

Proyecto Final

Estados cognitivo-afectivos en línea

Asignatura: Reconocimiento de Patrones

Contenido

Introducción	1
1. Recopilación de videos y etiquetado.....	1
2. Preprocesamiento de imágenes	2
3. Modelos a entrenar	3
4. Protocolo de entrenamiento y prueba	4
5 Dataset sintético de CAS.....	5
6 Entrega del proyecto	5

Introducción

Para el contexto educativo, en lugar de las emociones básicas (Ekman 1992), se consideran más adecuados los estados que enfatizan la conexión entre los procesos cognitivos y las emociones.

Descripción del proyecto:

En este proyecto vas a entrenar dos modelos ligeros utilizando un dataset de expresiones faciales sintéticas que representan 5 CAS: Joy, Neutral, Distraction, Fatigue, Confusion. Para ello contarás con un dataset sintético con splits de entrenamiento, validación y prueba.

Además, vas a generar un conjunto de datos de prueba con expresiones reales de cada uno de los integrantes del equipo.

A continuación, se describen las etapas del proyecto:

1. Recopilación de videos y etiquetado

1.1 Participantes y videos

Cada uno de los integrantes del equipo debe crear un video para cada uno de los 5 CAS. La longitud de los videos resultantes no debe ser mayor a 3 minutos.

Instrucciones para la recopilación de videos

- Iniciar una sesión en Zoom,
- Preparar el clip que corresponda dependiendo del CAS que se quiera obtener (joy, neutral, distraction, fatigue y confusion).
- Comenzar a grabar en Zoom
- Iniciar la reproducción del clip.
- Detener la grabación cuando considere que cuenta con 2 o 3 ejemplos del CAS objetivo)

Clips para inducir CAS

- Joy

Título: 🤖 3 MEMES DE PROGRAMACIÓN explicados por un DESARROLLADOR DE SOFTWARE

Enlace: <https://www.youtube.com/watch?v=MaqlceTDYU4>

Comenzar a verlo a partir de 1min con 35 segundos

- **Neutral, Distraction and Fatigue**

Título: Cómo crear un documento de Especificación de Requerimientos de Software

Enlace: <https://youtu.be/AotyBHVkp8I?si=PIWGVtwmBFYixvXo>

Comenzar a verlo desde el inicio y avanzar hasta completar los videos de los tres CAS objetivo: neutral, distraction y fatigue (recopilar videos en ese orden para que fatigue sea el último)

- **Confusion**

Título: Diffusion Models | Paper Explanation | Math Explained

Enlace: <https://youtu.be/HoKDTa5jHvg?si=C3cUJAU7HMwGninj>

Verlo desde el minuto 15:00 al minuto 20:00

Características de la grabación:

- Utilizar la calidad de grabación por default de Zoom.
- Mantener webcam frontal, fondo relativamente neutro (Puede ser el difuminado de Zoom).

Consejos para buenas tomas

- Alejar la cámara para que la toma no salga tan centrada en el rostro.
- Para distracción, pueden voltear hacia cualquier lugar distinto a la cámara, pueden ver el celular, pueden fingir que alguien les habla y ustedes voltean.
- Para confusión, se pueden fruncir el ceño, se puede levantar una ceja o ambas, se puede colocar una mano o ambas sobre el mentón o la mejilla. Si la confusión se convierte en frustración, se puede cubrir la cara con ambas manos o colocar una o ambas manos sobre las sienes.
- Para fatiga, se puede expresar somnolencia parpadeando lentamente (como si nos estuviera ganando el sueño), podemos quedarnos dormidos, podemos colocar una o ambas manos en el menton y cerrar los ojos, podemos bostezar.
- Para Joy, combinar sonrisas sutiles con sonrisa plenas como si les hubieran contado una broma.

1.2 Etiquetado de los videos

Segmentación en clips

- Divide los videos en ventanas de tiempo (p.ej. 2–4 segundos)
- Marca las ventanas donde se supone que el CAS inducido está más claro.

Etiquetado

- Etiqueta cada clip con la condición para la que fue diseñado el clip; p.ej. “confusion” en los clips obtenidos al ver el video con la explicación matemática.

2. Preprocesamiento de imágenes

2.1 Extracción de frames de video

- Extrae 1–5 frames por segundo en la parte central de cada clip elegido.
- Tomar frames de alta calidad. (sin desenfoque).

- Tratar frames de un mismo clip como ejemplos correlacionados (importante para análisis por clip).

subject_id	clip_id	frame_index	label	ruta_imagen
S01	C01	0	confusion	S01_C01_f000_confusion.png
S01	C01	1	confusion	S01_C01_f001_confusion.png
S01	C02	0	Joy	S01_C02_f000_joy.png
S02	C03	0	fatigue	S02_C03_f000_fatigue.png
...

2.2 Detección y alineación de rostro

Usa un pipeline genérico (Haar, MediaPipe, etc.) para:

- Detectar el rostro.
- Alinear (ojos horizontales, recorte en torno al rostro).

2.3 Normalización y resizing

- Para Fisherfaces:
Convertir a escala de grises.
Redimensionar a 100×100.
Estandarizar (media 0, varianza 1).
- Para MobileNet:
Tamaño típico: 224×224 RGB.
Normalización [0,1] o según la implementación de Keras/PyTorch.

***Asegúrate de aplicar exactamente el mismo pipeline de preprocesamiento tanto a sintético como a real.**

3. Modelos a entrenar

Para obtener la calificación máxima es necesario generar los dos modelos.

3.1 Fisherfaces

Input: imágenes alineadas en gris.

Proceso:

PCA → reducción de dimensionalidad.

LDA (Fisher) → proyección que maximiza separación entre las 5 clases.

Clasificador: por ejemplo softmax sencillo en el espacio Fisher. Puede ser también KNN, SVM, entre otros.

Ventaja: sencillo, interpretable, buena “baseline tradicional”.

3.2 MobileNet (Efficient network)

Arquitectura: MobileNetV2 o V3.

Inicialización:

Desde pesos preentrenados en ImageNet.

Reemplazas la última capa por una densa de 5 neuronas con softmax.

Entrenamiento:

Congelar la mayor parte de las capas al inicio, luego fine-tuning ligero (comenzar a probar con 20 a 30 épocas)

Optimizer: Probar varios

Loss: categorical cross-entropy.

Early stopping con validación en el conjunto sintético (**val split**).

4. Protocolo de entrenamiento y prueba

4.1 Splits para el dataset sintético

Entrenamiento / Validación / Prueba sintético:

Split por sujeto.

70% entrenamiento (28 sujetos), 15% validación (6 sujetos), 15% prueba (6 sujetos).

4.2 Entrenamientos

- i. Entrenamiento1: entrenamiento (entrenamientop) + validación; test en el test sintético.
- ii. Entrenamiento2: entrenas igual (Ya lo tienes en E1); luego pruebas en con el conjunto de imágenes obtenidas a partir de videos reales.

4.3 Evalúa a dos niveles:

- a) Nivel frame:
Considerando cada frame como ejemplo independiente.
- b) Nivel clip:
Agrega las predicciones de todos los frames de un clip (p.ej. mayoría de votos o promedio de probabilidades).

4.4 Métricas principales

- Accuracy.
- macro-F1 (por si hay desbalance entre clases).
- Precision, recall, F1 por clase.
- Matriz de confusión

5 Dataset sintético de CAS

Enlace:

https://drive.google.com/drive/folders/14UDQ2gzo2d09Mhlh0JQVhR2NSAA79swM?usp=drive_link

Características de la DB Sintética de CAS

Número de sujetos en la base balanceada: 40

Tamaño de la base = $40 * 5 * 3 = 600$

(40 sujetos * 5 CAS * 3 ejemplos)

- **Distribución por CAS:**

confused 120

distracted 120

fatigued 120

joyful 120

neutral 120

- **Distribución por género:**

female 300

male 300

- **Distribución de sujetos en splits**

Sujetos train: 28

Sujetos val: 6

Sujetos test: 6

- **Distribución final por split (imágenes):**

train 420

val 90

test 90

6 Entrega del proyecto

Carpeta de Drive en la que se deben colocar los entregables:

<https://drive.google.com/drive/folders/1xh9UI7x2QTjnCRcS77HcOoHgtLcMskaA?usp=sharing>

Entregables

Coloca en la carpeta de drive correspondiente a tu equipo los siguientes entregables:

- i. Videos completos
- ii. Clips separados (Etiquetados con el CAS correspondiente)
- iii. Base de datos de expresiones faciales (CAS)
- iv. Código de entrenamiento, validación y prueba con Fisher Faces
- v. Código de entrenamiento, validación y prueba con MobileNet
- vi. Reporte completo. A parte de lo que normalmente incluye, debe reportar hallazgos, configuración de los modelos, métricas obtenidas y matriz de confusión
- vii. Presentación de Powerpoint.
- viii. Código