

INTELIGENCIA ARTIFICIAL

PROYECTO

GUILLERMO RIVO VÁZQUEZ



GUILLERMO RIVO VÁZQUEZ

RESPONSABLE DEL DESARROLLO DE LA APLICACIÓN

Fecha: 14, 9, 2021

DEFINICIÓN DEL PROBLEMA

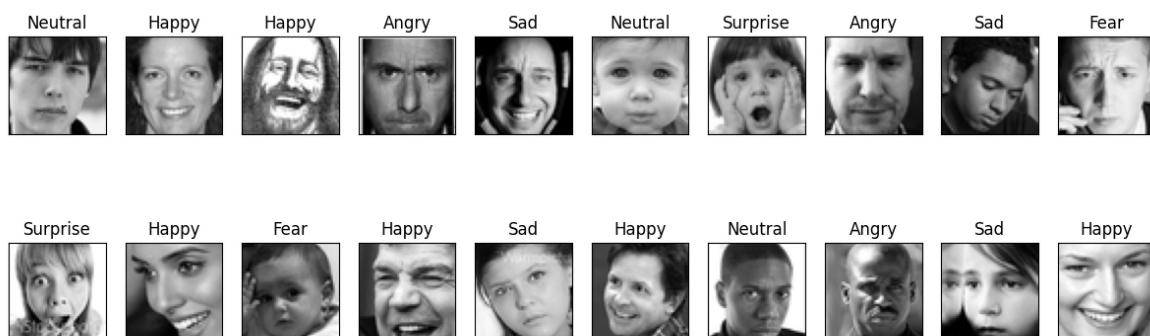
El objetivo de la aplicación es reconocer la cara del usuario a través de la grabación de la webcam del ordenador, y la clasificación de la emoción del usuario a partir de su cara. El dataset de Kaggle que voy a utilizar es el siguiente:

<https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge>

Es un challenge en el cual el objetivo era la clasificación de expresiones faciales con la máxima precisión posible.

DATOS DEL PROBLEMA

Los datos que voy a utilizar son los del dataset anterior, se trata de un archivo csv que contiene las tags con la respues de qué emoción presenta el registro (columna 'emotion'), el uso previsto de cada resgistro, si se va usar para entrenamiento o para test (columna 'usage') y por último la columna 'pixels' que contiene varios ints separados por espacios, cada uno con el valor en blanco y negro de cada pixel de la imagen (la imagen tiene 2304 pixels, 48x48).





LIMPIEZA Y TRATAMIENTO DE DATOS

La limpieza de datos que he hecho en este proyecto consiste en lo siguiente:

- Para los datos del csv:
 - Separar el dataset en varias arrays de numpy según el valor de la columna 'usage'.
 - Convertir la columna 'emotion' a un int, que puede tomar hasta 7 valores distintos del 0 al 6, su equivalencia se define en el diccionario 'emotions'.
 - La columna 'pixels', que en el csv original contenía una string con los valores de los pixels separados por espacios, se transforma en un numpy array de dimensiones 48x48.
 - El valor de grayscale en el csv original se definía con un int que podía tomar valores entre 0 y 255, no es recomendable introducir esto en la red neuronal, entonces se divide por 255 para que todos los valores se encuentren entre el 0 y el 1.
 - Se hace one-hot encoding de la variable emoción, por la misma razón de antes, el input de la red neuronal debe estar entre el 0 y el 1, el resultado es que por en cada fila la columna emoción va a pasar a ser una lista de 7 elementos en la cual todos van a ser 0 excepto el de la posición = emoción, que será 1.
 - Una modificación que al final acabé descartando era unificar las emociones 0 y 1, (correspondientes a angry y disgust respectivamente), pero esto no resultó ser beneficioso para el accuracy del modelo.
- Para el input de la webcam:
 - La imagen RGB que se recibe desde la webcam se convierte a grayscale.
 - Se aplica un clasificador Haar Cascade, en este caso utilizo un modelo creado para reconocer caras, este modelo devuelve una lista con 4 datos, (x, y, w, h), esto permite definir un rectángulo sobre la imagen e ignorar todo lo que se encuentre fuera del mismo, esto es beneficioso porque permite que al modelo le pasemos una imagen en la que la mayoría del espacio está ocupada por la cara a clasificar.
 - Como en el caso anterior se hace one-hot encoding de la variable emotion.
 - Cambiamos la forma de la imagen para que se adapte a los inputs del modelo, que deben ser de la forma (48, 48, 1).



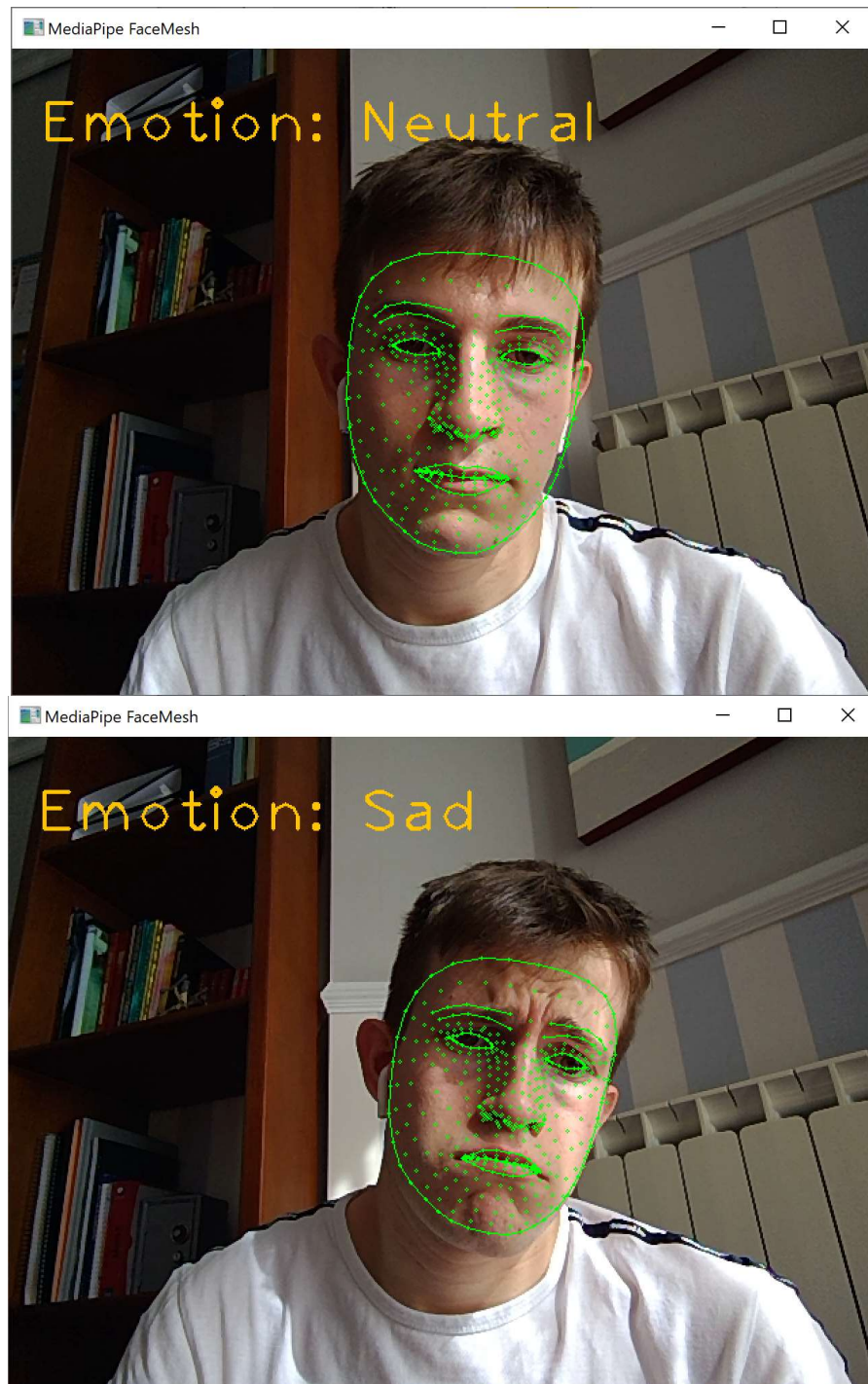
OBSERVACIONES SOBRE LOS DATOS

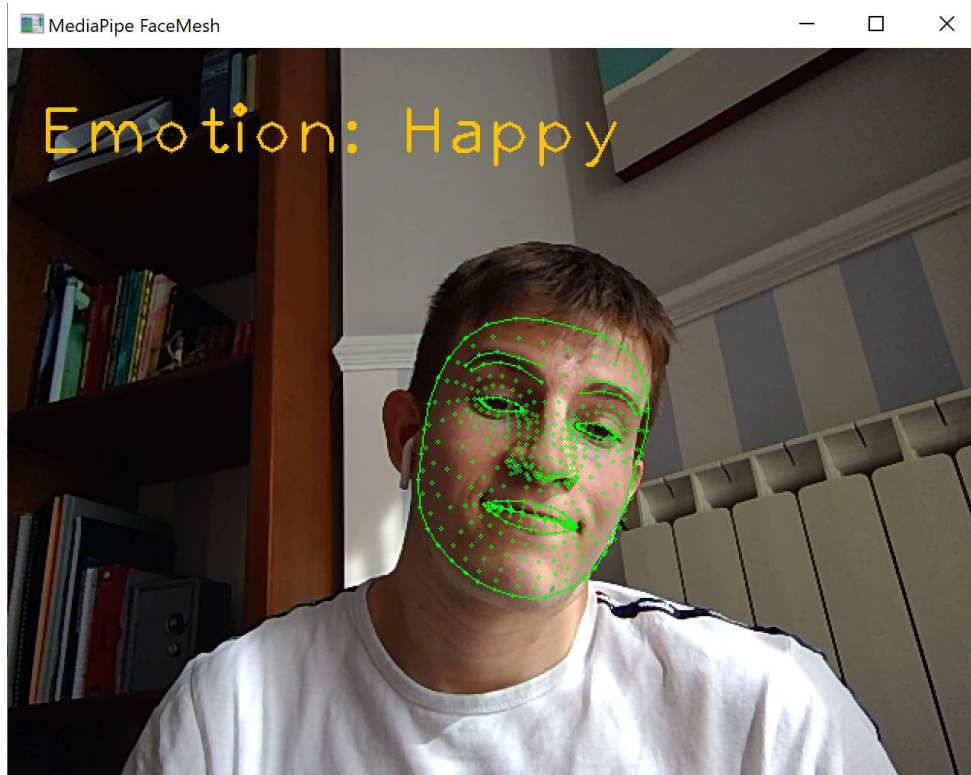
Los datos para este proyecto son bastante numerosos, tenemos un tamaño del dataset de entrenamiento de alrededor de 30.000 imágenes, sin embargo hay una emoción que es muy minoritaria con respecto al resto, disgust, y happy es claramente mayoritaria, esta disparidad se puede traducir en un bias en la clasificación, además hay algunas imágenes cuyo tag no parece corresponderse con la emoción que yo interpreto tiene la persona en ese momento.

MODELO PARA LA RESOLUCIÓN DEL PROBLEMA

Para resolver este problema se va a usar una red neuronal con varias capas de convolución 2D, la operación de convolución consiste en desplazar una serie de filtros ventana (kernels) sobre una imagen para operar sobre ambos y así obtener información, es una técnica muy utilizada en el tratamiento de imagen.

REPRESENTACIÓN DE RESULTADOS





CONCLUSIONES

Este proyecto se basa en el machine learning, utiliza una red neuronal para realizar una tarea de clasificación de imagen. El programa tiene un claro bias, sabe reconocer mucho mejor la emoción happy que otras emociones mucho menos documentadas como disgust, además falla mucho más si la imagen de webcam presenta una mala iluminación.