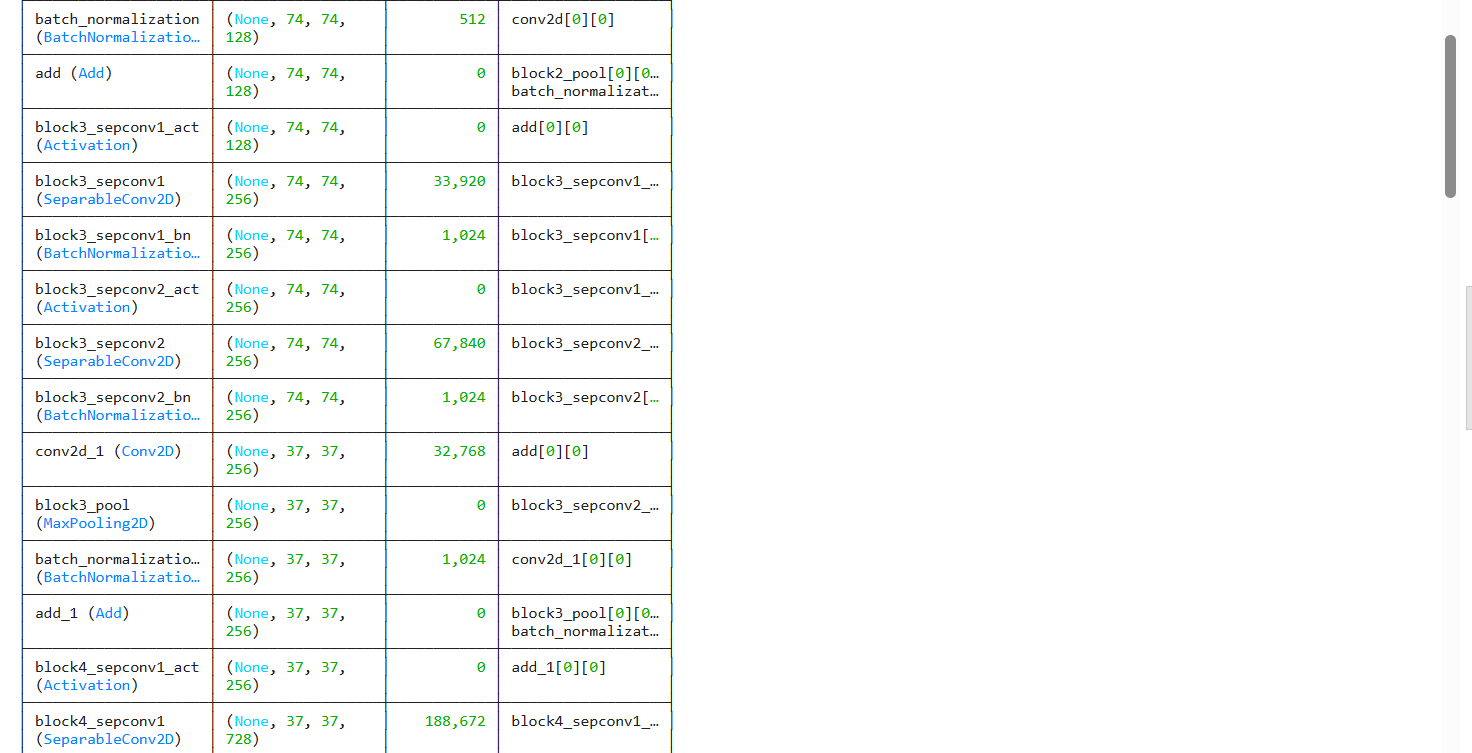


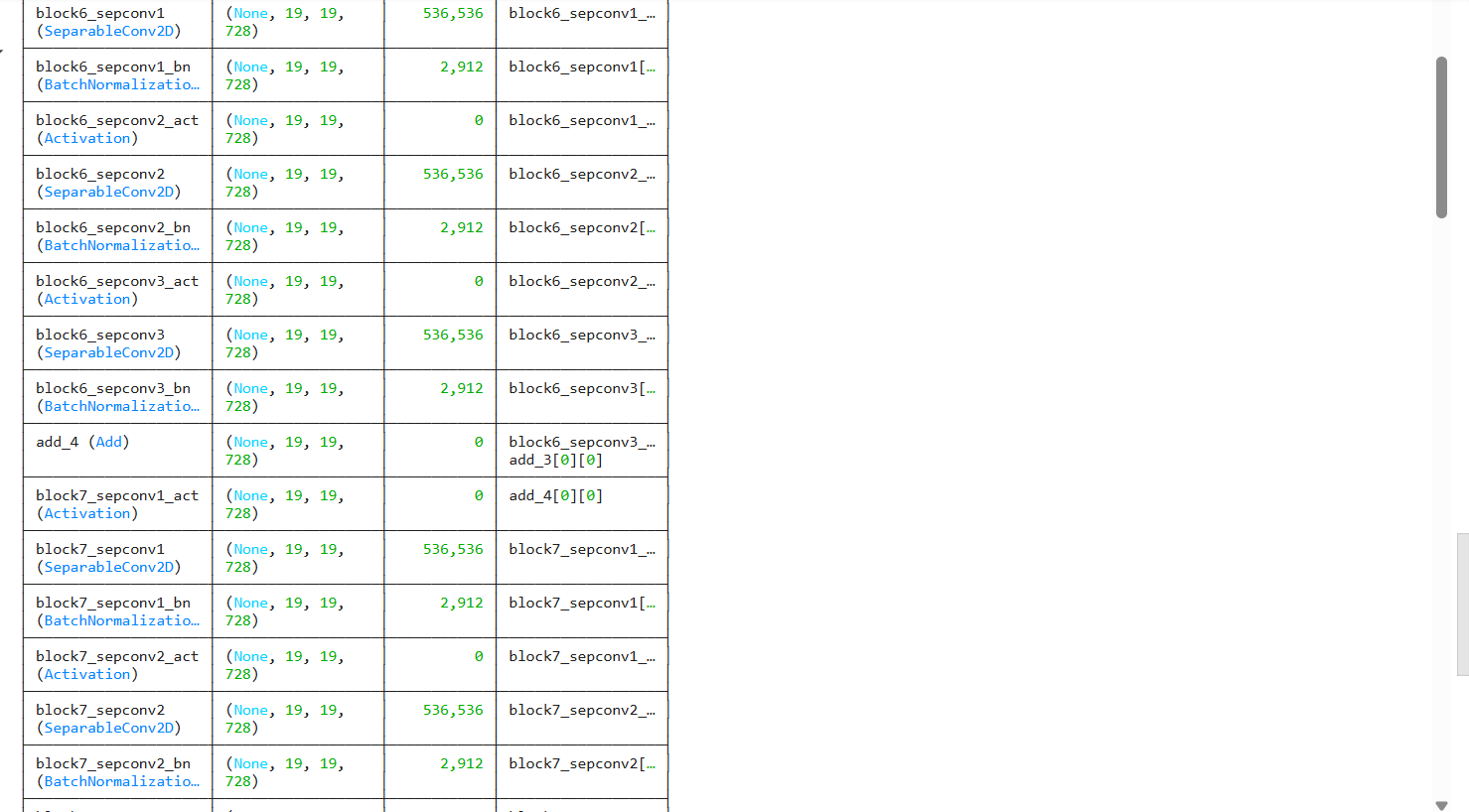


* Bước 5:Tải mô hình đã huấn luyện để sử dụng

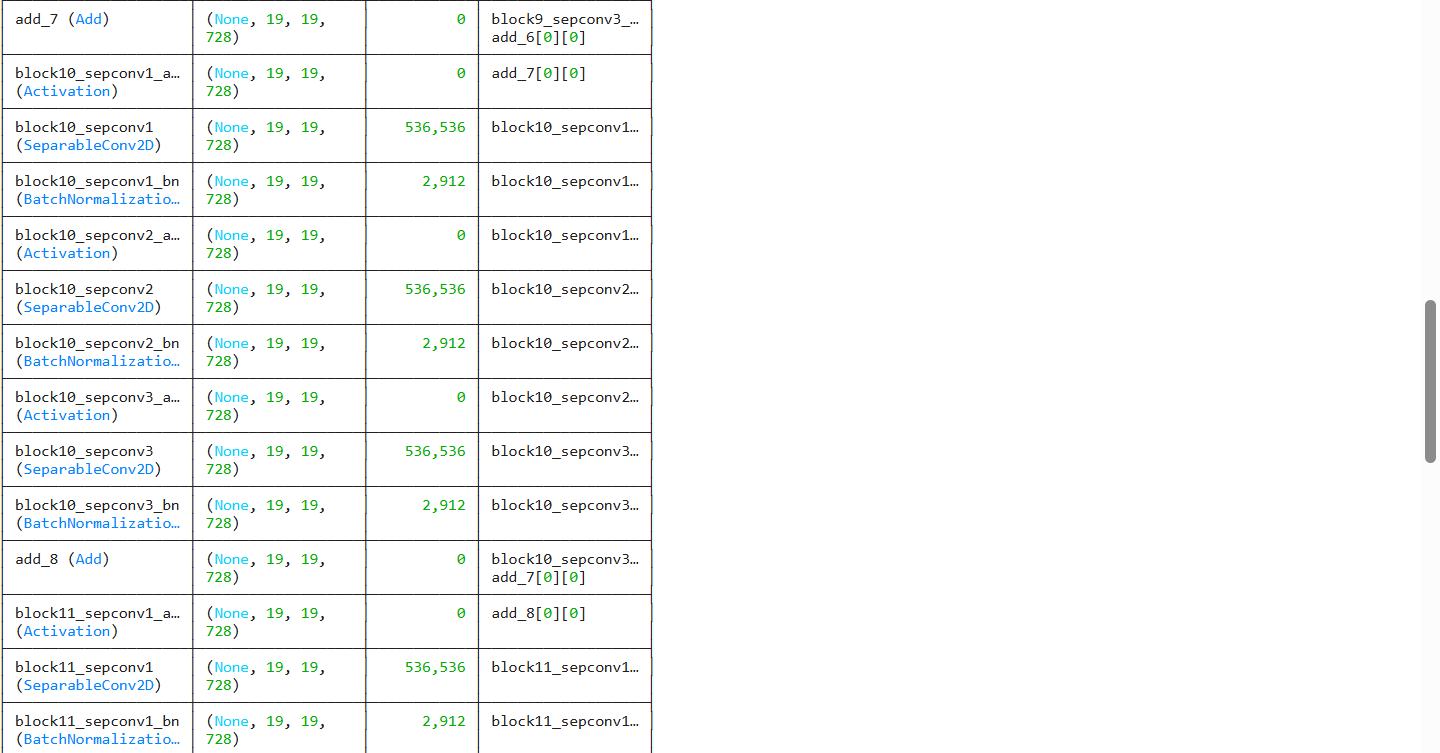




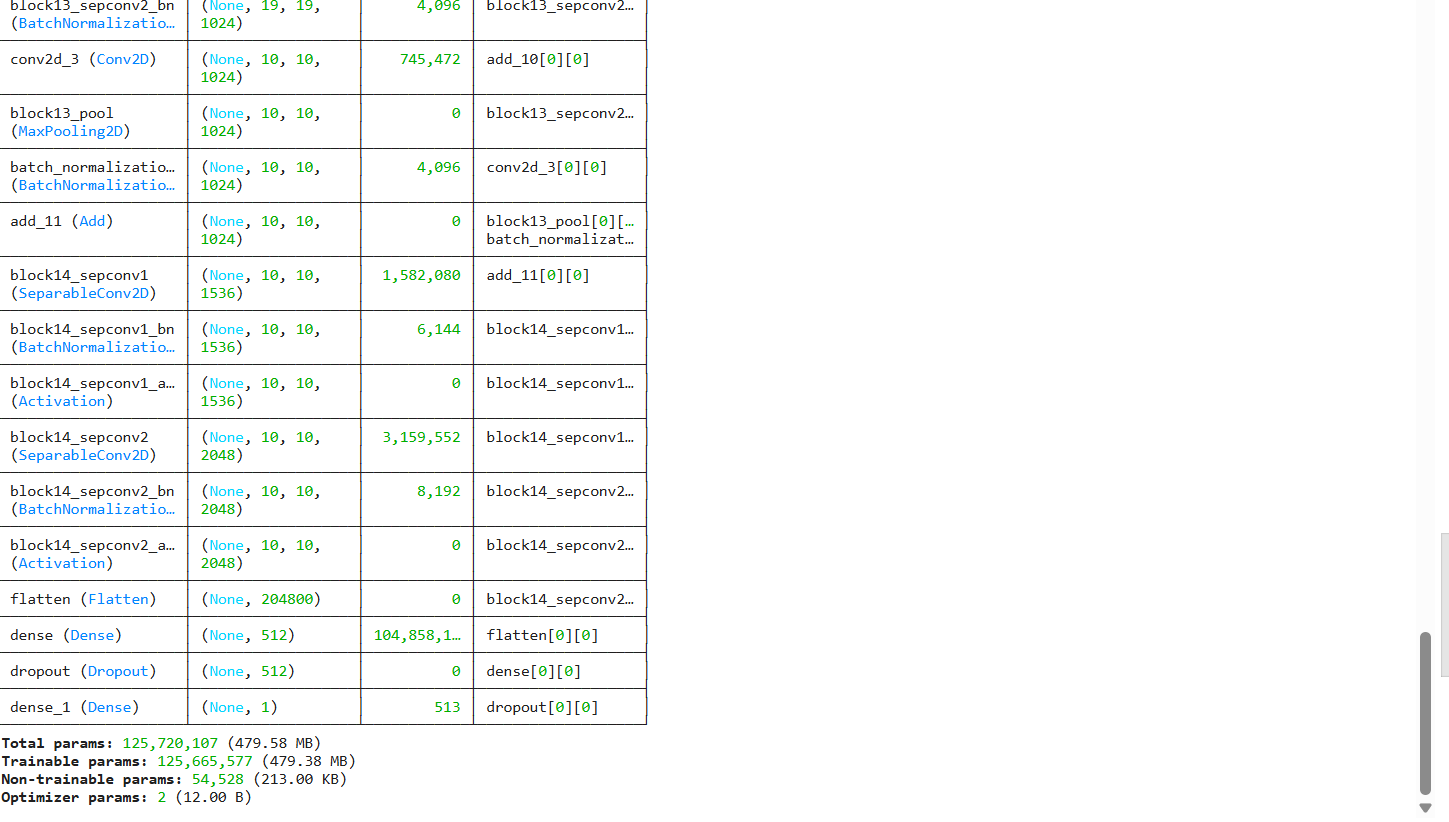




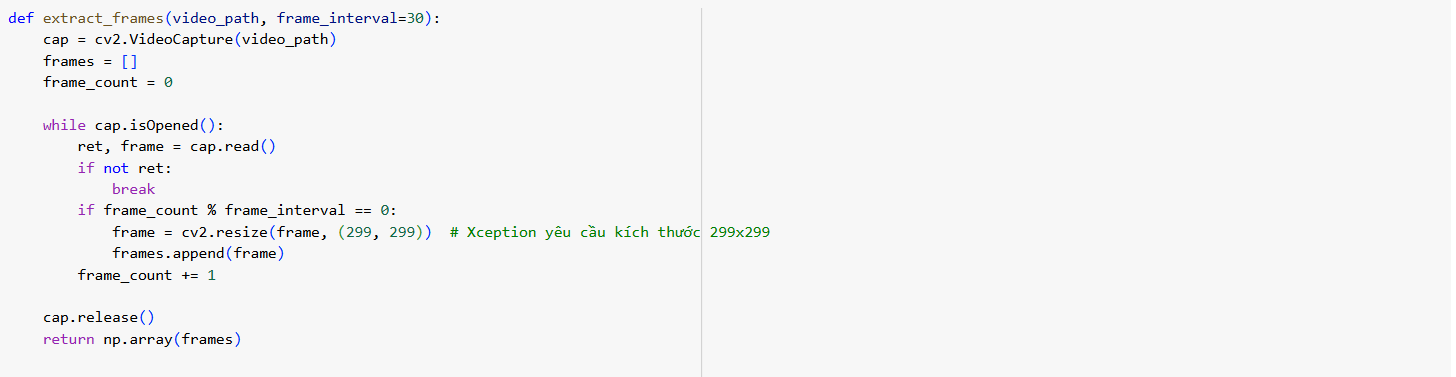




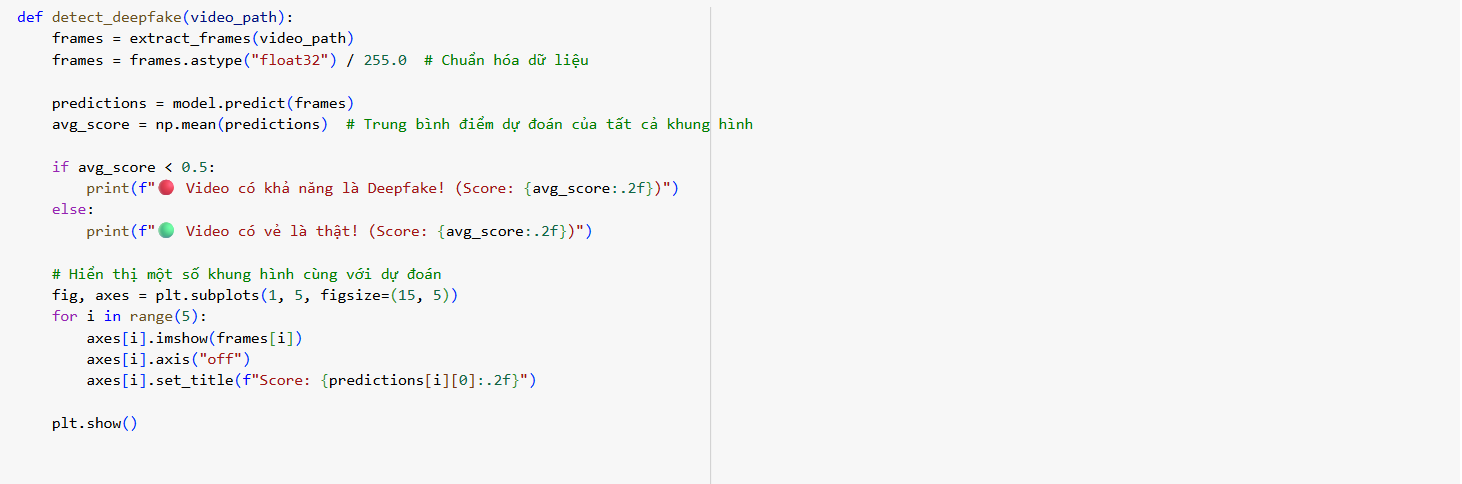




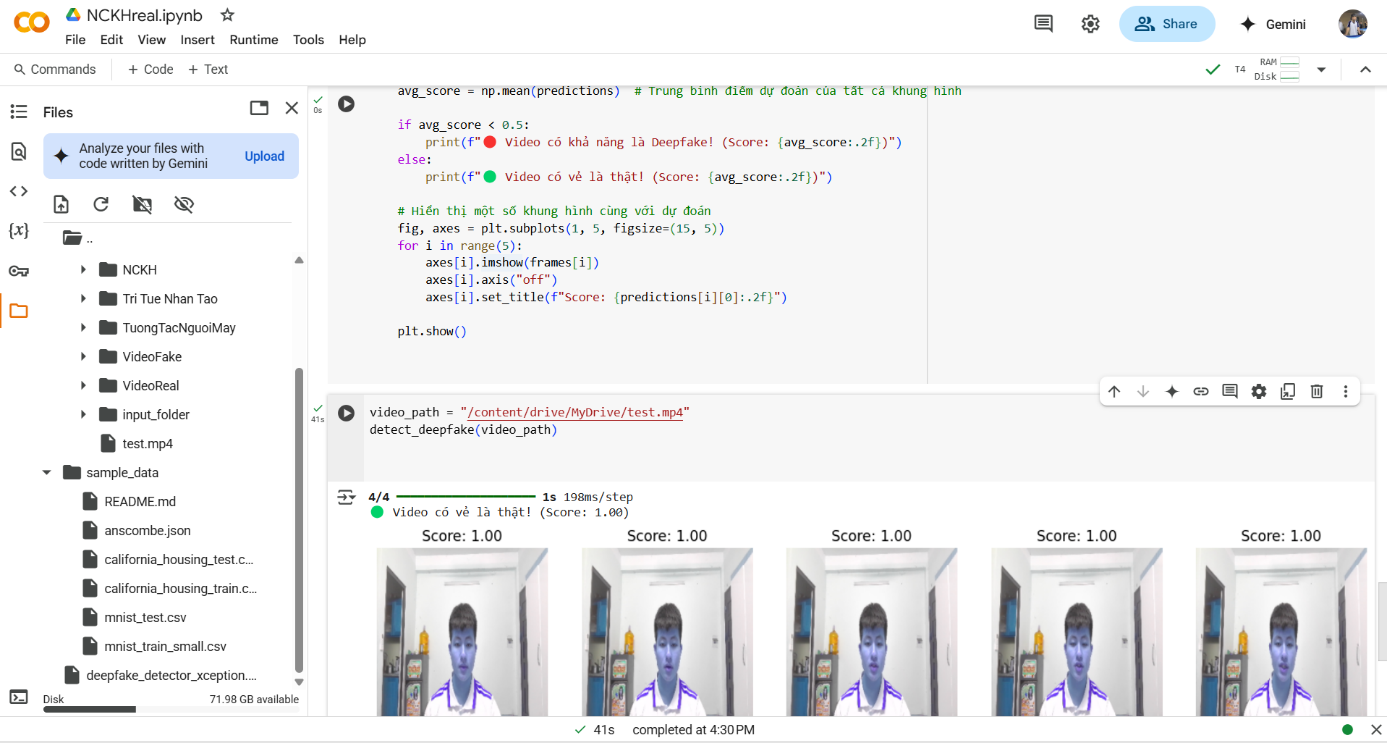
* Bước 6:Hàm trích xuất khung hình từ video đầu vào để dự đoán



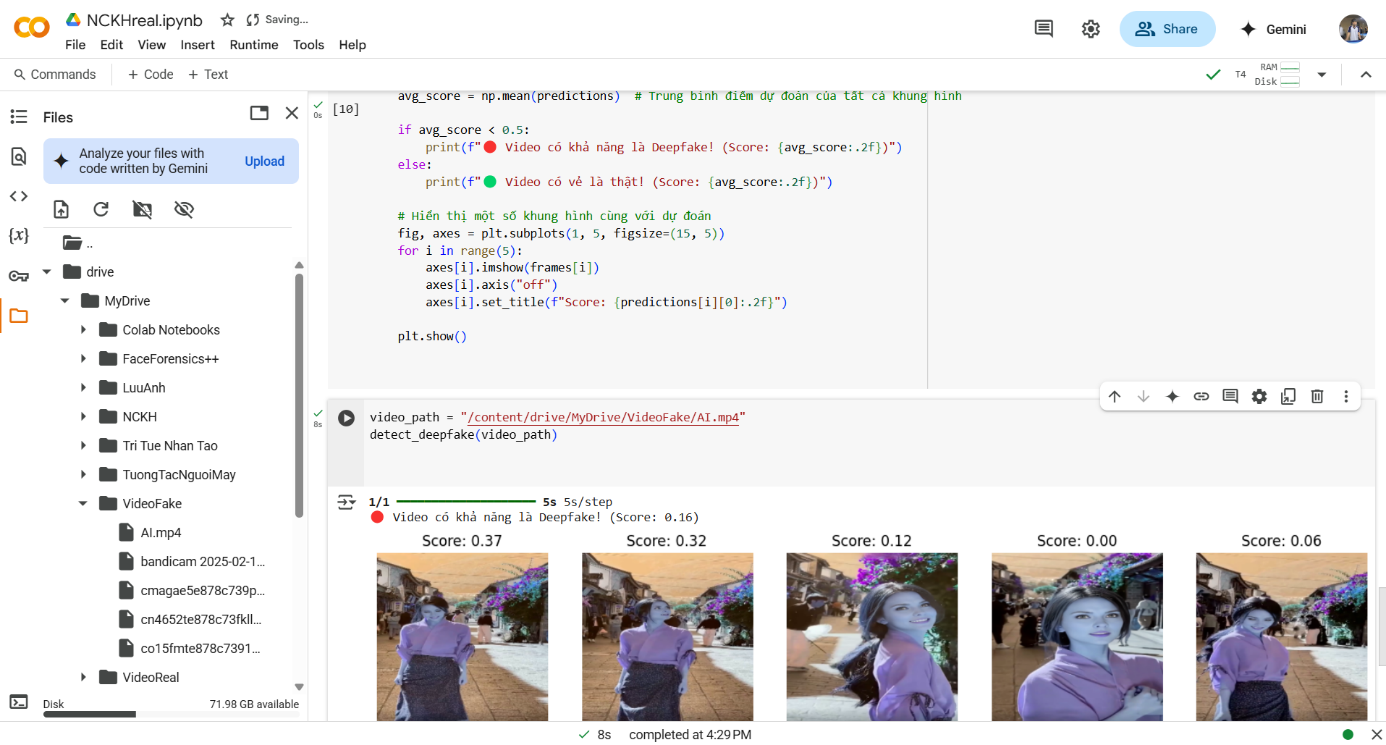
* Bước 7:Hàm phát hiện Deepfake từ video đầu vào



* Bước 8:Thử nghiệm mô hình
* Hiển thị một số khung hình kèm điểm dự đoán.



*Thử nghiệm video người thật*



*Thử nghiệm video do AI Tạo ra*

# THỬ NGHIỆM VÀ ĐÁNH GIÁ HỆ THỐNG

Độ chính xác của mô hình được đo lường dựa trên khả năng phân biệt video thật và video Deepfake trên các tập dữ liệu chuẩn hóa. Các chỉ số đánh giá chính bao gồm:

* Accuracy (Độ chính xác tổng thể): Phần trăm số video được phân loại đúng.
* Precision (Độ chính xác khi phát hiện Deepfake): Tỷ lệ video được dự đoán là giả mạo thực sự là giả.
* Recall (Khả năng phát hiện video giả): Tỷ lệ video giả được nhận diện đúng.
* F1-score: Thước đo tổng hợp giữa Precision và Recall.

Bên cạnh độ chính xác, tốc độ xử lý cũng là yếu tố quan trọng, đặc biệt đối với các ứng dụng thực tế như kiểm duyệt nội dung video trực tuyến. Tốc độ nhận diện Deepfake được đánh giá qua:

* Thời gian xử lý trung bình trên mỗi khung hình (ms/frame).
* Thời gian xử lý toàn bộ video (s/video).
* Khả năng chạy trên GPU so với CPU.

Các mô hình Deepfake ngày càng trở nên tinh vi, đòi hỏi hệ thống phát hiện phải đủ mạnh để nhận diện các video giả mạo chất lượng cao. Để đánh giá khả năng chống lại các kỹ thuật làm giả, hệ thống sẽ được thử nghiệm trên:

* Deepfake với độ phân giải cao (4K).
* Deepfake có điều chỉnh ánh sáng và màu sắc để che giấu dấu vết giả mạo.
* Deepfake sử dụng AI nâng cao như StyleGAN, First Order Motion Model.
* Deepfake áp dụng phương pháp làm mờ, nén video để tránh bị phát hiện.

# KẾT LUẬN

Trong bối cảnh công nghệ Deepfake đang phát triển nhanh chóng và ngày càng tinh vi, việc xây dựng một hệ thống có khả năng phát hiện video giả là vô cùng quan trọng nhằm bảo vệ sự minh bạch thông tin và an toàn cộng đồng. **Qua quá trình triển khai và thử nghiệm, hệ thống AI được phát triển trong nghiên cứu này đã đạt được độ chính xác trên 90%** khi phát hiện video giả mạo. Đây là kết quả tích cực, khẳng định tiềm năng lớn của trí tuệ nhân tạo trong việc hỗ trợ con người phát hiện và xử lý các nội dung sai lệch trên môi trường số.

Tuy nhiên, độ chính xác hiện tại không phải là đích đến cuối cùng. **Các mô hình học sâu (Deep Learning) vẫn còn nhiều điểm cần cải tiến để bắt kịp sự biến hóa không ngừng của công nghệ Deepfake.** Các kỹ thuật giả mạo mới thường được tối ưu hóa để vượt qua các bộ phát hiện hiện có, tạo ra một cuộc chạy đua liên tục giữa "người tạo giả" và "người phát hiện giả". Do đó, để duy trì hiệu quả lâu dài, hệ thống AI cần được cập nhật thường xuyên với các dữ liệu huấn luyện mới, cải tiến về kiến trúc mạng nơ-ron, cũng như nâng cao khả năng nhận diện các chi tiết vi mô trong video – nơi các dấu hiệu giả mạo có thể ẩn giấu rất tinh vi.

Bên cạnh yếu tố kỹ thuật, việc phát hiện và ngăn chặn video giả cũng cần được nhìn nhận dưới góc độ toàn diện hơn. Không một giải pháp công nghệ đơn lẻ nào có thể đủ sức đối đầu với vấn đề này nếu thiếu sự phối hợp đồng bộ từ các lĩnh vực khác. **Sự hợp tác giữa công nghệ AI, hệ thống pháp luật và truyền thông đóng vai trò đặc biệt quan trọng.** Cụ thể, AI đóng vai trò phát hiện và cảnh báo sớm, luật pháp cần xây dựng các quy định rõ ràng về xử phạt hành vi sử dụng Deepfake vào mục đích xấu, trong khi các cơ quan truyền thông cần kiểm duyệt nội dung, cung cấp thông tin chính xác và nâng cao nhận thức cho người dân về các thủ đoạn của tin giả.

Từ ba trọng tâm này – công nghệ, mô hình học sâu, và sự phối hợp liên ngành – có thể thấy rằng việc phòng chống video giả không phải là một nhiệm vụ có thể giải quyết trong thời gian ngắn. Nó đòi hỏi một chiến lược dài hạn, kết hợp giữa đổi mới công nghệ, xây dựng chính sách phù hợp và giáo dục cộng đồng.

# TÀI LIỆU THAM KHẢO

* **FaceForensics++**
* Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2019). *FaceForensics++: Learning to detect manipulated facial images*. arXiv preprint arXiv:1901.08971.
* Website: <https://github.com/ondyari/FaceForensics>
* **DeepFake Detection Challenge Dataset (DFDC)**
* Dolhansky, B., et al. (2020). *The Deepfake Detection Challenge Dataset*. Facebook AI Research.
* Website: <https://ai.facebook.com/datasets/dfdc>
* **Celeb-DF (Celeb-Deepfake Dataset)**
* Li, Y., Sun, P., Qi, H., Lyu, S. (2020). *Celeb-DF: A Large-scale Challenging Dataset for DeepFake Forensics*. IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
* Website: http://www.cs.albany.edu/~lsw/celeb-deepfakeforensics.html
* **DF-TIMIT Dataset**
* Korshunov, P., & Marcel, S. (2018). *Deepfakes: a new threat to face recognition? Assessment and detection*. arXiv preprint arXiv:1812.08685.
* Website: <https://www.idiap.ch/en/dataset/df-timit>

1. Korshunov, P., & Marcel, S. (2019). *DeepFakes: a new threat to face recognition? Assessment and detection*. arXiv preprint arXiv:1812.08685.
2. Afchar, D., Nozick, V., Yamagishi, J., & Echizen, I. (2018). *MesoNet: a compact facial video forgery detection network*. 2018 IEEE International Workshop on Information Forensics and Security (WIFS), pp. 1-7.