TRƯỜNG ĐẠI HỌC BÁC KHOA HÀ NỘI

VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

-------------------------------------------------------

Báo cáo môn học máy

Đề tài: Phân loại cảm xúc trong video

***Giáo viên hướng dẫn****:* **Thầy Ngô Văn Linh**

**Sinh viên:**

1. Phạm Đức Tuệ MSSV: 20164432
2. Phạm Duy Tiên MSSV: 20164038
3. Hồ Xuân Cường MSSV: 20160537
4. Trịnh Khánh Linh MSSV: 2016
5. Nguyễn Thị Diệu Mơ MSSV: 2016

**Hà nội**, ngày 14/06/2020

Tóm tắt nội dung

Trong khuôn khổ đề tài môn học máy, cũng nhưng mong muốn bắt đầu đi vào tìm hiểu mô hình Deep Learning cho bài toán xử lý ảnh. Bọn em lựa chọn đề tài **Nhận diễn cảm súc.** Báo cáo bọn em trình bày gồm 3 phần chính như sau:

* Phần 1: Đặt vấn đề
* Phần 2: Phương pháp tiếp cận (sử dụng MTCNN để phát hiện khuôn mặt sau đó dùng mô hình CNN để phân loại cảm súc)
* Phần 3: Kết quả đạt được

Mục lục

Phần 1: Đặt vấn đề

Convolutional Neural Network (CNNs – Mạng nơ-ron tích chập) là một trong những mô hình Deep Learning tiên tiến. Nó giúp cho chúng ta xây dựng được những hệ thống thông minh với độ chính xác cao như hiện nay. Như hệ thống xử lý ảnh lớn như Facebook, Google hay Amazon đã đưa vào sản phẩm của mình những chức năng thông minh như nhận diện khuôn mặt người dùng, phát triển xe hơi tự lái hay drone giao hàng tự động. CNN được sử dụng nhiều trong các bài toán nhận dạng các object trong ảnh…

Bài toán Nhận diện cảm súc là một bài toán phân loại ảnh, và có ứng dụng lớn trong bài toán Computer Vision. Phương pháp tiếp cận sử dụng MTCNN để phát hiện khuôn mặt có chính xác cao. (Recall = 82.60%, Precsion = 95,24%) và mô hình CNN cơ bản để phân loại cảm xúc.

Phần 2: Phương pháp tiếp cận

2.1 Bộ dữ liệu fer2013

Bọn em sử dụng bộ dữ liệu fer2013(<https://www.kaggle.com/deadskull7/fer2013>) được tạo bởi Pierre-Luc Carrier and Aaron Courville, được công khai trên Kaggle. Dữ liệu bao gồm 35.887 ảnh xám, kích thước 48 \* 48 với 7 cảm súc đã được gán nhãn.

**0:** -4593 ảnh- Angry  
**1:** -547 ảnh - Disgust  
**2:** -5121 ảnh - Fear  
**3:** -8989 ảnh - Happy  
**4:** -6077 ảnh - Sad  
**5:** -4002 ảnh - Surprise  
**6:** -6198 ảnh – Neutral

2.2 Kiến trúc tổng quan

emotion

image

image

image

Emotion Classification

Feature Extraction

Face Detection

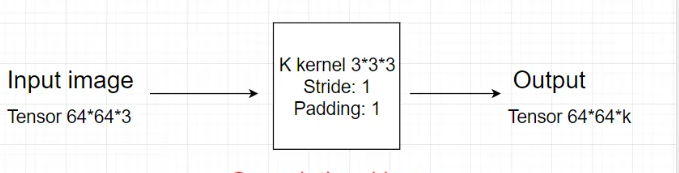
* Face Detection: MTCCN phát hiện khuôn mặt => thu được tọa độ bouding box, ảnh từ bouding box được chuyển sang dạng ảnh xám, resize về 48\*48
* Feature Extraction & Emotion Classification: Ảnh được trích rút đặc trưng thông qua các lớp tích chập sau đó tính toán xác suất vào một trong 7 lớp.

2.3 Mô hình CNN sử dụng

2.3.1 Về mô hình CNN (Convolutional Neural Network)

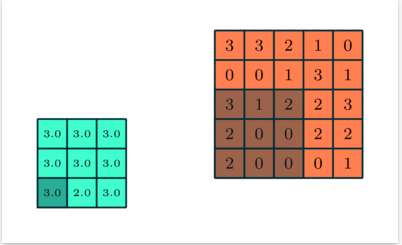
Trong mạng ANN bình thường một ảnh đầu vào (32 \* 32 \* 3) (RGB) cần tới 32 \* 32 \*3 = 3072 weight (fully connection), nhìn có vẻ không tốt nến kích thước bức ảnh lớn, kích thước tham số sẽ rất lớn. CNN không giống như ANN giảm thiểu số lượng tham số thông qua các lớp ConvNet. Mạng CNN gồm 3 layer chính: Convolutional Layer, Pooling Layer, Fully-Connected Layer. Mạng CNN gồm 3 tính chất: **các trường tiếp nhận cục bộ** (local receptive field), **trọng số chia sẻ** (shared weights)**, tổng hợp** (pooling).

* Convolution Layer: gồm các kernel với kích thước nhỏ (thường 3 \* 3) là cửa sổ trượt, mỗi kernel khác nhau ta sẽ học được những đặt trưng khác nhau của ảnh, vì ứng với kernel output là một matrix nên độ sâu output là bằng số kernel



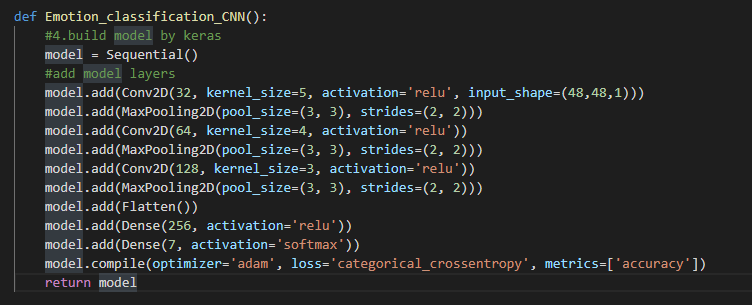
Convolution Layer dùng để học những đặc trưng ảnh

* Pooling layer: Thường được đặt giữa các lớp Convolution layer, để giảm kích thước dữ liệu mà vẫn giữ được các thuộc tính quan trọng. Có 2 loại Pooling là MaxPooling và Average Pooling



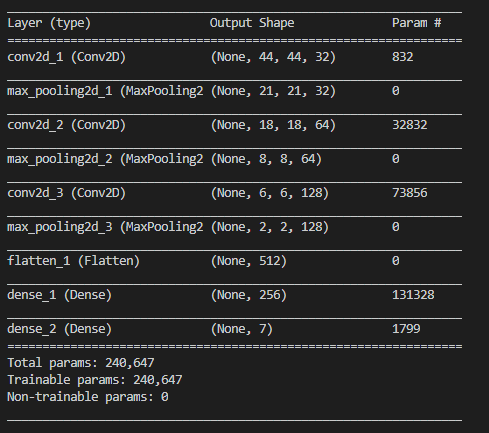
* Fully-connection Layer: Hoạt động như ANN bình thường, cuối cùng thu được xác xuất vào các lớp

2.3.2 Mô hình CNN đề xuất



Mô hình sử dụng hàm xác suất softmax, hàm lỗi categorical\_crossentropy, optimizer adam cho quá trình điều chỉnh tham số.

* Kích thước tham số qua các lớp conv:



2.4 MTCNN

Phần 3: Kết quả đạt được

Phần 4: Tài liệu tham khảo