

University of Engineering and Technology

NHẬN BIẾT NGŨ NGŨ KÝ HIỆU

Sử dụng mạng thần kinh

Thành viên:

Nguyễn Đam Trường - 18021333

Nguyễn Hoàng Vũ - 18020453

GIỚI THIỆU

Bài thuyết trình, trình bày những vấn đề nào ?

1

- 1. Tình trạng thực tế
- 2. Mục tiêu

2

- 1. Cấu trúc mạng thần kinh
- 2. So sánh các cấu trúc

3

- 1. Kết quả
- 2. Định hướng

MỞ ĐẦU

+ Theo WHO, Gần 2,5 tỷ người trên toàn thế giới (hoặc cứ 4 người thì có 1 người) sẽ sống với tình trạng khiếm thính ở một mức độ nào đó vào năm 2050.

+ Theo một số báo cáo, gần 1,3 triệu người khuyết tật câm điếc và khiếm thính (theo thống kê chưa đầy đủ ở nước ta) và con số thực tế có thể còn lớn hơn.



NGÔN NGỮ KÝ HIỆU



Các quốc gia khác nhau có giao tiếp cử chỉ ký hiệu gần giống nhau

Những ngôn ngữ có liên quan trong cùng một họ ngôn ngữ có thể được mong đợi chia sẻ từ 36% đến 79% từ vựng cơ bản (ngôn ngữ ký hiệu của Mỹ và Pháp, những ngôn ngữ được xem là có liên quan trong cùng một họ ngôn ngữ chia sẻ khoảng từ 61% từ vựng cơ bản

Tuy nhiên, ngôn ngữ ký hiệu của Mỹ và Anh không có quan hệ gần nhau

không chung một họ ngôn ngữ giống nhau, chúng chỉ có 31% cùng nguồn gốc trong từ vựng cơ bản. Cấu trúc ngôn ngữ ký hiệu thay đổi theo không gian và thời gian.

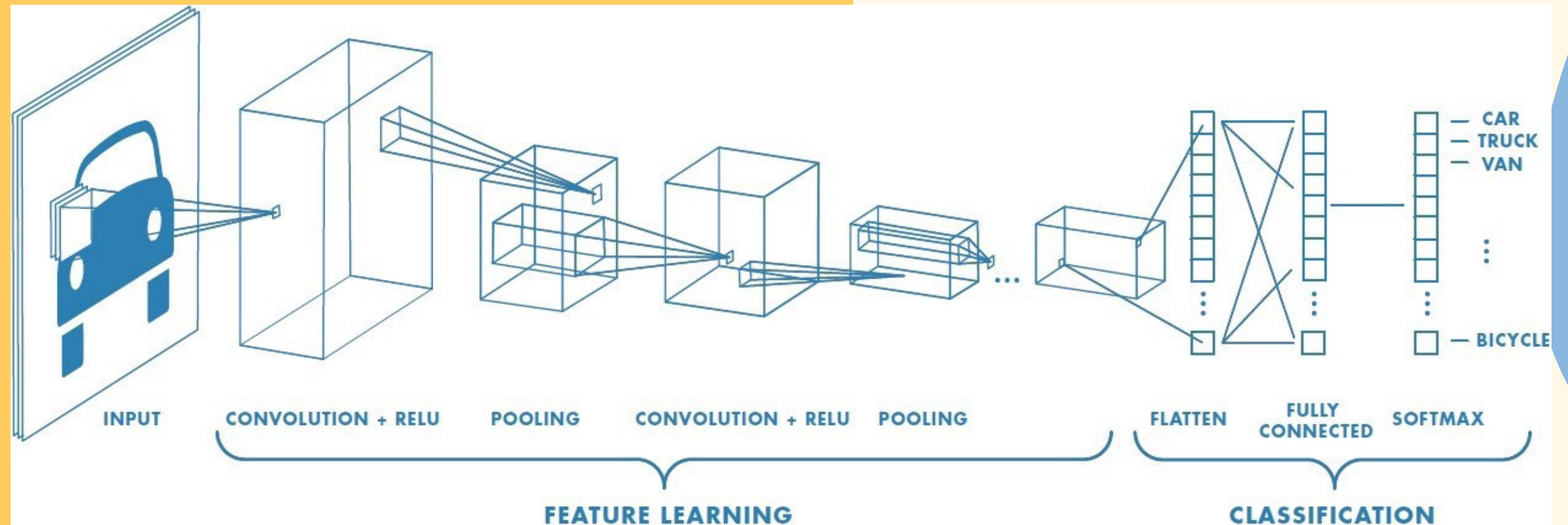
Phương pháp dựa trên vision

Một máy ảnh sẽ ghi lại chuyển động của cơ thể người, đặc biệt là cử chỉ tay để đưa ra ngôn ngữ ký hiệu.

Phương pháp dựa trên cảm biến

Các chuyển động của bàn tay và ngón tay trong thời gian thực có thể được theo dõi bằng cách sử dụng cảm biến chuyển động Leap.

MẠNG CNN

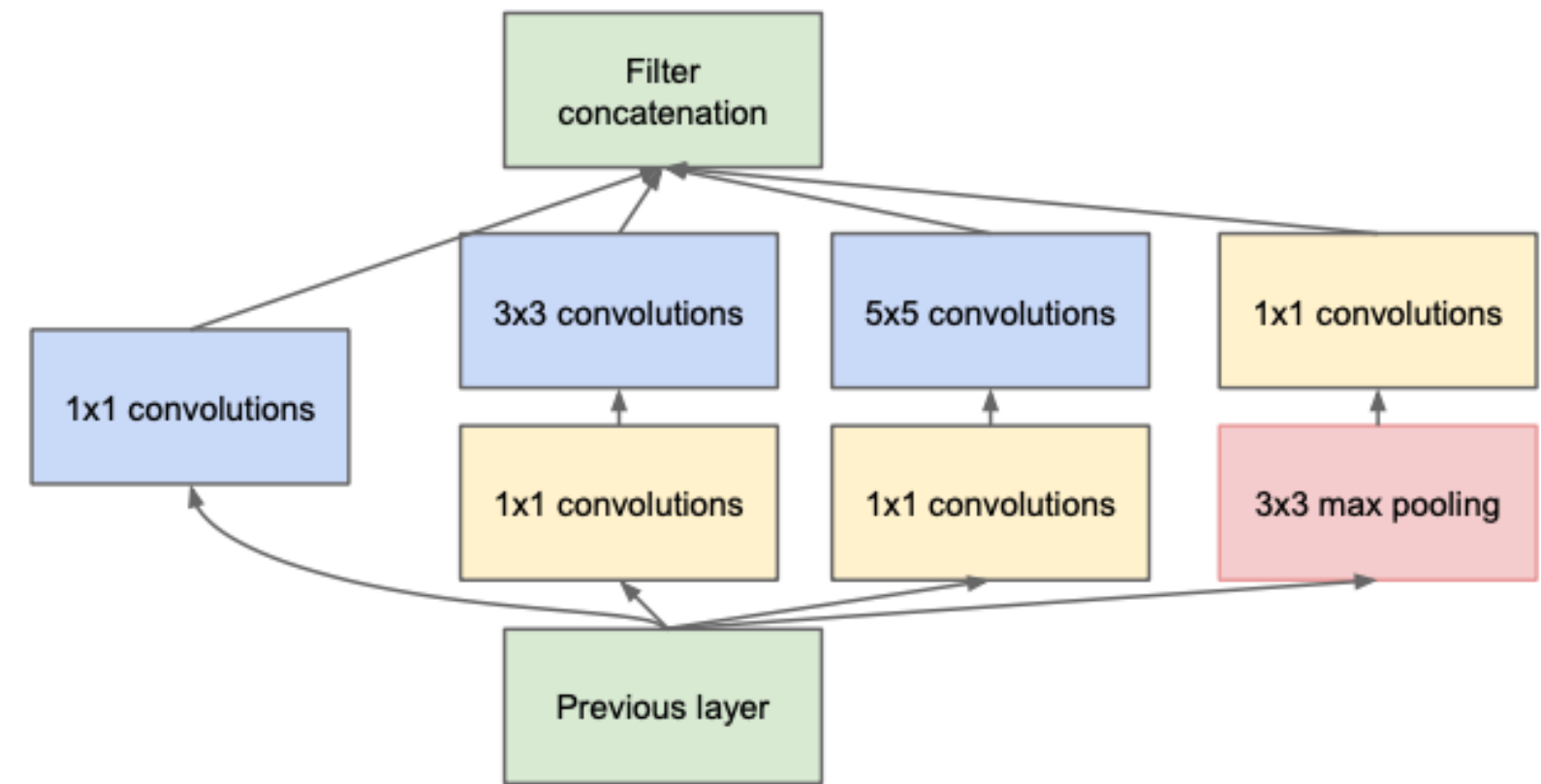


CNN rất hiệu quả trong việc giảm số lượng các tham số mà không làm giảm chất lượng của các mô hình. Hình ảnh có kích thước cao (vì mỗi pixel được coi là một tính năng) phù hợp với khả năng được mô tả ở trên của CNN.

CNN giữ nguyên dạng hình ảnh không gian 2D.

Tất cả các lớp của CNN có nhiều bộ lọc tích hợp hoạt động và quét ma trận tính năng hoàn chỉnh và thực hiện giảm kích thước. Điều này cho phép CNN trở thành một mạng rất phù hợp và phù hợp để phân loại và xử lý hình ảnh.

INCEPTION V3 MODEL



Tại sao lại cần Transfer Learning Google Inception Net ?

- Nhiều lớp đối tượng hơn làm cho sự khác biệt giữa các lớp khó hơn. Ngoài ra, một mạng lưới thần kinh chỉ có thể chứa một lượng thông tin hạn chế, có nghĩa là nếu số lượng lớp trở nên lớn thì có thể không có đủ trọng số để đối phó với tất cả các lớp. Điều này làm giảm độ chính xác của mô hình sau khi thêm nhiều lớp và dữ liệu đào tạo trong bộ dữ liệu.
- Sự khác biệt chính giữa các Transfer Learning và CNN thông thường là các khối khởi đầu. Chúng liên quan đến việc kết hợp cùng một tenxơ đầu vào với nhiều bộ lọc và kết hợp kết quả của chúng. Ngược lại, CNN thông thường thực hiện một thao tác tích chập duy nhất trên mỗi tenxơ.

ĐẦU VÀO



Bảng chữ cái ASL (American sign language) bao gồm phân loại 51 kí tự.

Huấn luyện mô hình phân loại ngôn ngữ kí hiệu với bộ dataset ASL_and_same_words được lấy từ Kaggle, gồm 4000 ảnh cỡ 200x200x3 cho mỗi kí tự.

Mô hình Inception V3

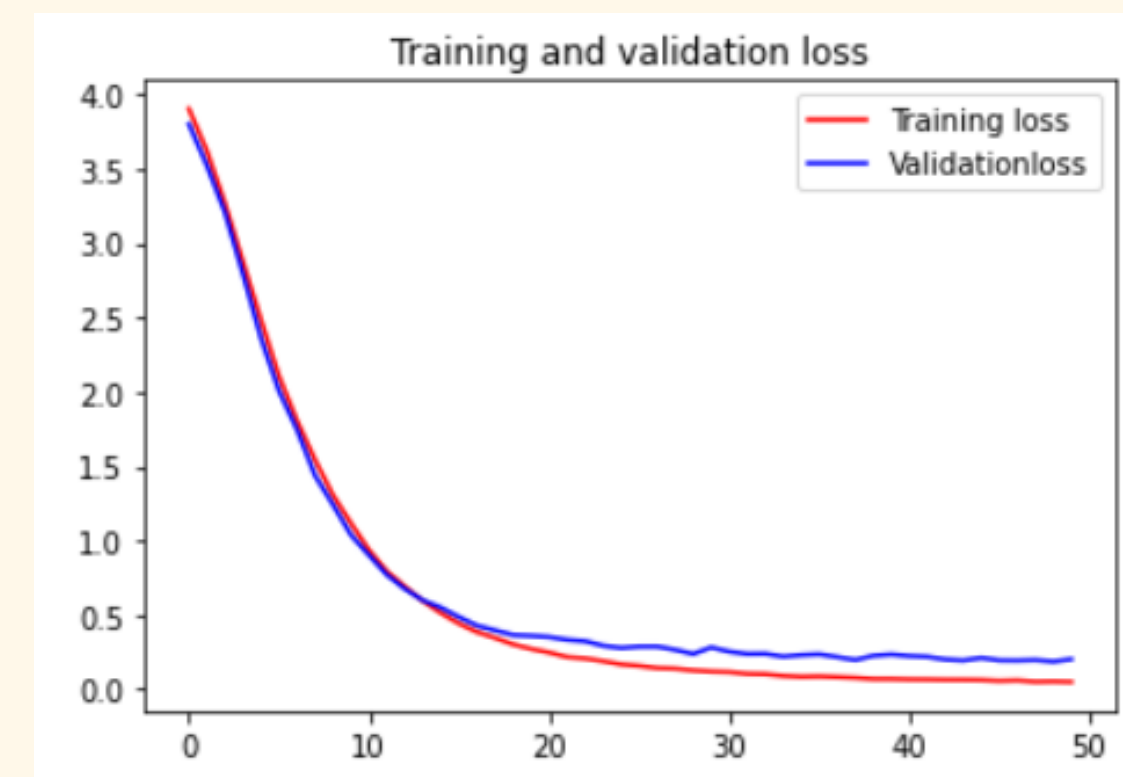
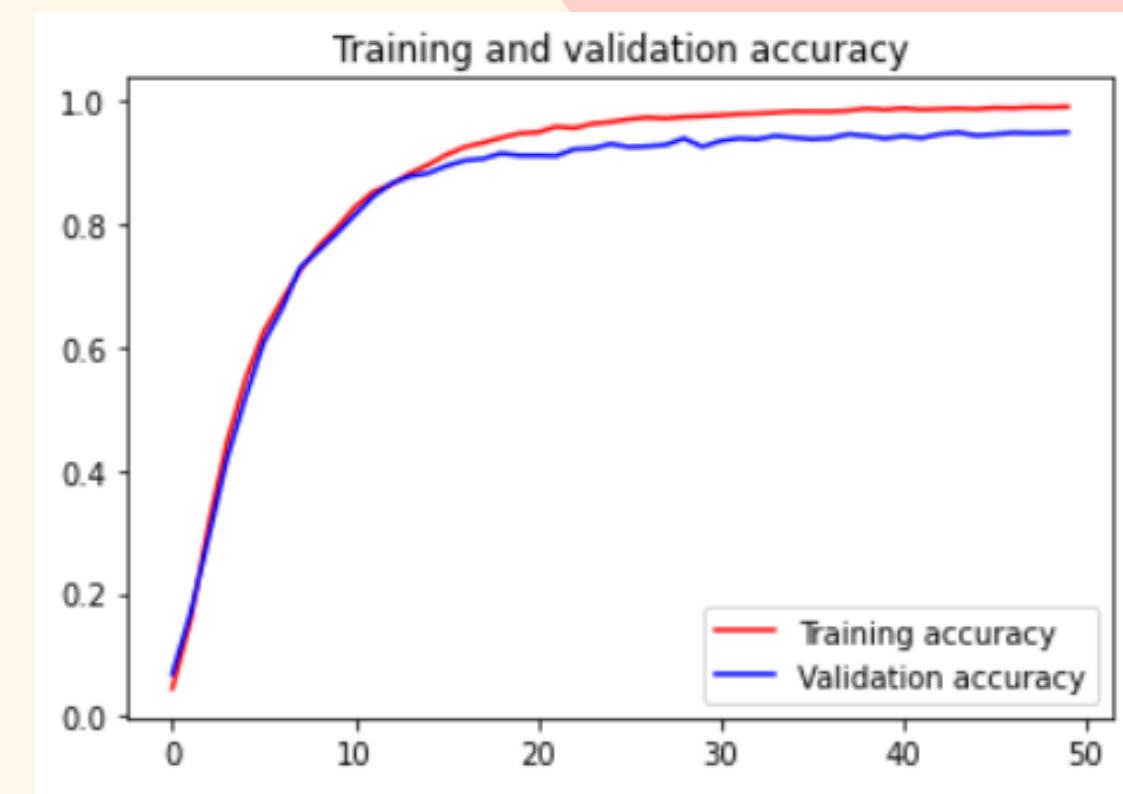
Mô hình đã được huấn luyện với dataset ImageNet gồm 1000 lớp với hơn 1 triệu ảnh huấn luyện.

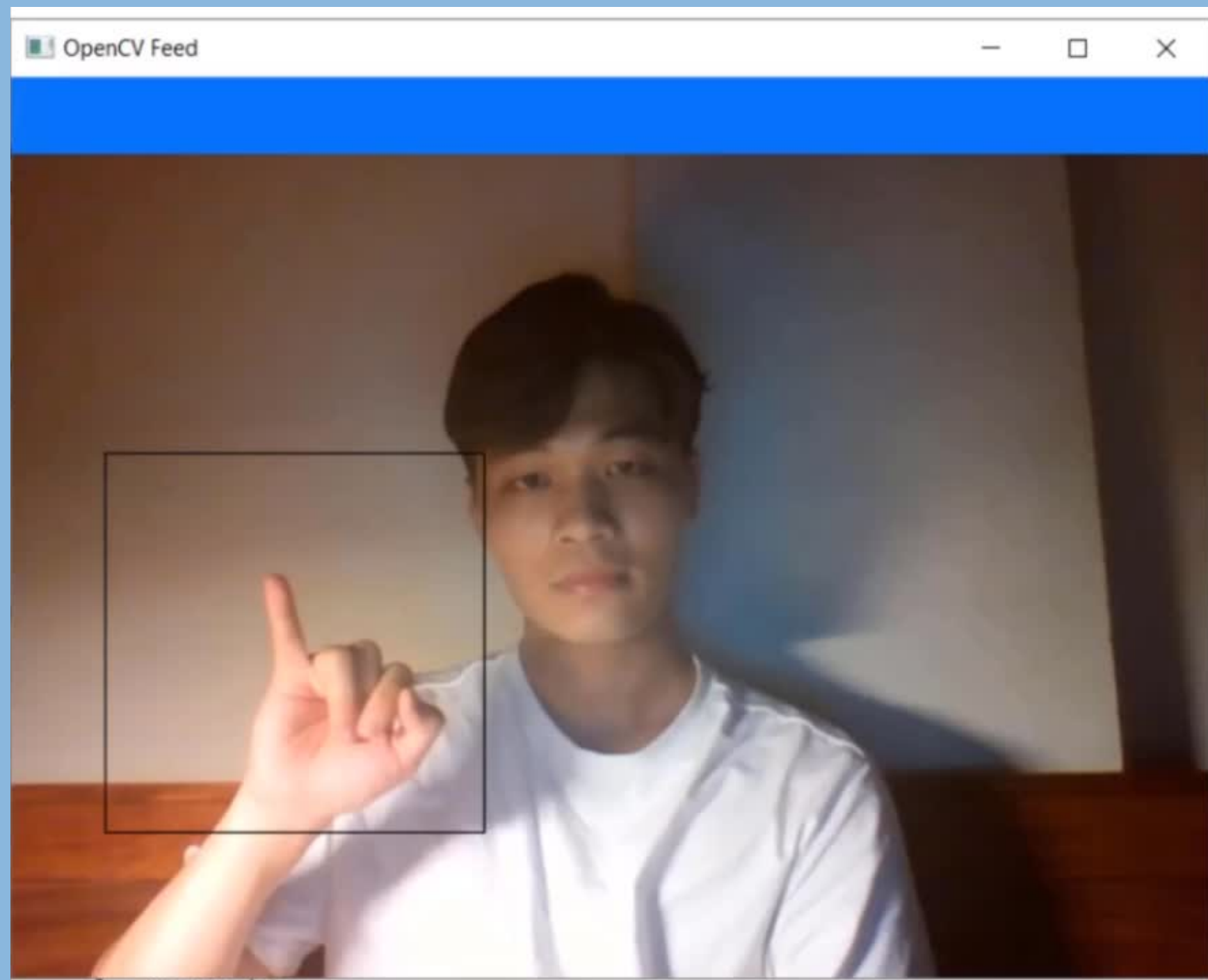
Mô hình InceptionV3 sau khi được lấy từ keras được xếp thêm một lớp dense với 1024 neuron, và đầu ra gồm 51 neuron sử dụng hàm kích hoạt softmax tương ứng với 51 kí tự. Mạng InceptionV3 bao gồm 314 lớp, trong đó chúng tôi đóng băng 249 lớp đầu tiên và thực hiện huấn luyện với tất cả các lớp còn lại

ĐẦU RA

Dữ liệu	Mô hình	Chính xác
DATA	3-CNN	Train: 79,25% Test: 64,84%
DATA	Inception V3	Train: 98,99% Test: 94,88%

- Mô hình V3 Inception đã giúp tăng độ chính xác của mô hình.





VIDEO KẾT QUẢ

CẢM ƠN ĐÃ LẮNG
NGHE PHẦN TRÌNH BÀY