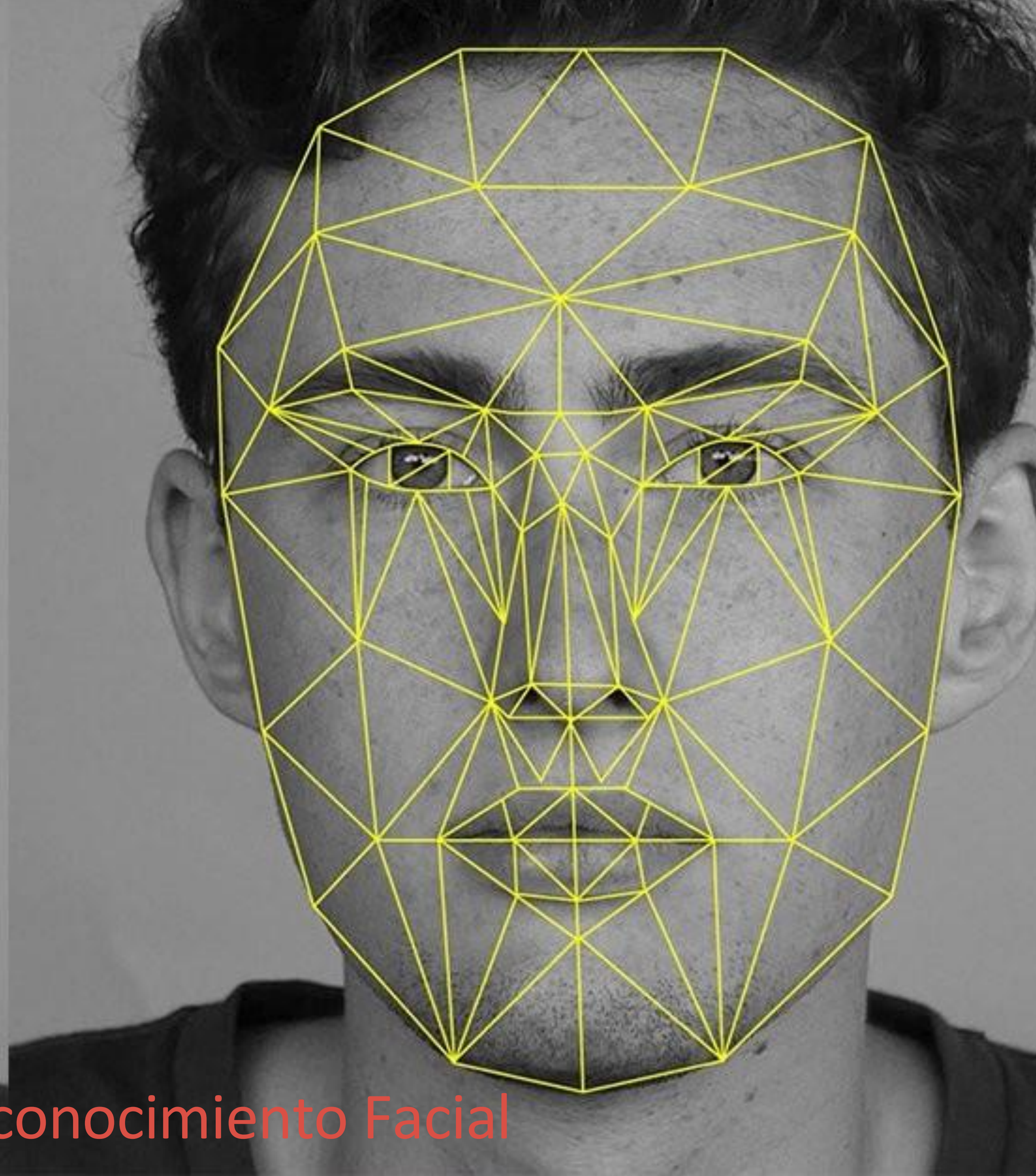
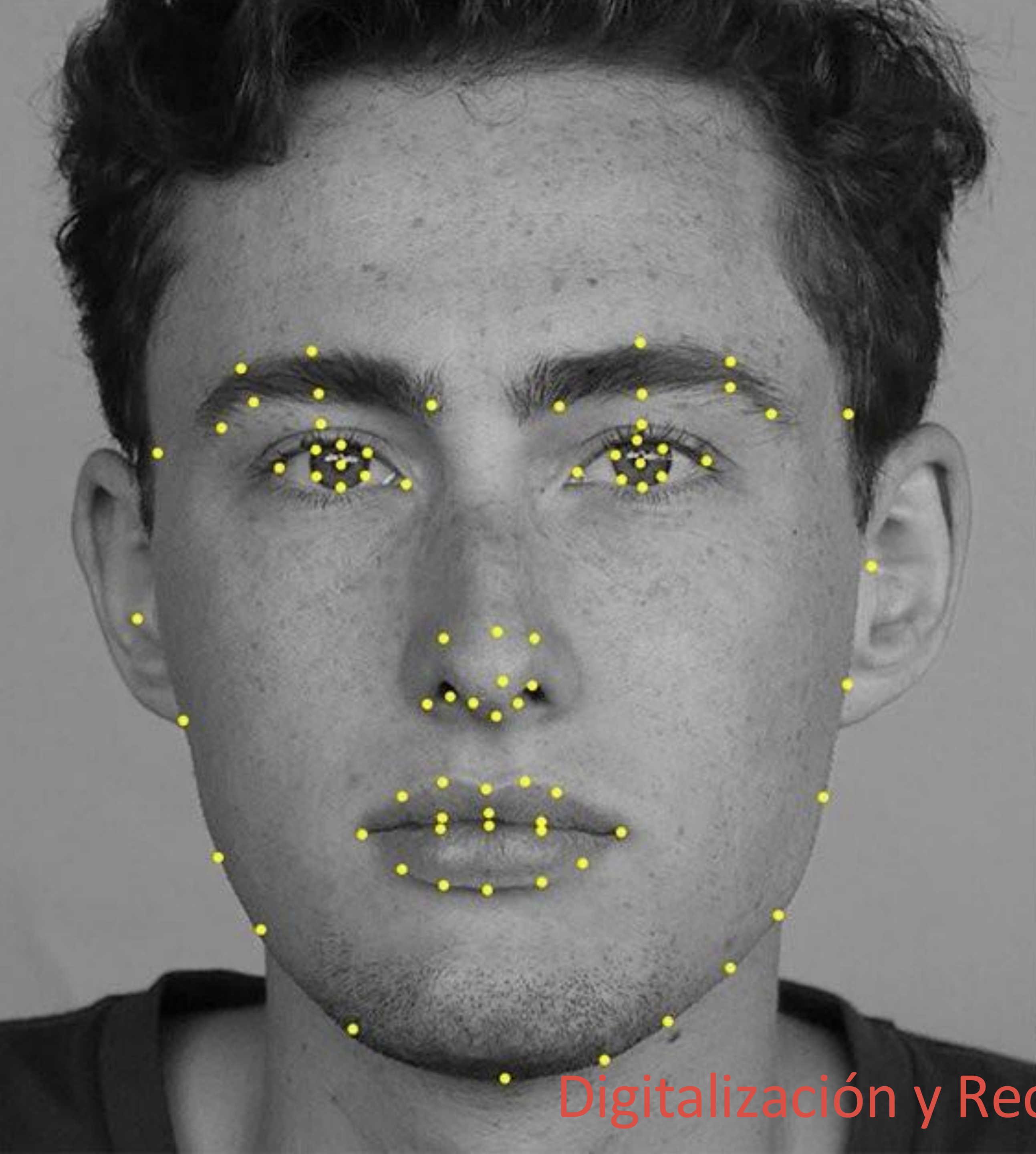


Introducción al reconocimiento facial y sus aplicaciones

Reconocimiento Facial para Archivo Fotográfico

GUILLEM CADEVALL, FRANCESC TARRÉS (UGIAT TECHNOLOGIES, UPC)



Digitalización y Reconocimiento Facial

Tecnologías de Reconocimiento Facial

Tipologías RF: 1. Problema biométrico

Verificación

Identificación

Clustering

Vigilancia

Deduplicación

Indexación Archivo Audiovisual

Tecnologías de Reconocimiento Facial

Tipologías RF: 2. Pipeline

Detección

Alineamiento

Embedding

Búsqueda

Decisión

Tecnologías de Reconocimiento Facial

Tipologías RF: 3. Familias tecnológicas

Modelos especializados

Modelos visuales genéricos

Modelos multimodales

Servicios Comerciales

Tecnologías de Reconocimiento Facial

Tipologías RF: 4. Criterios de selección

Precisión

Tamaño

Latencia

Coste

Privacidad

Explicabilidad

sesgo

Facilidad de despliegue

disponibilidad de pesos/licencias

Tecnologías de Reconocimiento Facial

Tipologías RF: 5. Captura de la imagen

Métodos 2D: Reconocimiento de imágenes planas. TV, foto, video

Métodos 3D: Adquisición de datos en 3D. Sistemas de seguridad acceso

Métodos Mixtos: Inferencia de la información 3D a partir de proyecciones

Tecnologías de Reconocimiento Facial

Tipologías RF: 6. Tecnologías de procesamiento

Detección de Landmarks biométricos

Métodos Holísticos (PCA, LDA, Adaboost, etc.)

Métodos basados en Deep Learning

Métodos basados en proyección de puntos

Métodos landmarks avanzados (3D)

Tecnologías de Reconocimiento Facial

Problemática a resolver: Verificación, reconocimiento, etc.

Arquitectura funcional del modelo: Detección -> Alineamiento -> embedding -> comparación/identificación

Restricciones aplicación: Precisión, coste computacional, Cloud/On-Premise, etc.

Grado de especialización del modelo: Modelos faciales específicos, modelos visuales genéricos y modelos multimodales.

Tipo de imágenes de entrada: 2D, 3D, 2.5 D

Tecnologías de Reconocimiento Facial

Métodos 2D

Son los más habituales en reconocimiento facial.

La entrada es una imagen plana en la que la cara aparece proyectada sobre el plano de la cámara.

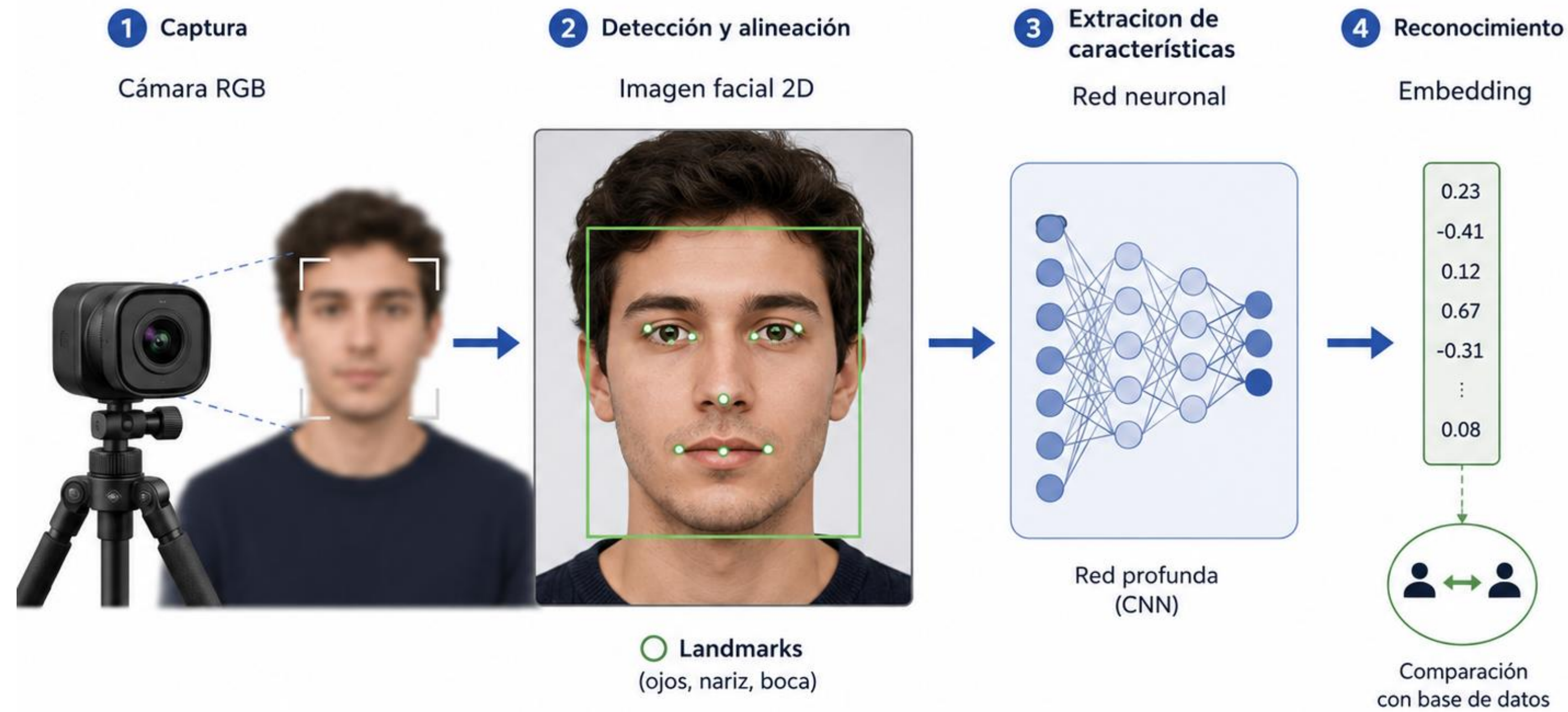
Son simples y no requieren sensores especiales.

Son sensibles a cambios de pose, iluminación, expresión facial, oclusiones, resolución, edad, maquillaje, gafas, calidad de captura.

Existen grandes bases de datos de imágenes para hacer los entrenamientos de los modelos.

Tecnologías de Reconocimiento Facial

Método 2D



- 1 La cámara RGB captura el rostro → 2 Se detecta y alinea el rostro en 2D y se extraen landmarks → 3 Una red neuronal extrae un vector de características (embedding) → 4 El embedding se compara con la base de datos para reconocer la identidad

Tecnologías de Reconocimiento Facial

Métodos 3D

Usan información tridimensional de la cara, capturando la geometría real del rostro: profundidad, forma de la nariz, órbitas de los ojos, pómulos, curvatura de la superficie facial.

Se puede obtener mediante escáneres 3D, cámaras de profundidad, sensores de luz estructurada, cámaras ToF, luz infraroja, reconstrucción, etc. (Android, iPhone).

Es más estable que 2D ya que la forma del rostro cambia menos. Son menos sensibles a ataques de suplantación mediante imágenes. Tienen mayor independencia ante la pose

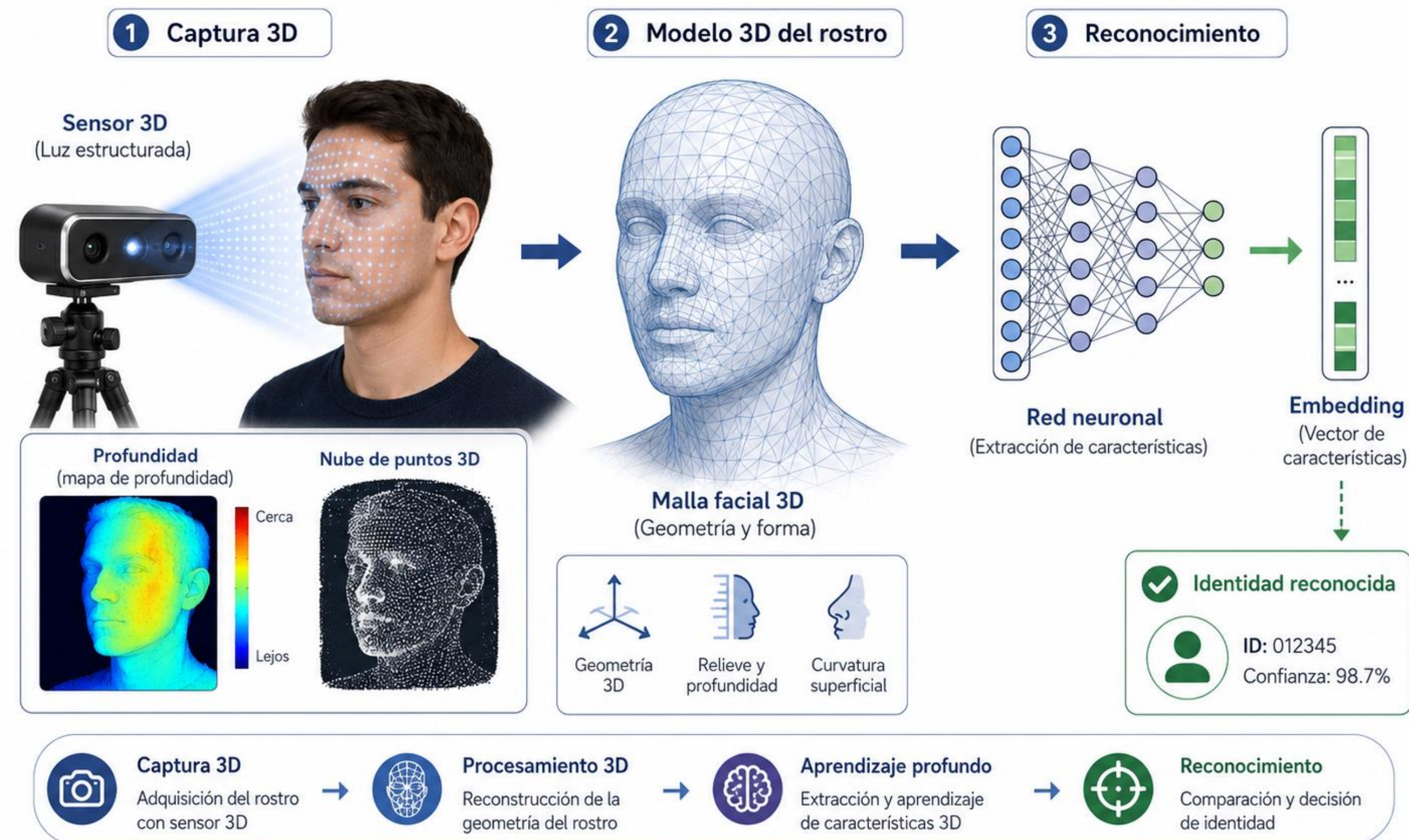
Se requieren sensores específicos por lo que son considerablemente más caros.

También puede verse afectada por oclusiones.

Tecnologías de Reconocimiento Facial

Método 3D

Reconocimiento facial basado en la geometría 3D del rostro



Tecnologías de Reconocimiento Facial

Métodos 2.5 D

Utilizar imágenes 2D para inferir la orientación de la cabeza y reconstruir una versión aproximada de la cara en 3D.

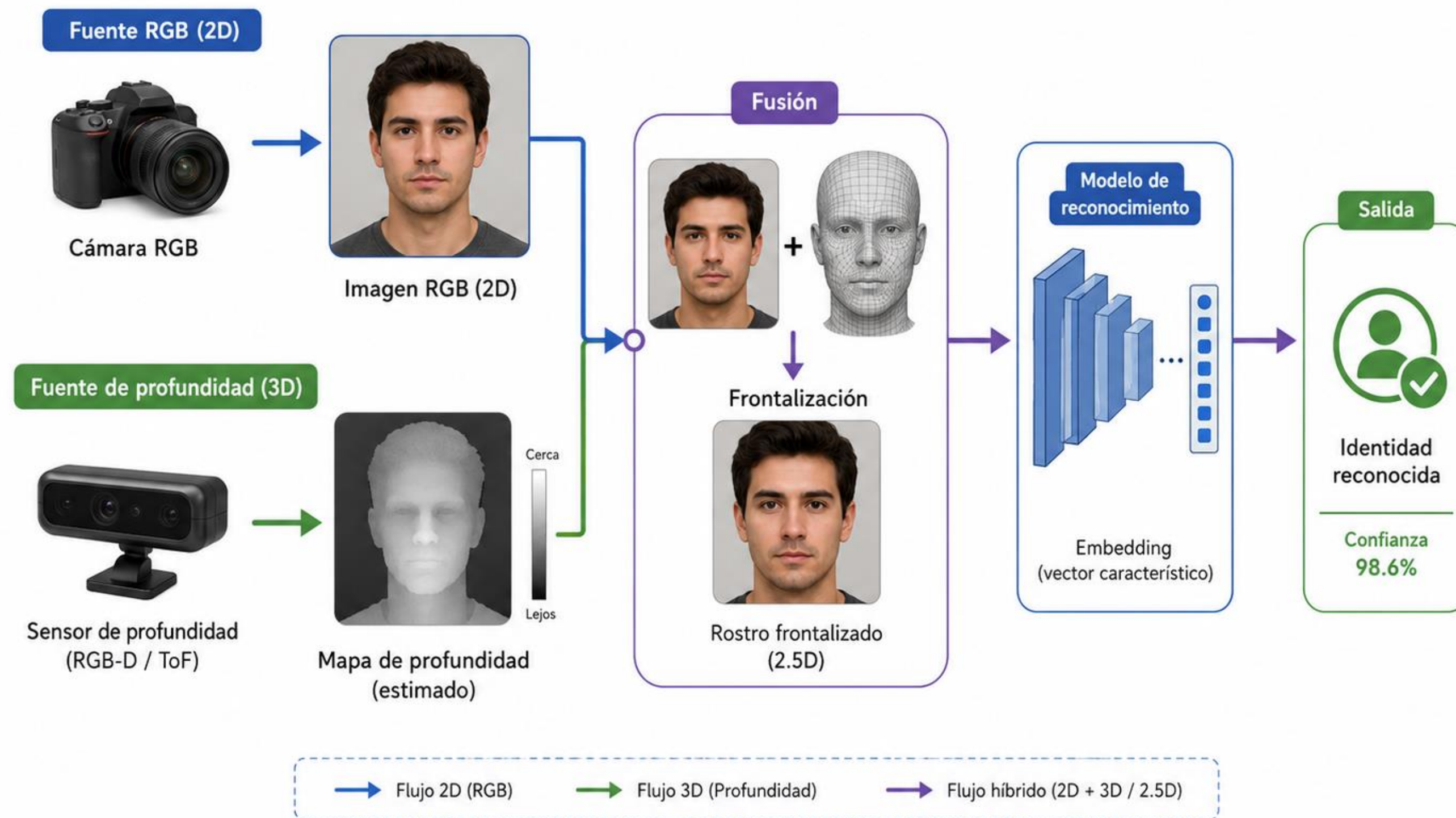
Uso de sensores RGB-Depth. Cada pixel tiene asociadas 4 componentes: R, G, B i D.

Pueden obtener algunas mejoras sobre los métodos 2D clásicos sin necesidad de equipos de adquisición complejos que pueden usarse en instalaciones de bajo coste.

Los métodos mixtos intentan obtener lo mejor de ambos mundos: la facilidad de captura de los métodos 2D y parte de la robustez geométrica de los métodos 3D. Para ello estiman profundidad, pose o modelos faciales 3D a partir de imágenes 2D, o combinan imágenes RGB con sensores de profundidad.

Tecnologías de Reconocimiento Facial

Método mixto (2D + 3D / 2.5D)



Idea clave:

El método mixto combina la información de apariencia (2D) con la geometría facial (3D) para obtener representaciones más robustas frente a pose, iluminación y oclusiones.

Tecnologías de Reconocimiento Facial

2.5 D for Pose Invariance

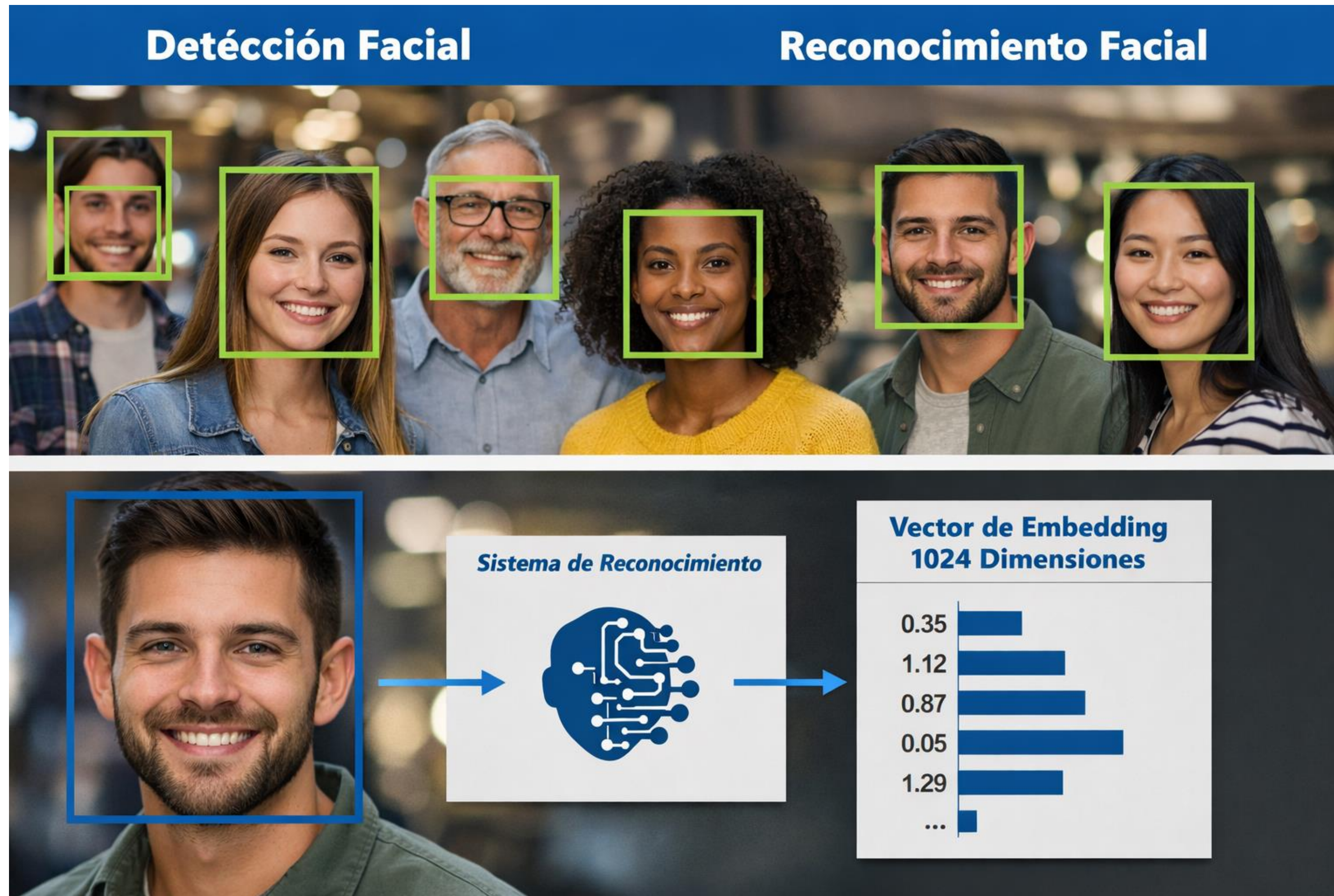


High Correlation in left Area



High Correlation in central Area

Fases del Reconocimiento Facial



Fases del Reconocimiento Facial

Pipeline Básico

DETECCIÓN FACIAL

RECONOCIMIENTO
FACIAL

Detección Facial

Reconocimiento facial

Fases del Reconocimiento Facial

Pipeline Básico



Detección Facial

Acondicionamiento

Reconocimiento facial

Detección Fiducials

Corrección frontal

Normalización

Fases del Reconocimiento Facial

Pipeline Básico (Face Recognition)

Detección Facial: Localizar las caras en el vídeo/imagen. Los detectores más usados son Adaboost (OpenCV), MTCNN, RetinaFace, SCRFD, Yolo-Face o variantes basadas en CNN/Transformers.

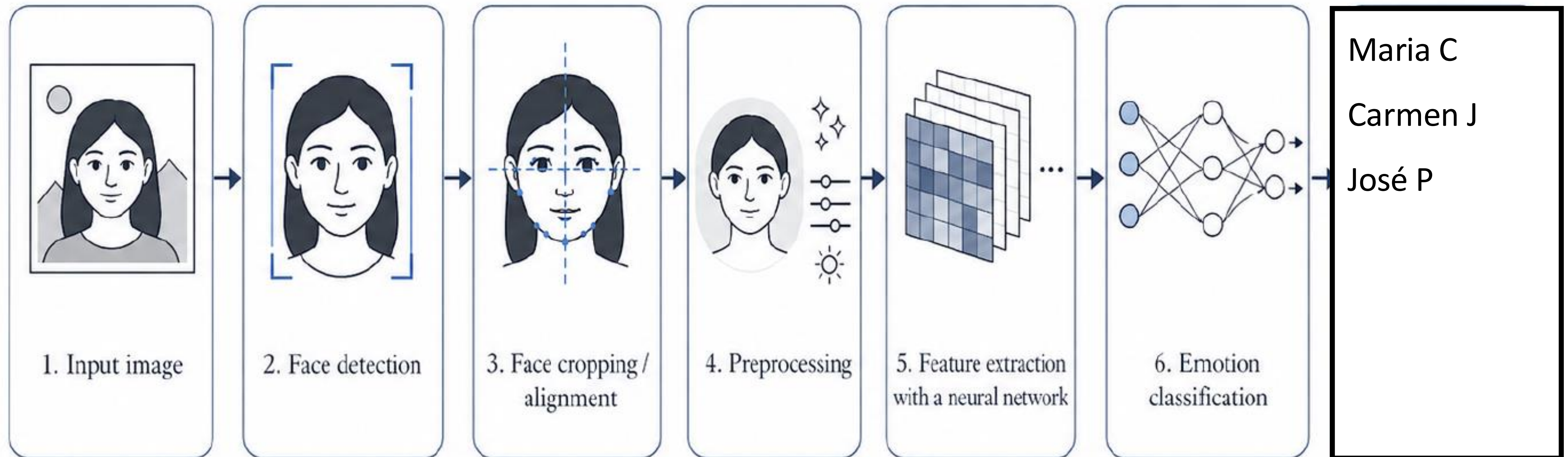
Alineamiento y normalización. A partir de los fiducials se corrige la alineación y la escala de las caras. Pequeños errores pueden afectar mucho al embedding.

Extracción de embedding. Se convierte la cara en un vector, que intenta conservar la identidad y descartar variaciones de pose, iluminación, expresión, edad, cámara, etc. FaceNet fue uno de los algoritmos fundacionales

Comparación, verificación o identificación: Con el embedding se puede hacer verificación 1:1, verificación 1:N o clustering - agrupar caras de una misma persona sin saber su nombre.

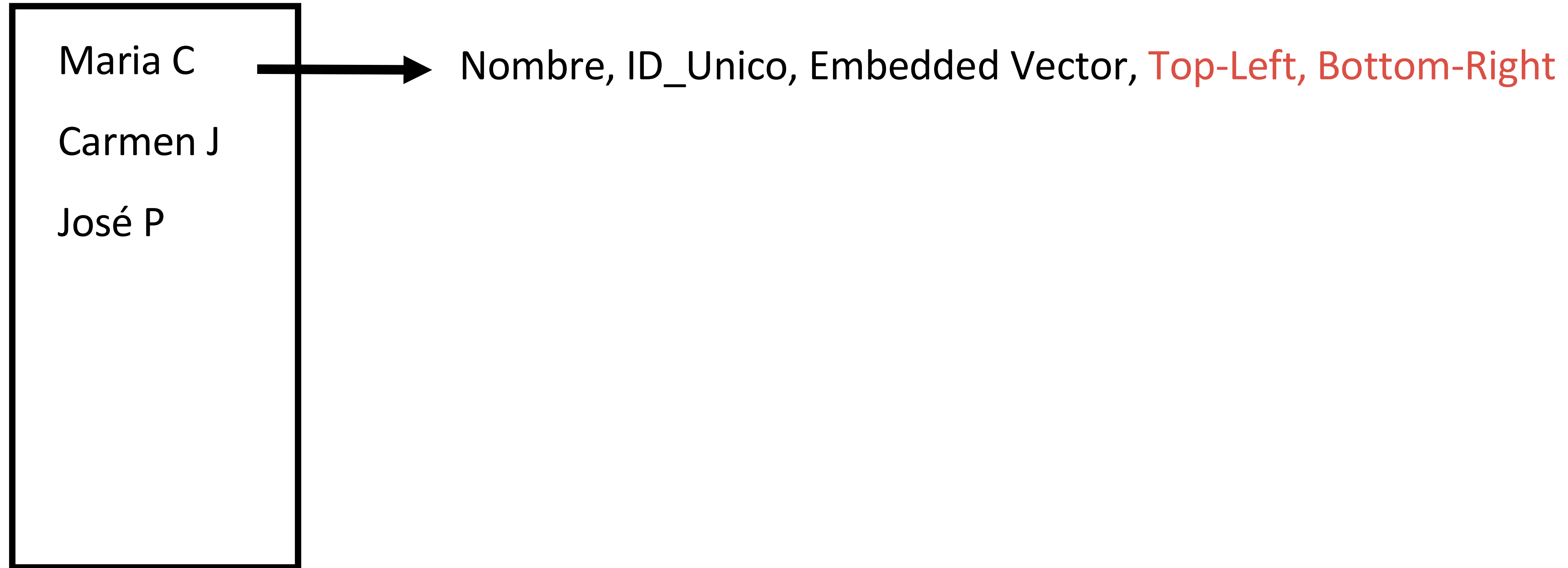
Fases del Reconocimiento Facial

Pipeline Básico (Face Recognition)



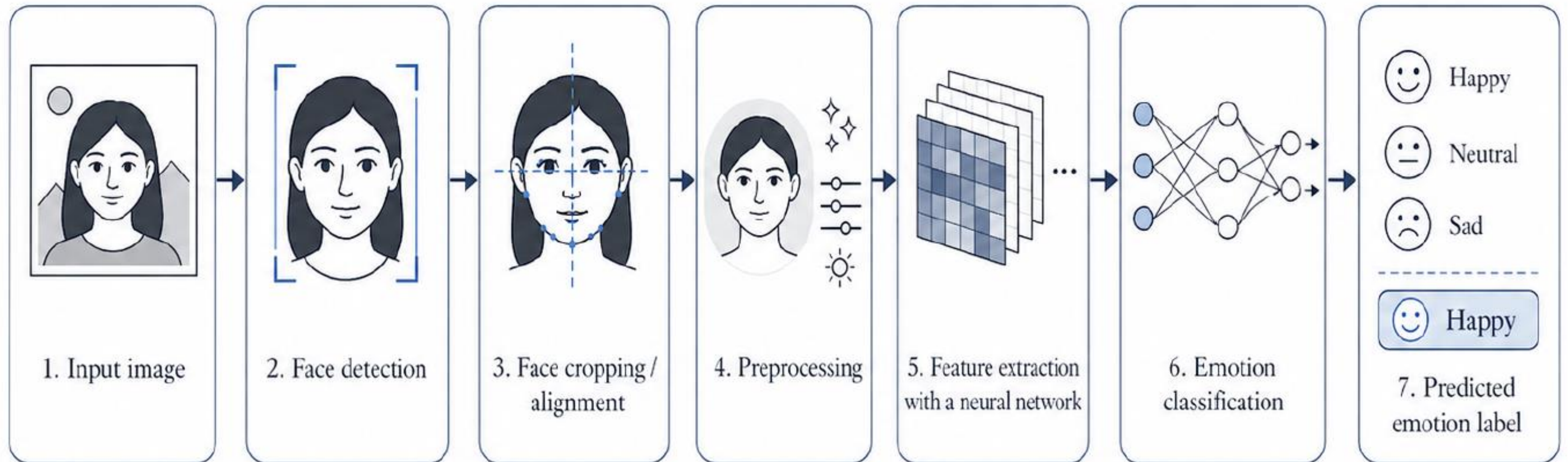
Fases del Reconocimiento Facial

Pipeline Básico



Fases del Reconocimiento Facial

Pipeline Básico (Emotion Recognition)



Face Detection

- Adaboost
- MTCNN
- Retina Face (BB + puntos faciales)
- SCRFD
- YOLO-Face



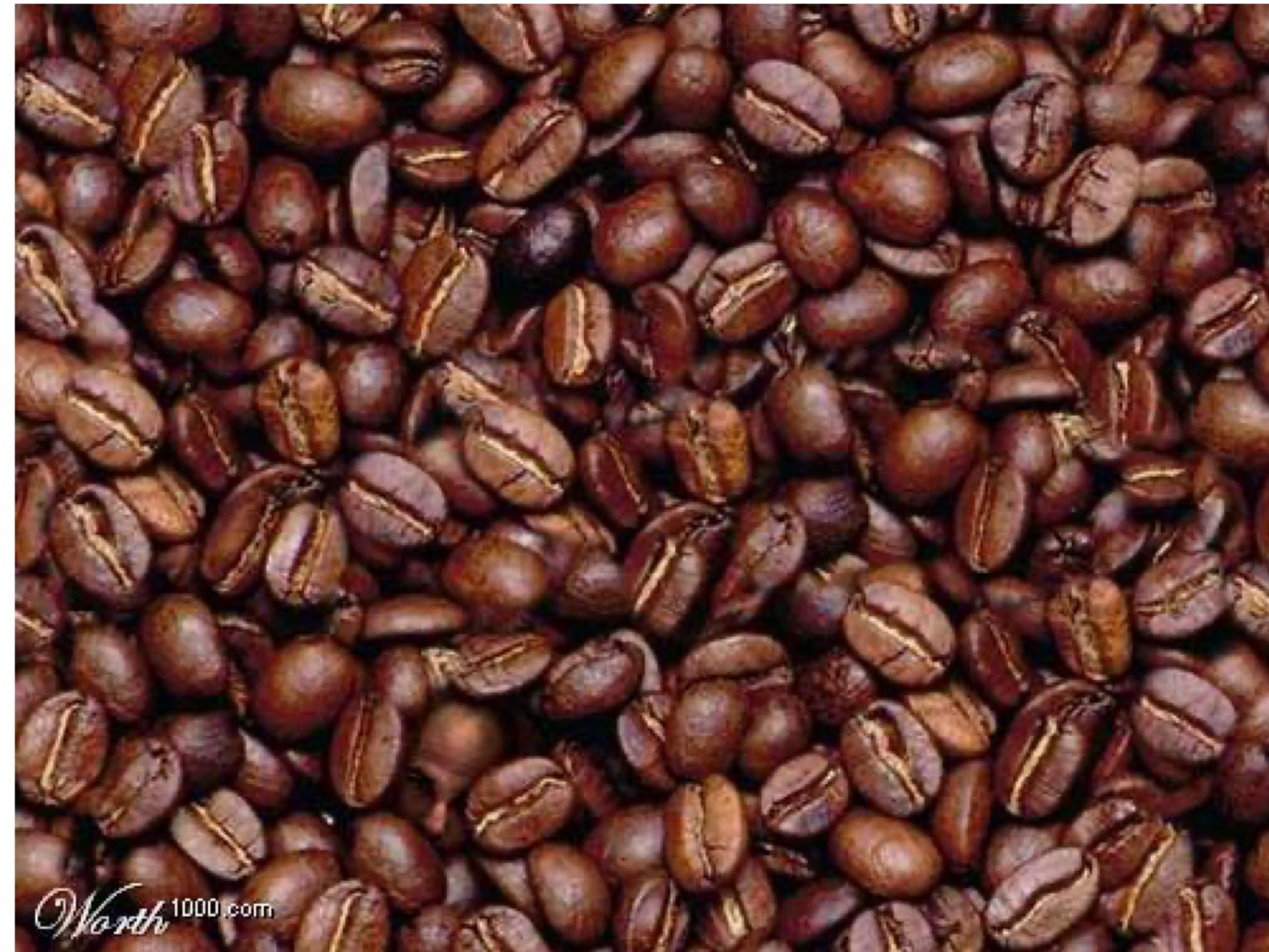
The Concept of Face Detection

Adaboost Face Detection, Viola & Jones, 2001



The Concept of Face Detection

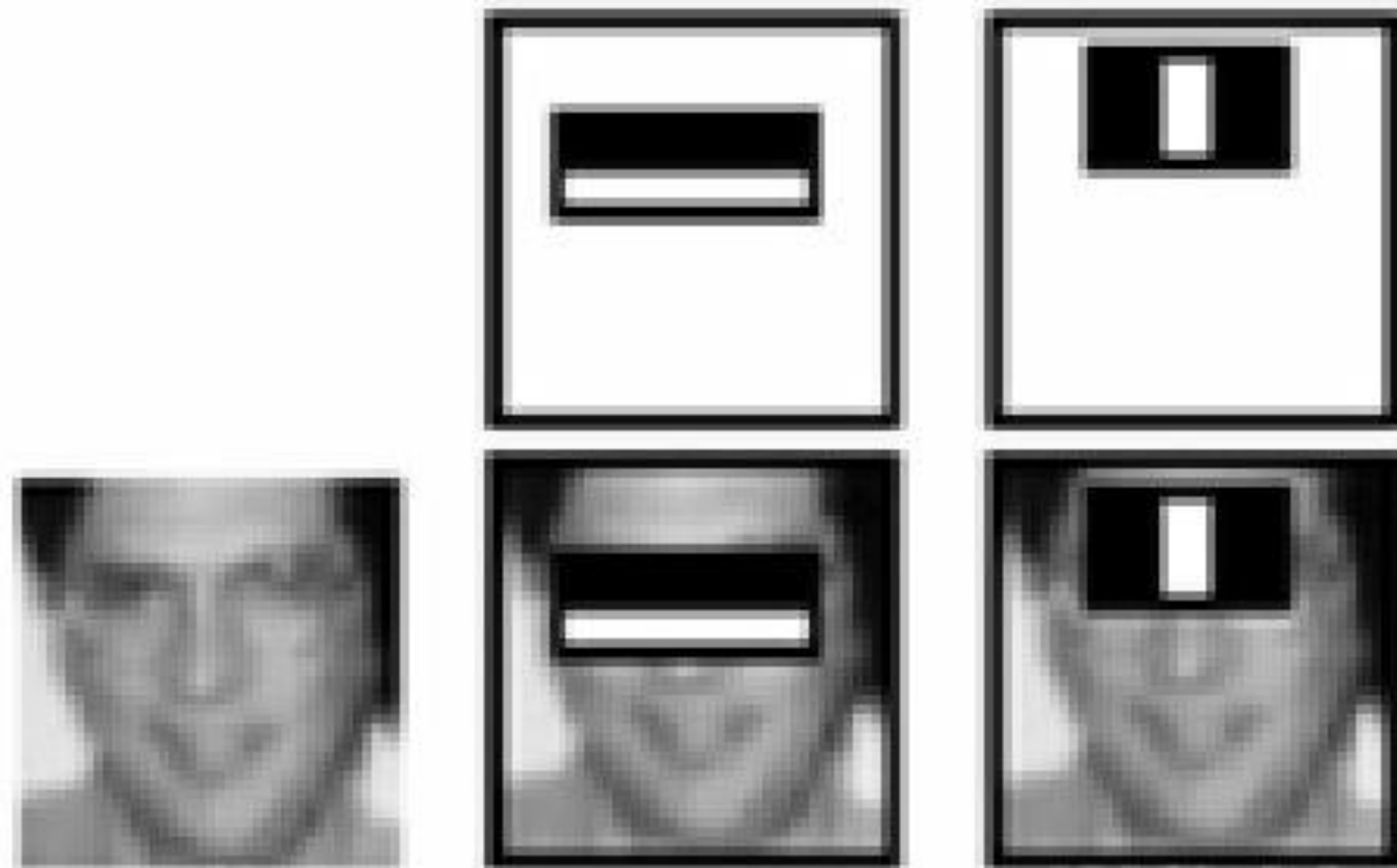
Adaboost Face Detection, Viola & Jones, 2001



The original image was created 2002 by user 'Furitsu' (a real Photoshop wizard) at Worth1000.com

Deteción Facial

Adaboost (OpenCV)



A very short history of Computer Vision

Adaboost Face Detection, Viola & Jones, 2001

Non-natural faces looking as faces

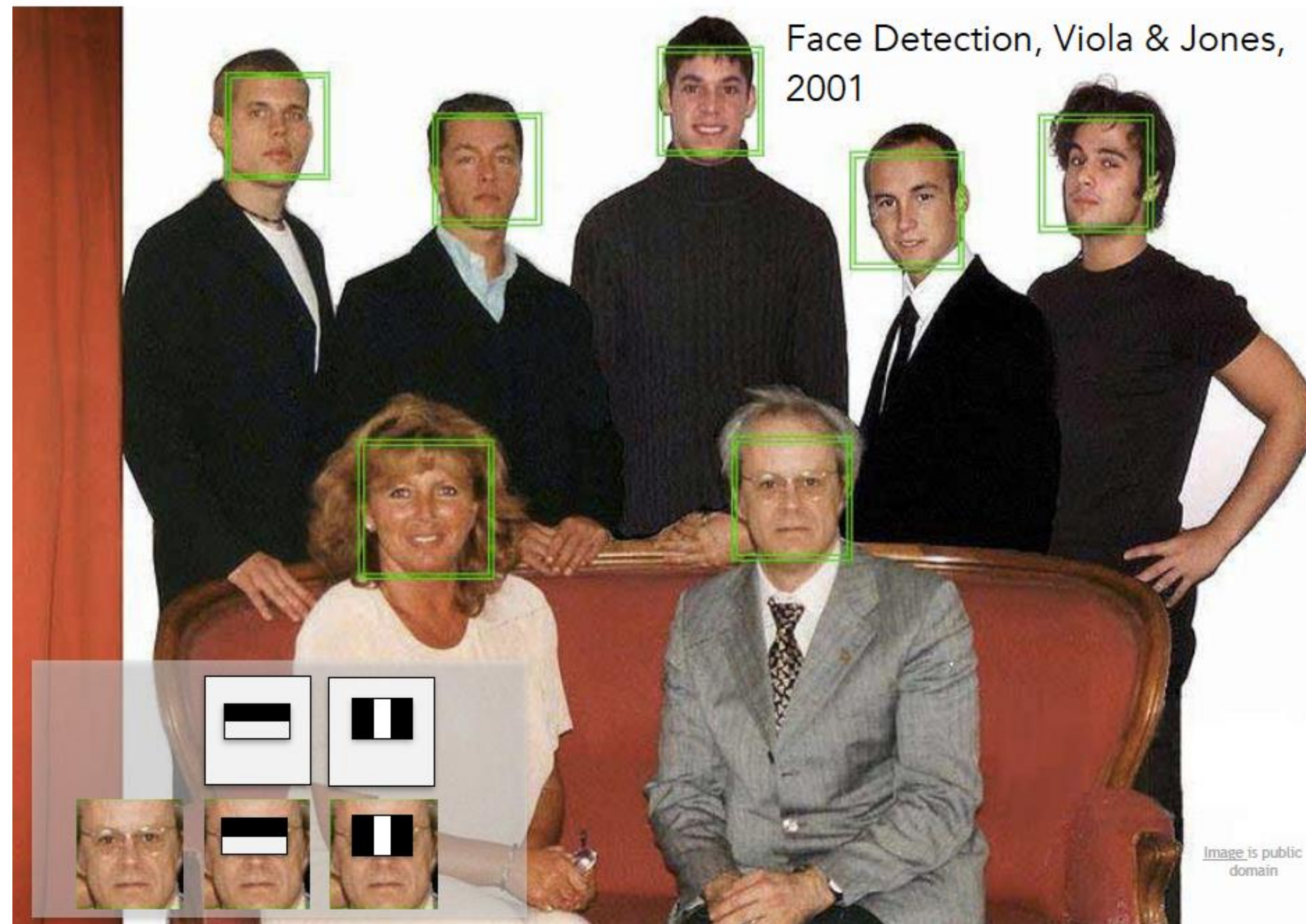


A very short history of Computer Vision

Adaboost Face Detection, Viola & Jones, 2001



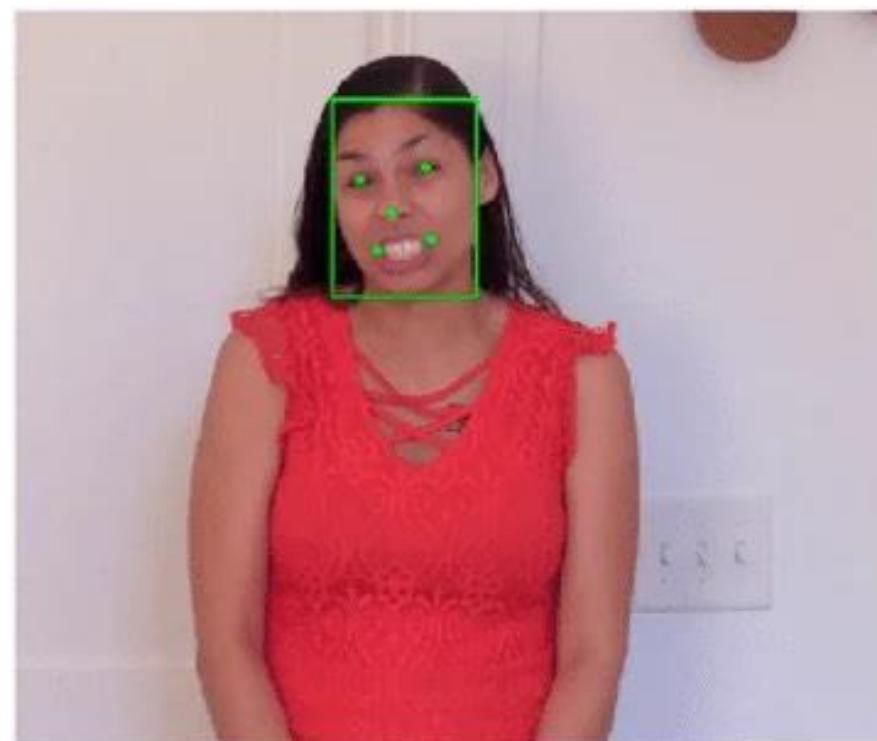
A very short history of Computer Vision



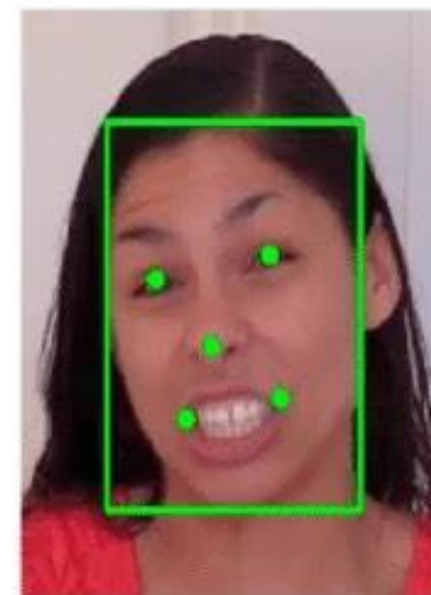
Face Detection

RetinaFACE

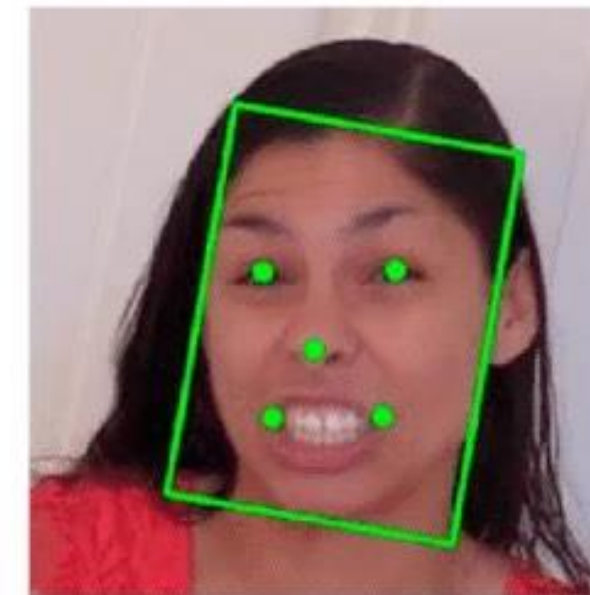
- ★ Inspired in RetinaNet an object detection architecture
- ★ CNN based architecture.
- ★ Trained to detect face, eyes, mouth and nose



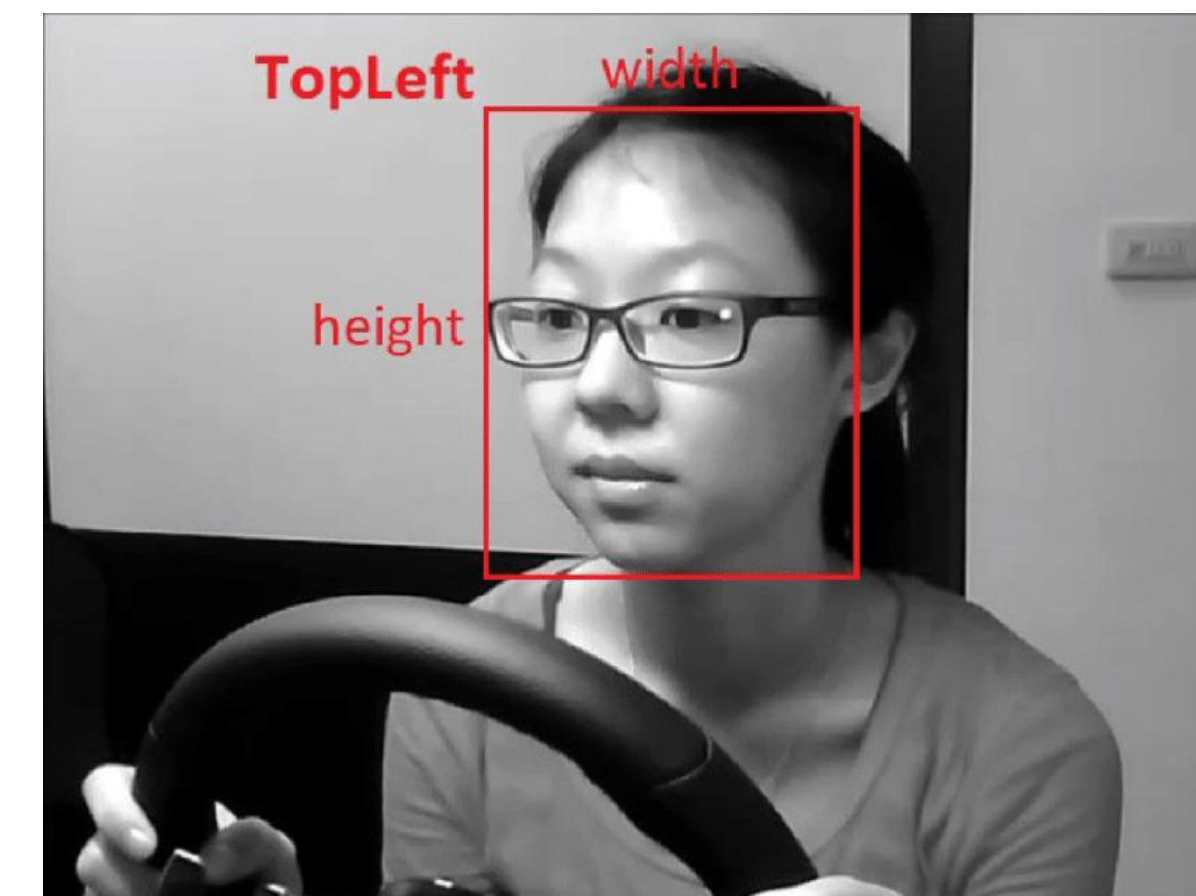
(a) - Face Landmark Detection



(b) - Face Crop



(c) - Face Alignment



RetinaFace

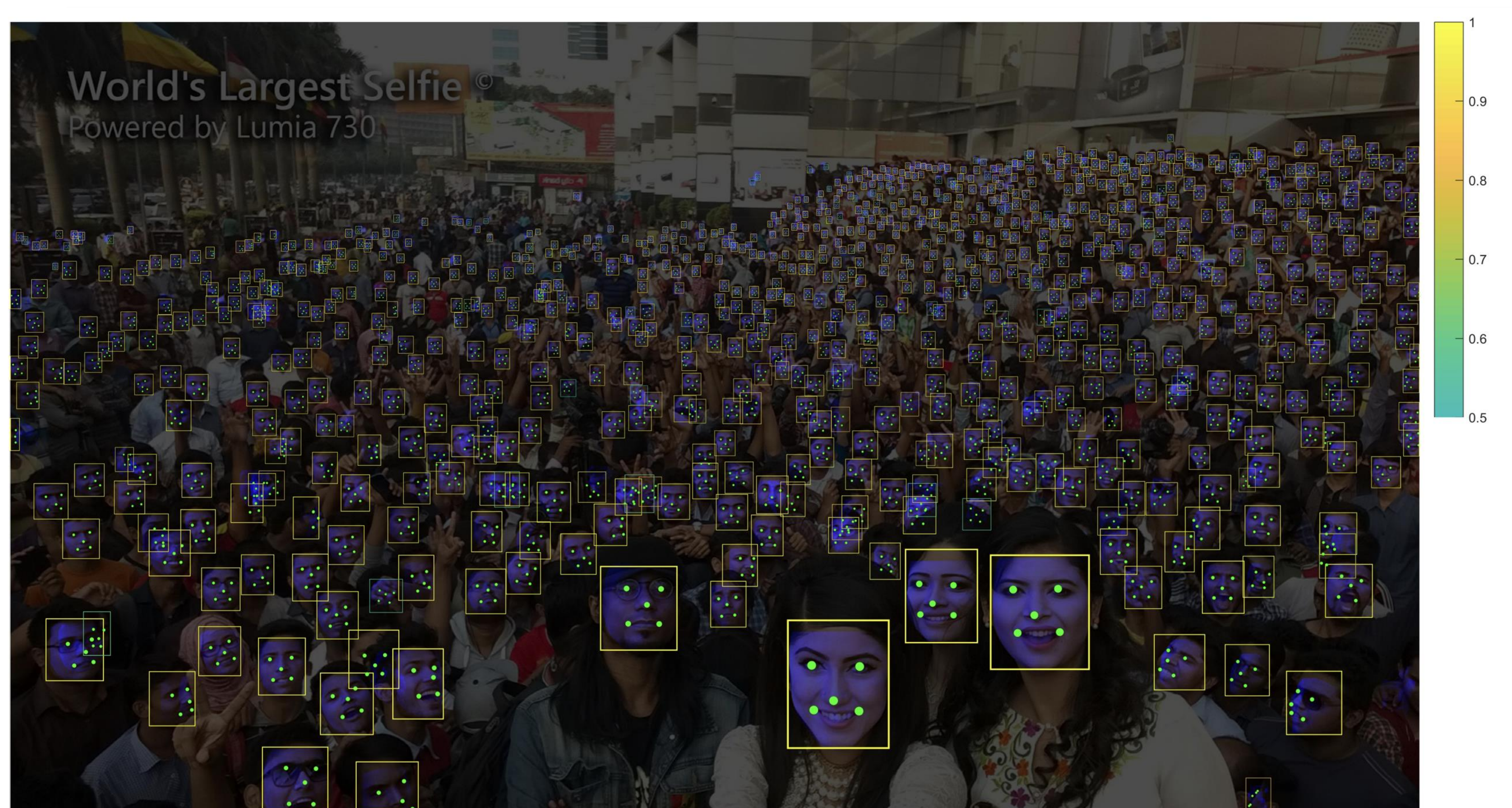
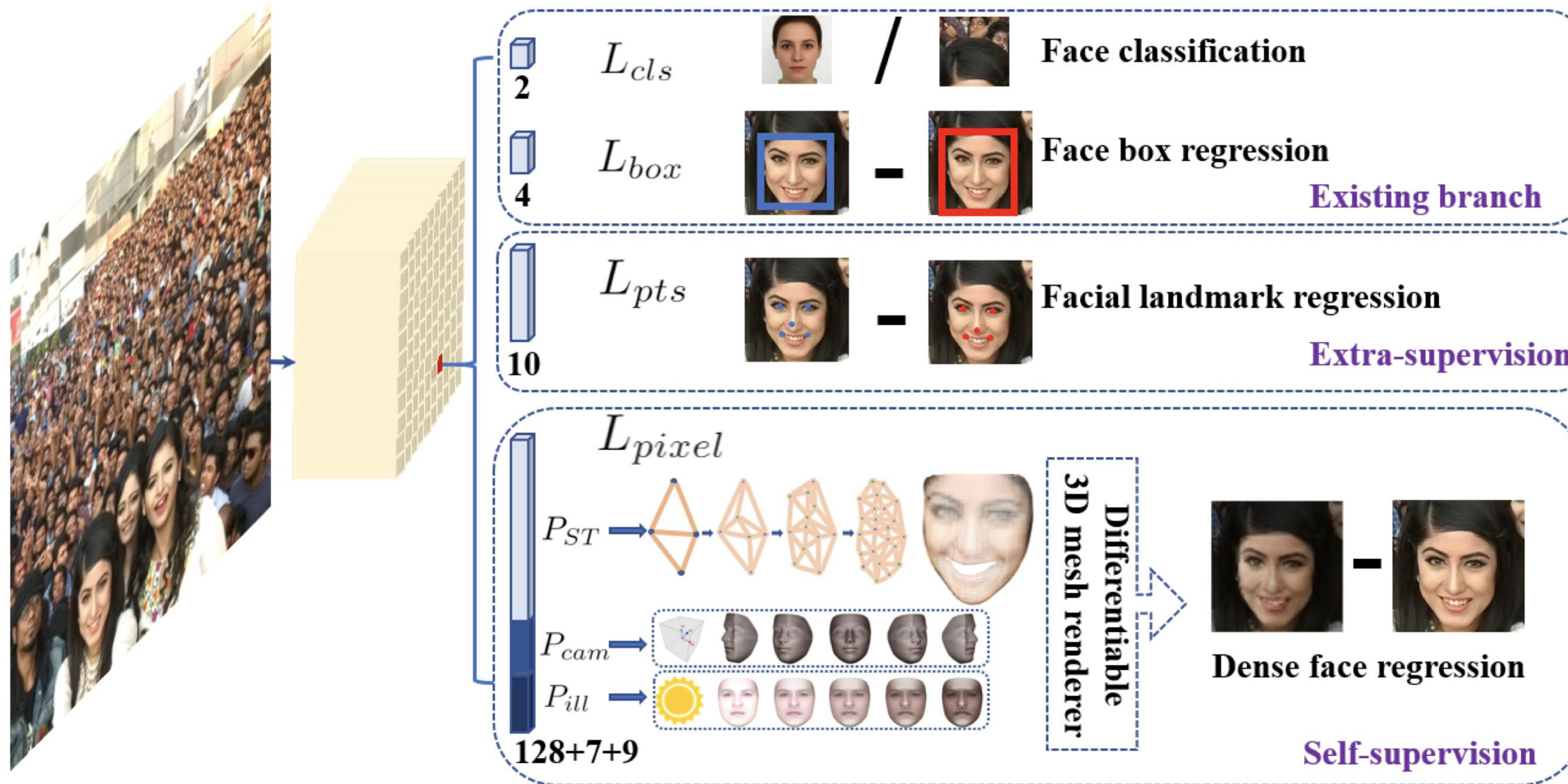


Figure 6. RetinaFace can find around 900 faces (threshold at 0.5) out of the reported 1151 people, by taking advantages of the proposed joint extra-supervised and self-supervised multi-task learning. Detector confidence is given by the colour bar on the right. Dense localisation masks are drawn in blue. Please zoom in to check the detailed detection, alignment and dense regression results on tiny faces.

RetinaFace



RetinaFace

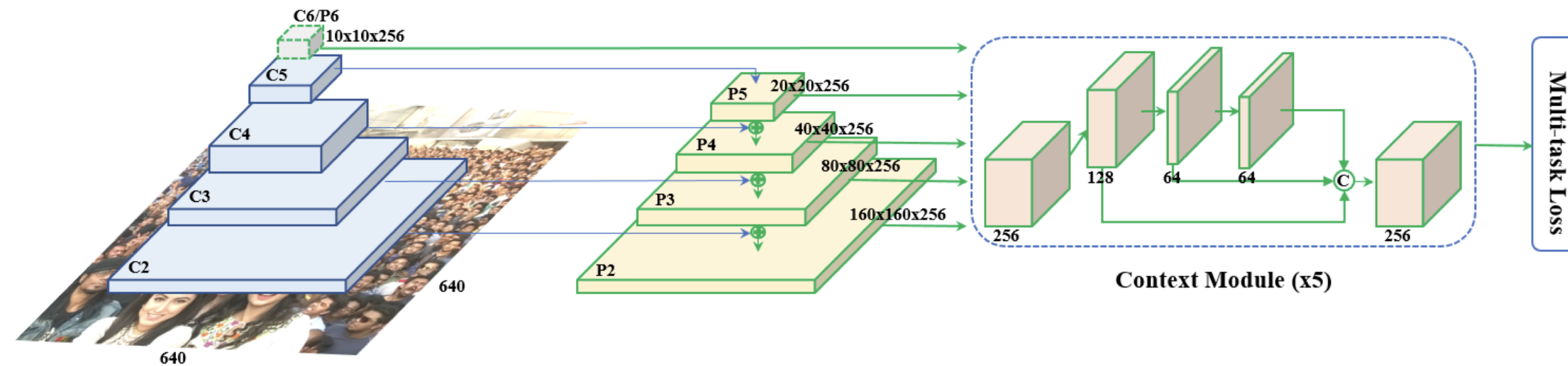
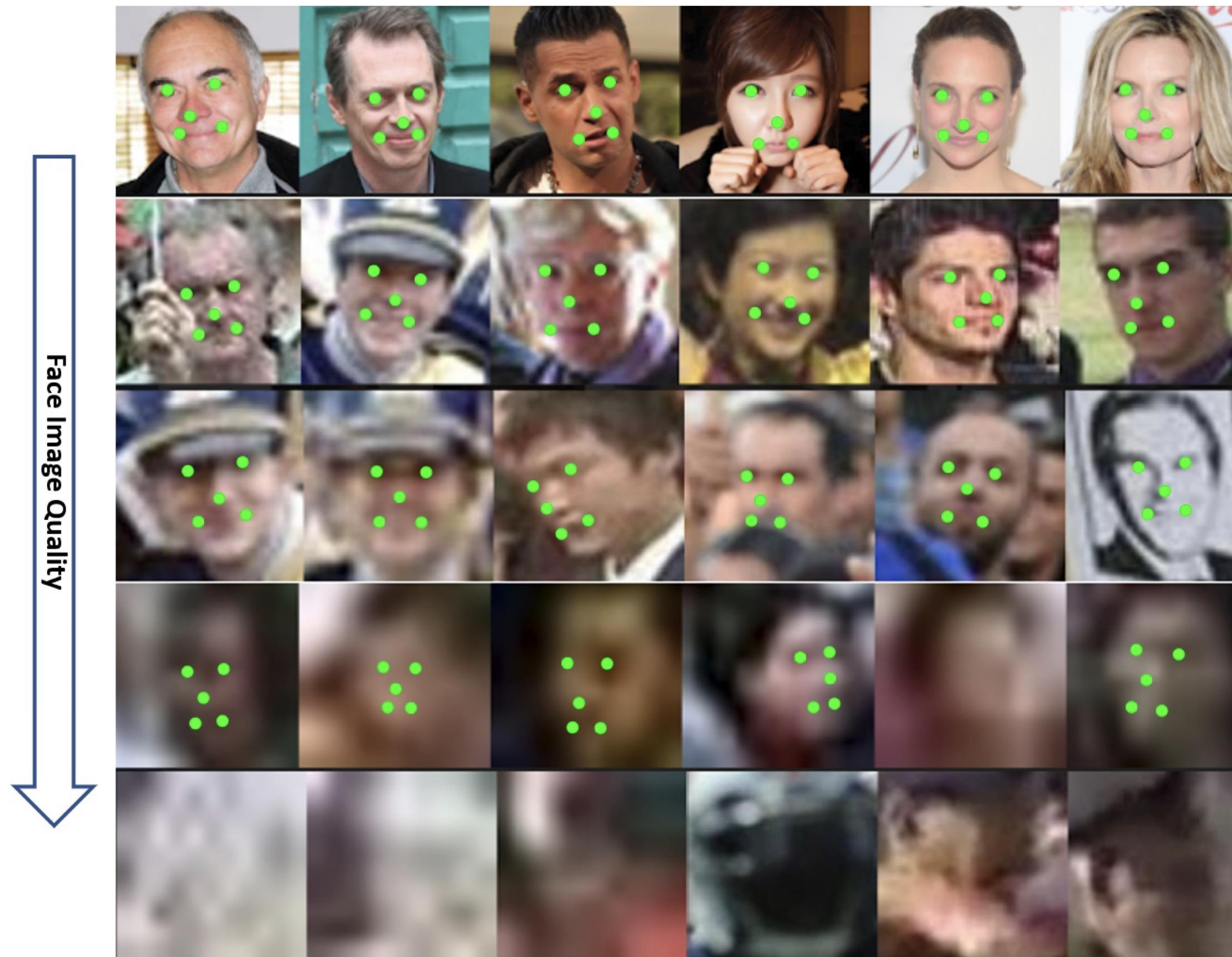


Figure 2. An overview of the proposed single-stage dense face localisation approach. RetinaFace is designed based on the feature pyramids with independent context modules. Following the context modules, we calculate a multi-task loss for each anchor.

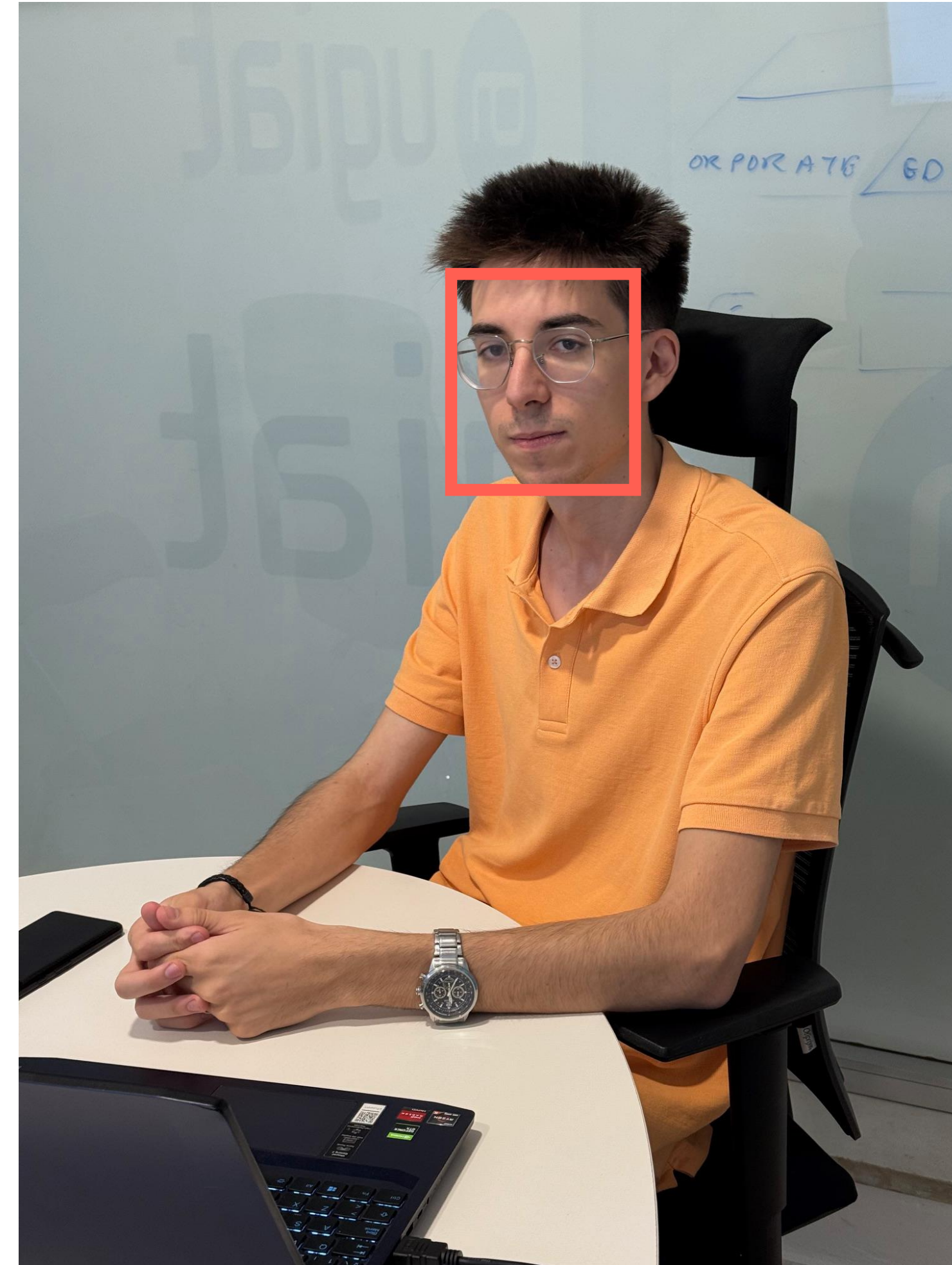
RetinaFace



Detección Facial



Face
Detector



Detecção Facial: YuNet Milisecond Face Detector



Reconocimiento Facial: Entrenamiento



Red Neuronal
Embedding Vectors

QUEREMOS UN VECTOR
que represente a Guillem

Face Detection & Recognition

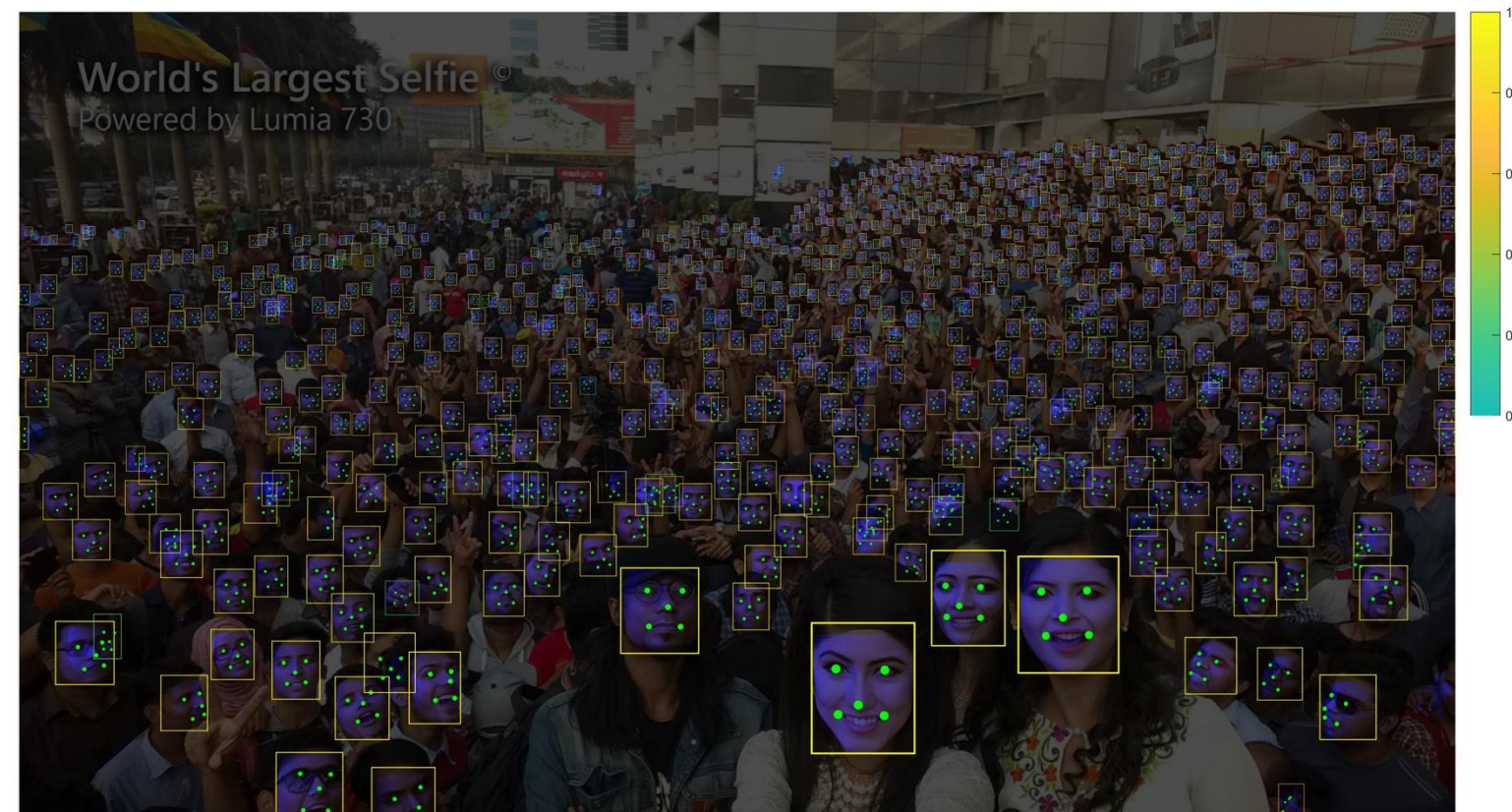
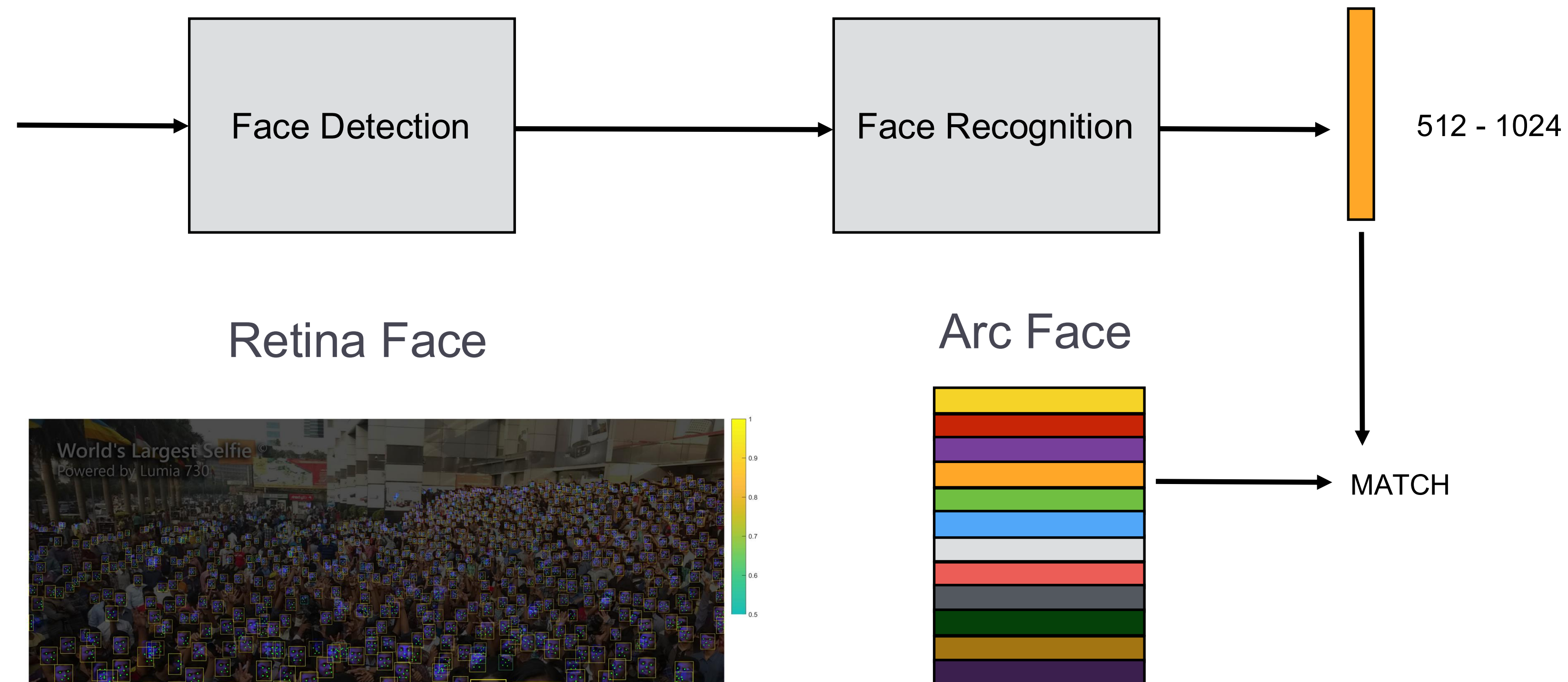
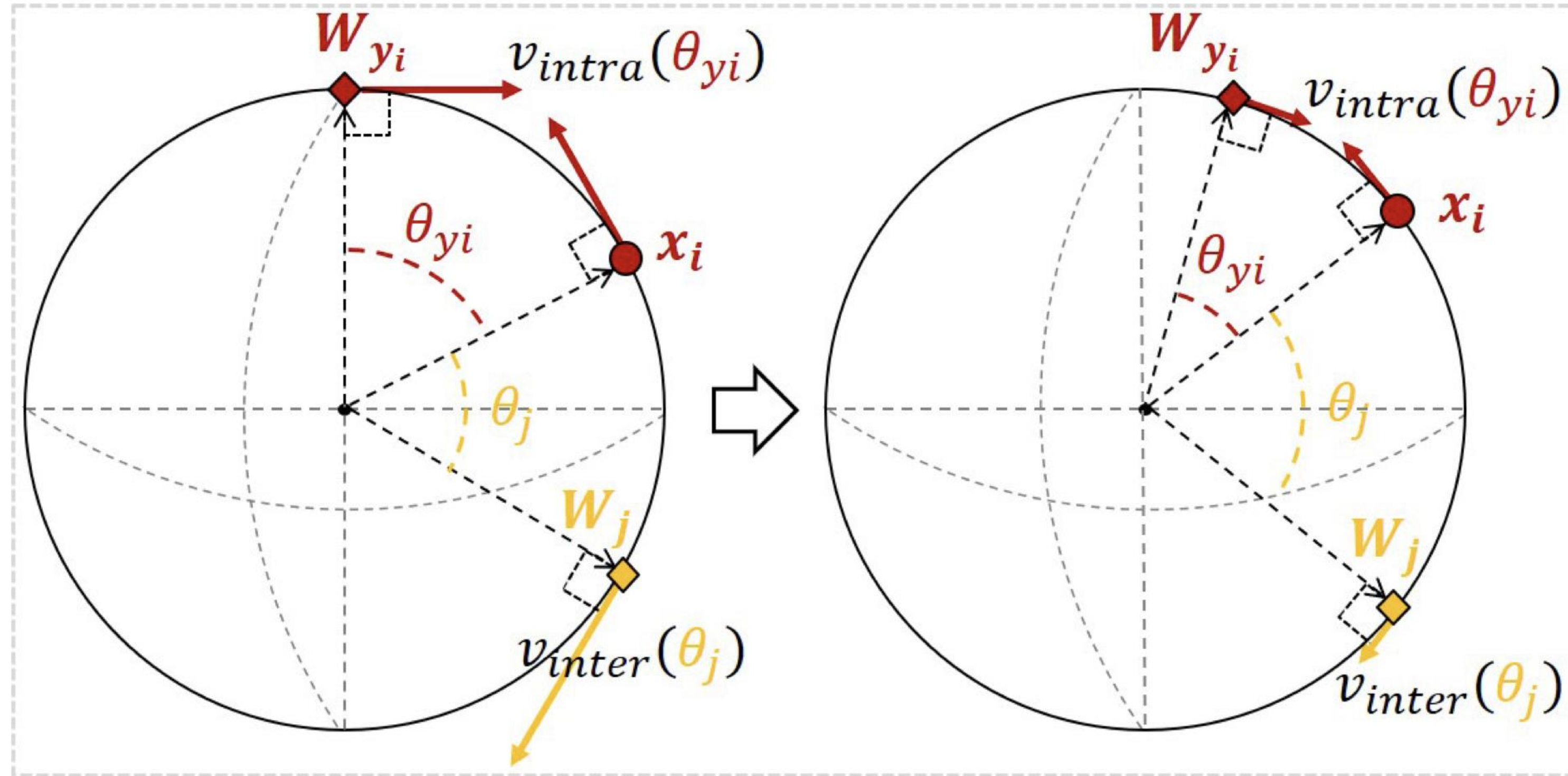
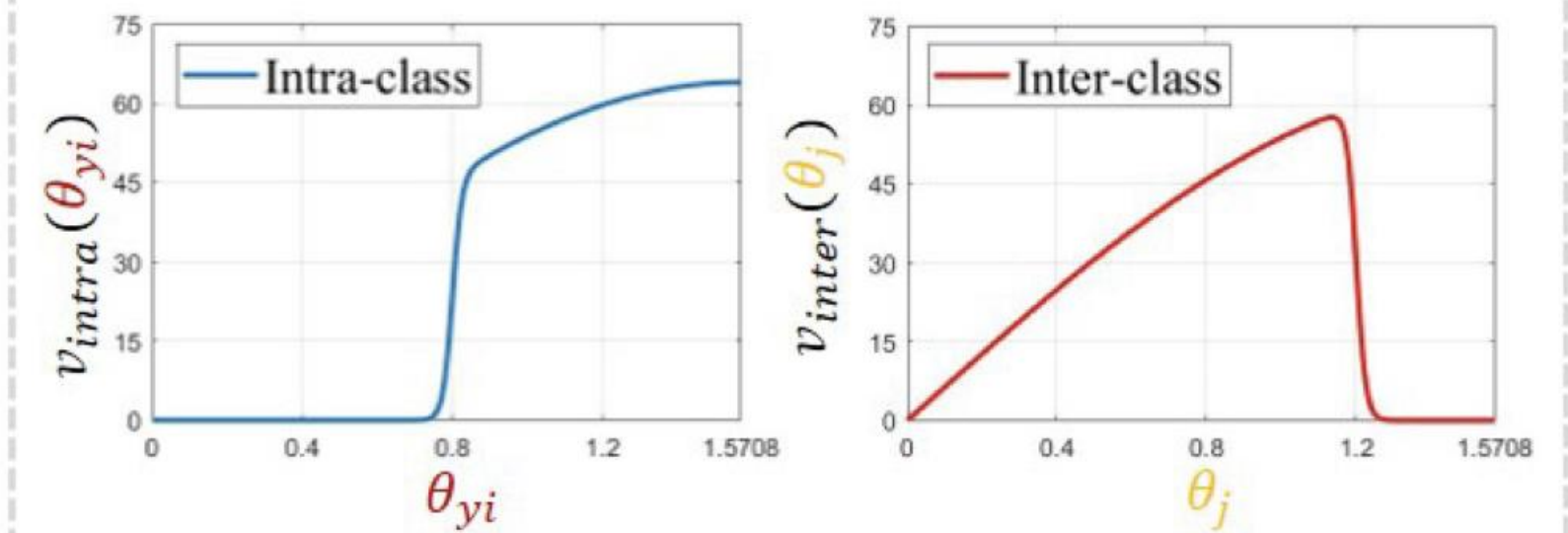


Figure 6. RetinaFace can find around 900 faces (threshold at 0.5) out of the reported 1151 people, by taking advantages of the proposed joint extra-supervised and self-supervised multi-task learning. Detector confidence is given by the colour bar on the right. Dense localisation masks are drawn in blue. Please zoom in to check the detailed detection, alignment and dense regression results on tiny faces.

Reconocimiento Facial



- x_i : The i -th training sample with label y_i .
- ◆ W_{y_i} : The target embedding of identity y_i .
- - θ_{yi} : The angle between x_i and target W_{y_i} .
- $v_{intra}(\theta_{yi})$: The scale of intra-class gradient.
- ◆ W_j : Targets of other identities $j \neq y_i$.
- - θ_j : Angles between x_i and other targets W_j .
- $v_{inter}(\theta_j)$: The scale of inter-class gradient.



Face Recognition Basic Idea

Ejemplos en reconocimiento facial (entrenamiento por duplets)



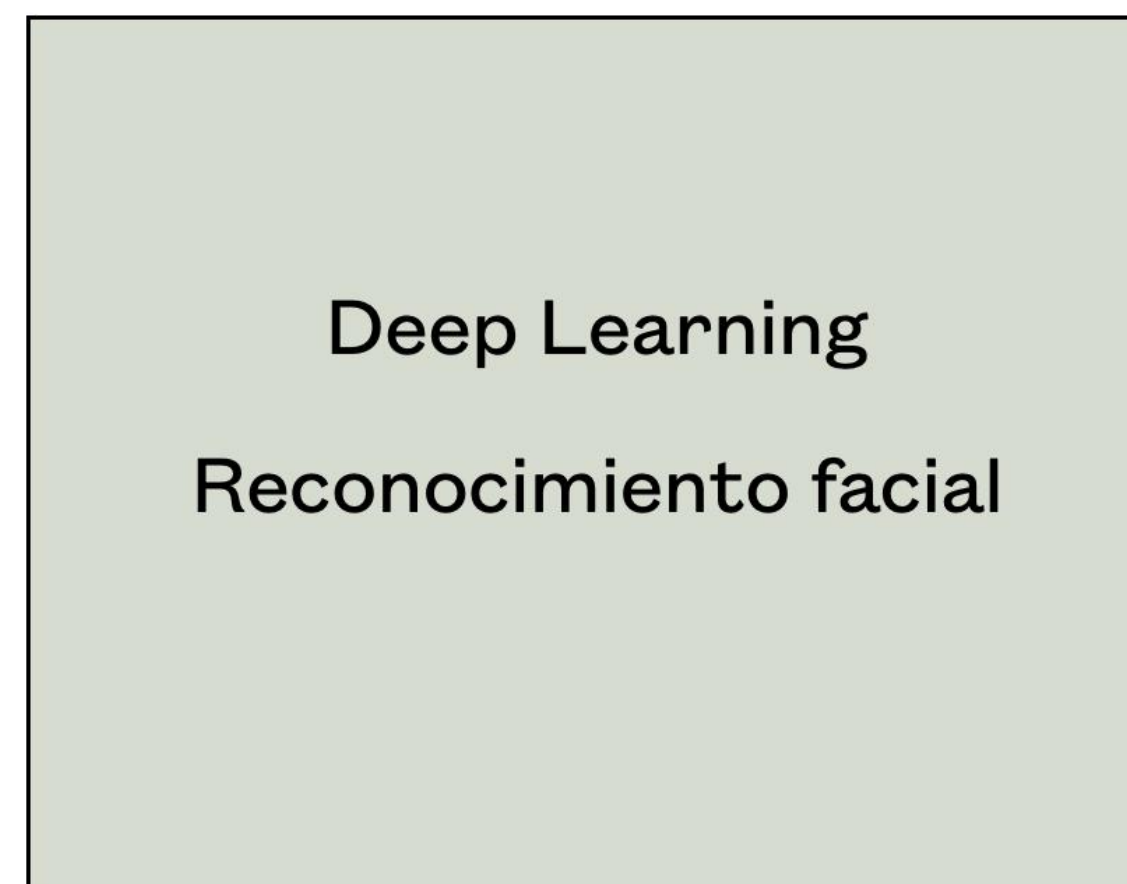
Deep Learning
Reconocimiento facial



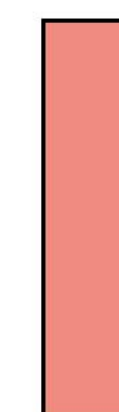
Los dos
vectores
deben ser lo
más
parecidos
posible

Face Recognition Basic Idea

Ejemplos en reconocimiento facial (entrenamiento por triplets)



Estos dos
vectores
deben ser lo
más
parecidos
posible

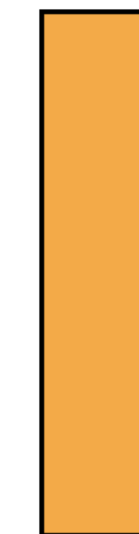


El tercer
vector debe
ser muy
diferente a
los otros dos

Embeddings Concept

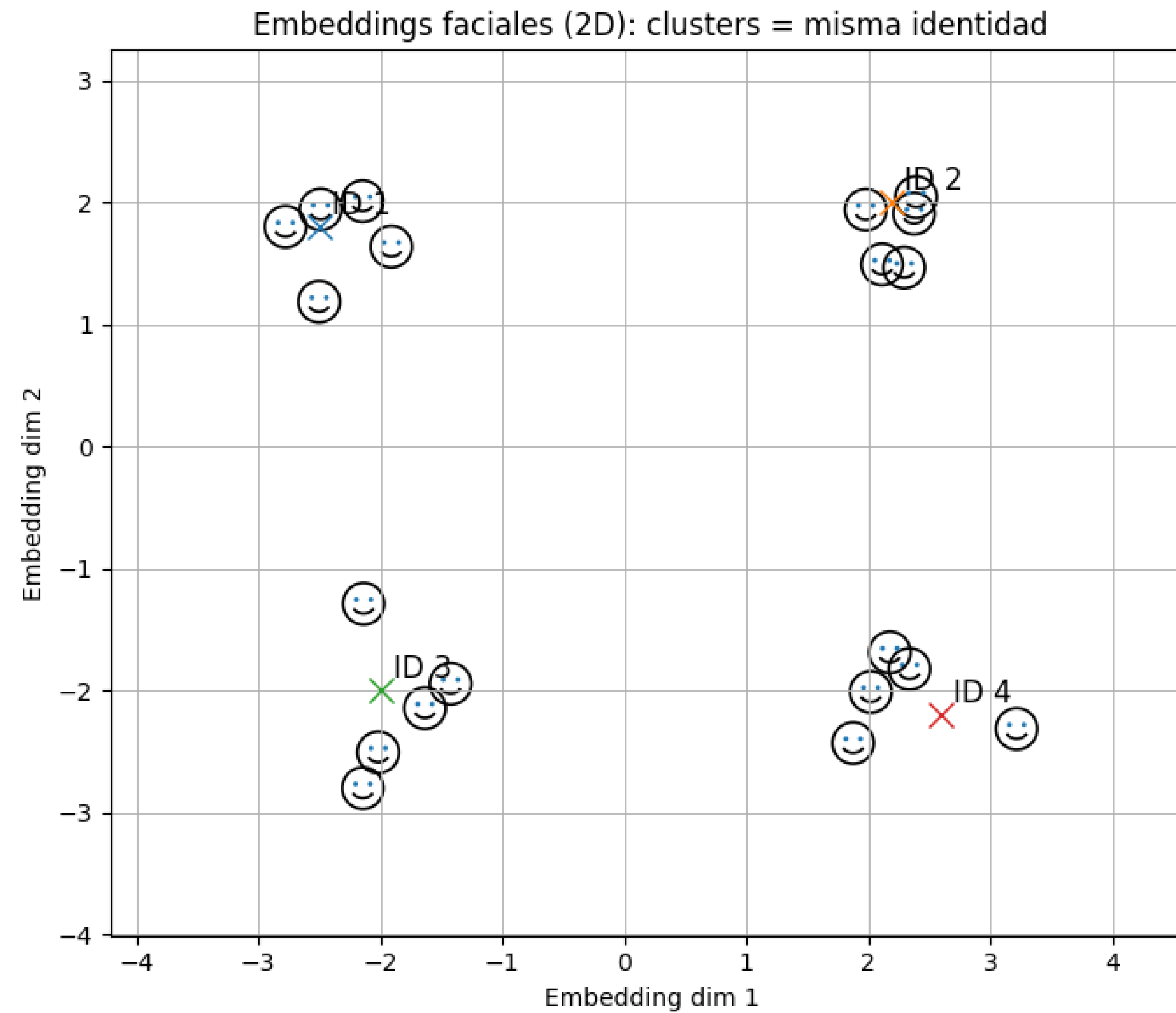


Deep Learning
Reconocimiento facial



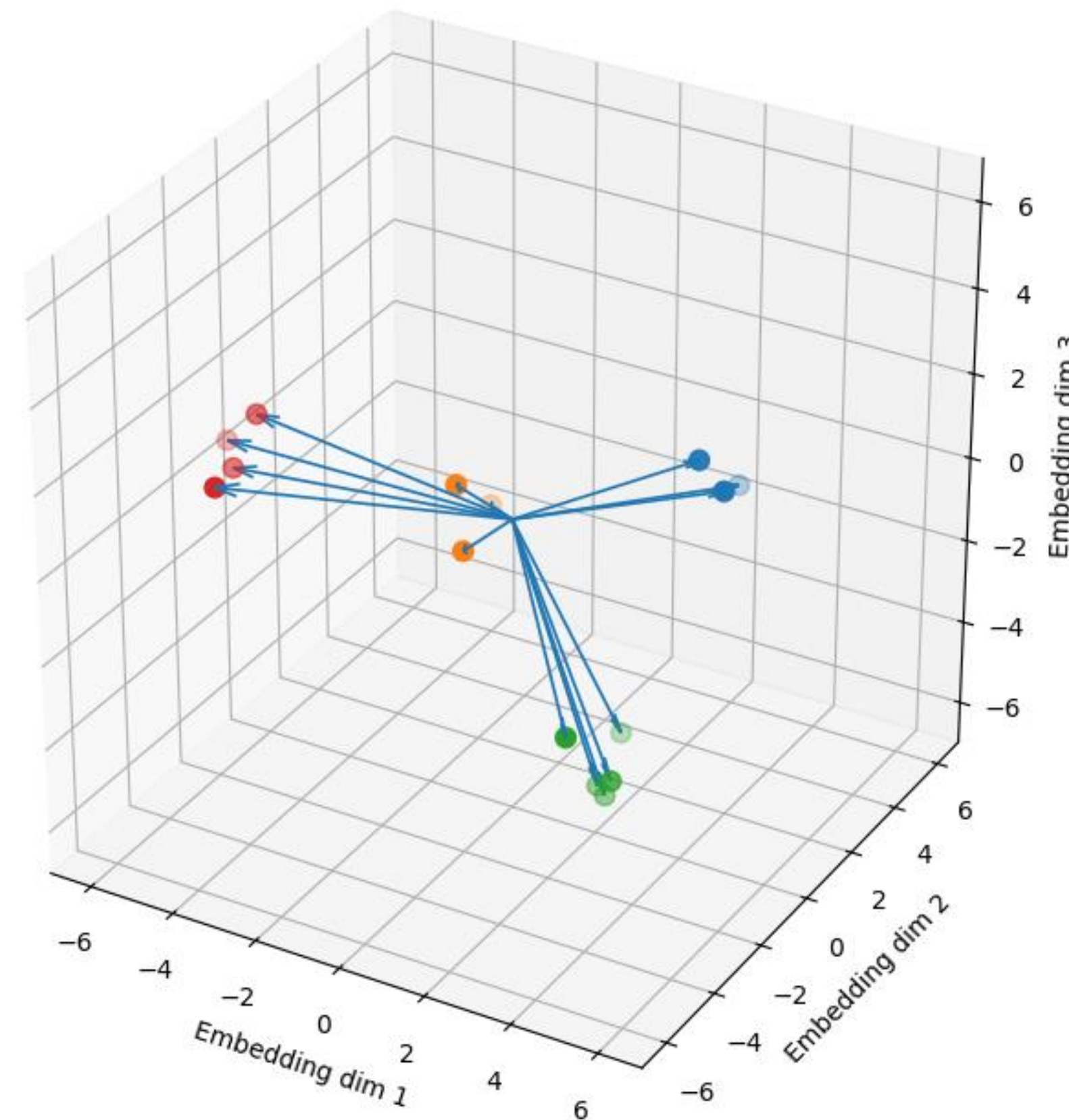
Después del
entrenamiento,
cada cara tiene un
vector asociado

Embeddings Concept



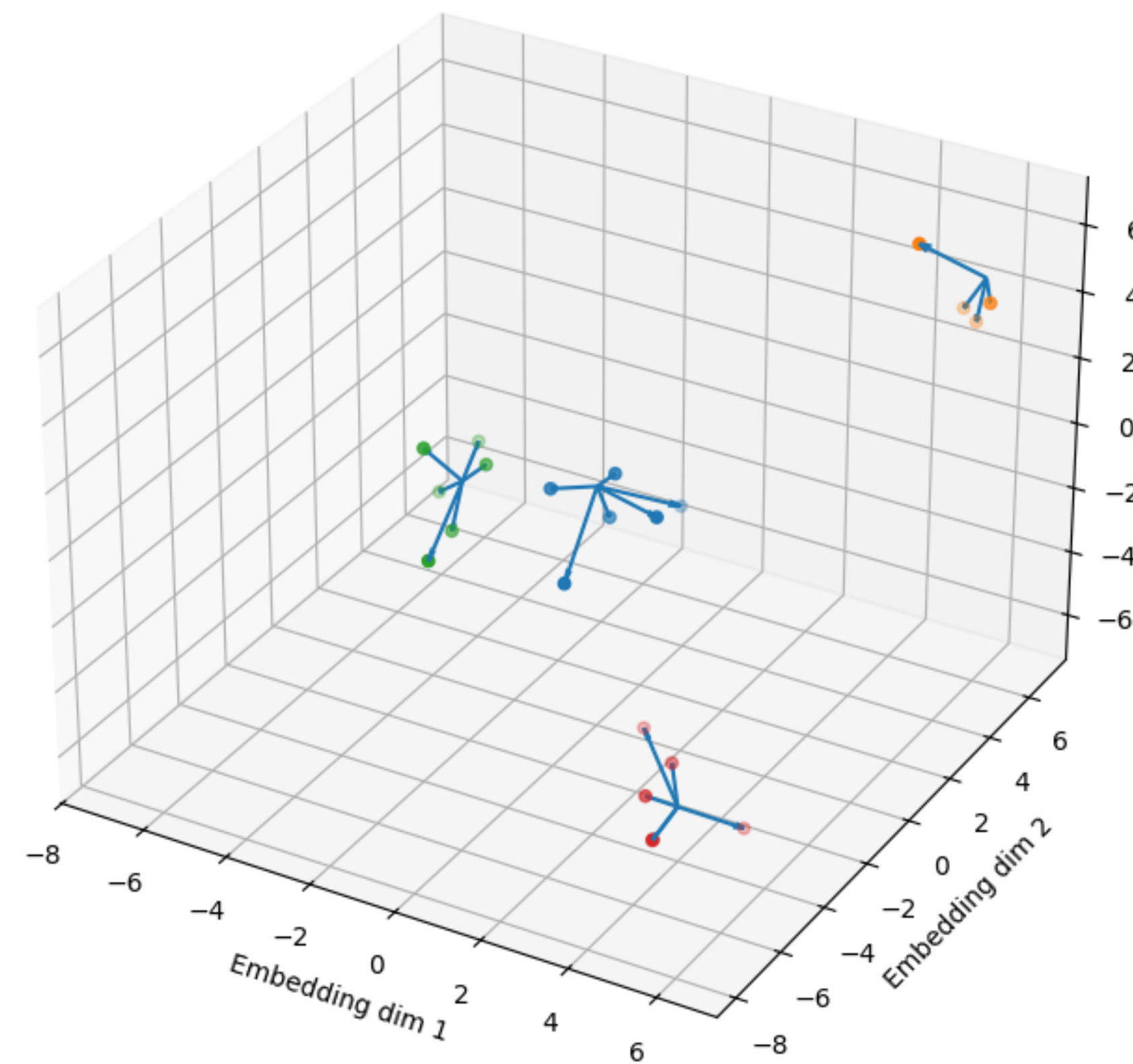
Embeddings Concept

Embeddings faciales
Vectores en espacio 3D (todos desde el origen)

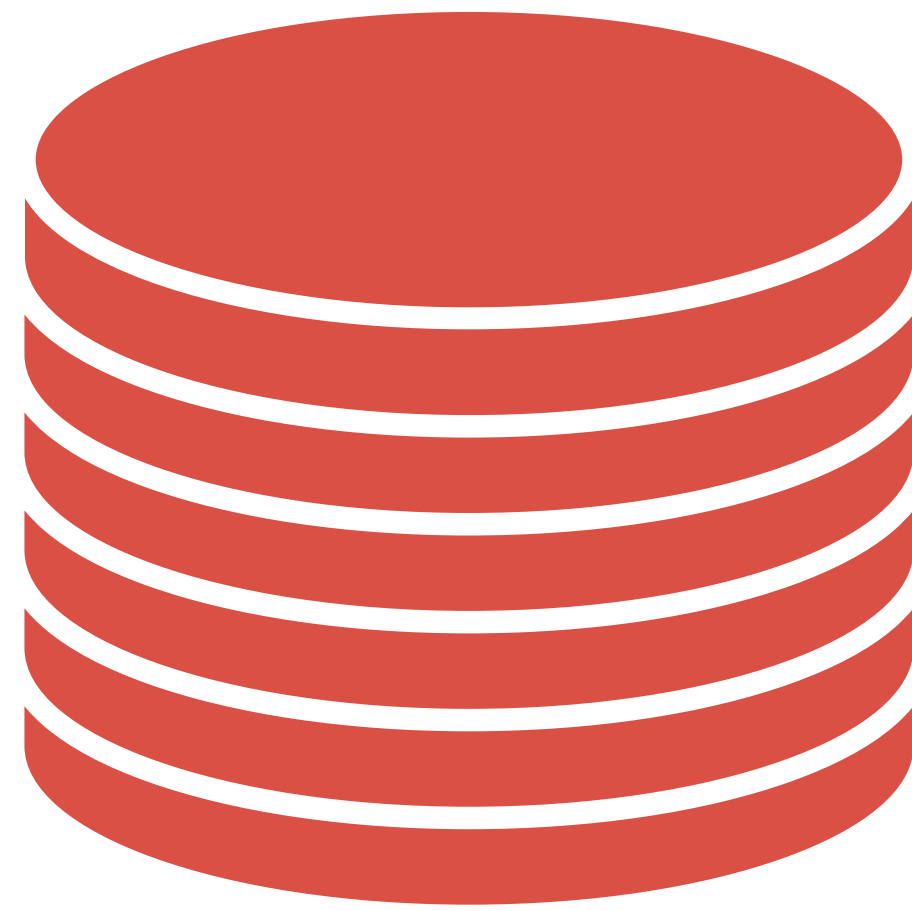


Embeddings Concept

Embeddings faciales en espacio 3D
Cada cluster representa una identidad



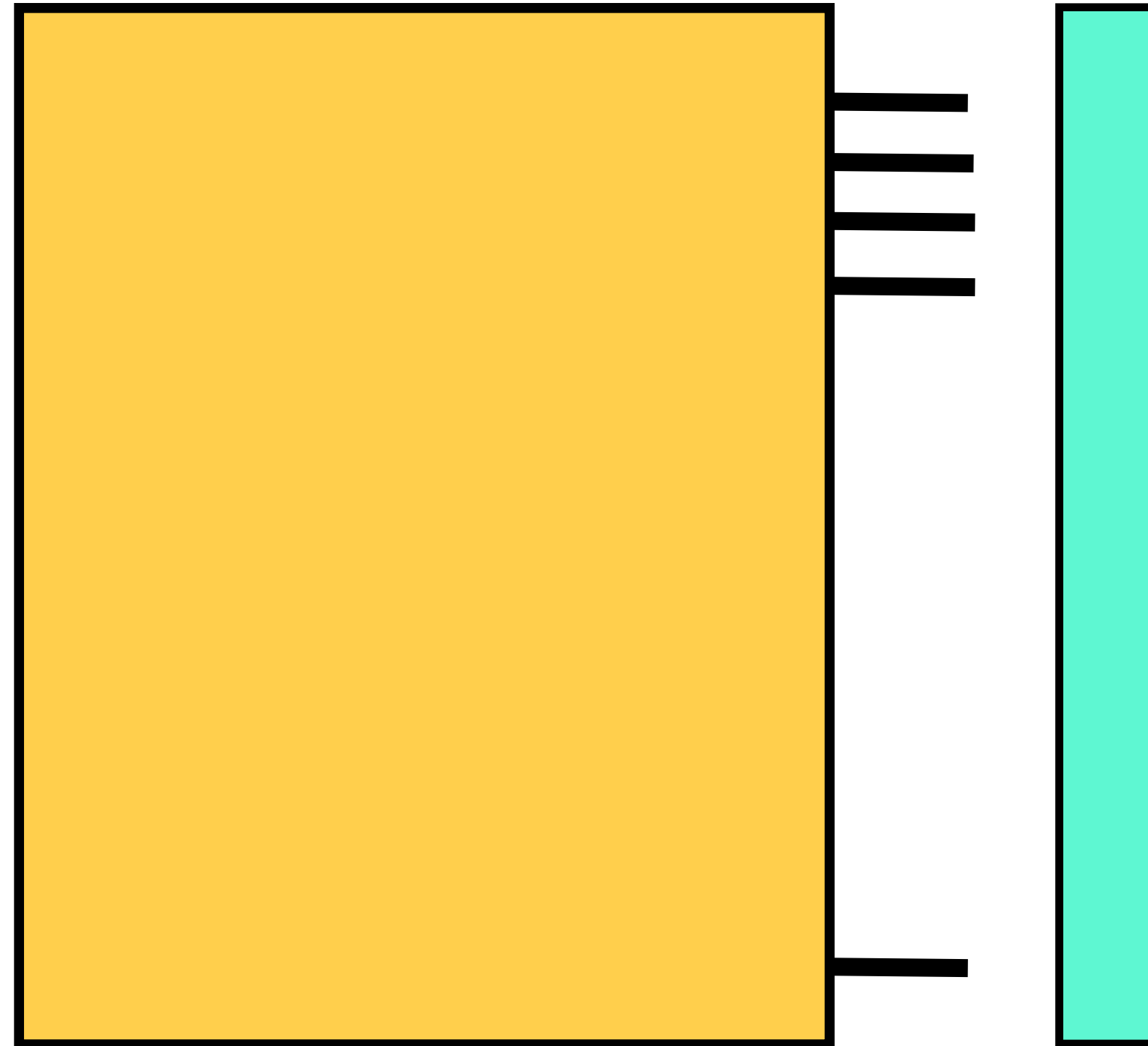
Reconocimiento Facial: Entrenamiento



Base de Datos

Muchas personas

‘Muchas’ fotos de cada persona



QUEREMOS:

1. Entrenar al sistema para que las fotos de una misma persona proporcionen vectores muy ‘parecidos’.
2. Las fotos de diferentes personas proporcionen vectores muy ‘distintos’.

Reconocimiento Facial: Entrenamiento

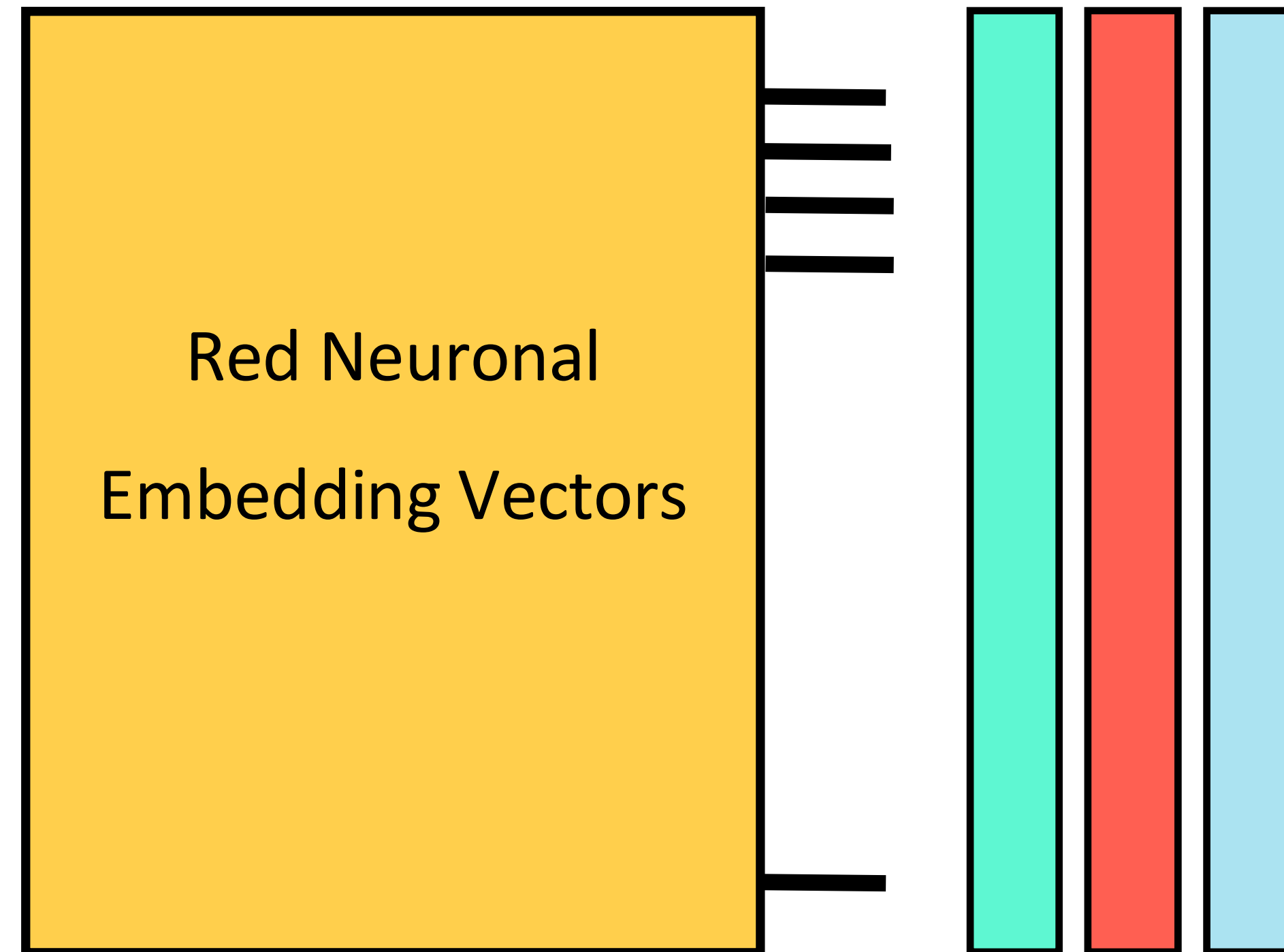
Entrenamiento por Triplets



G1

G2

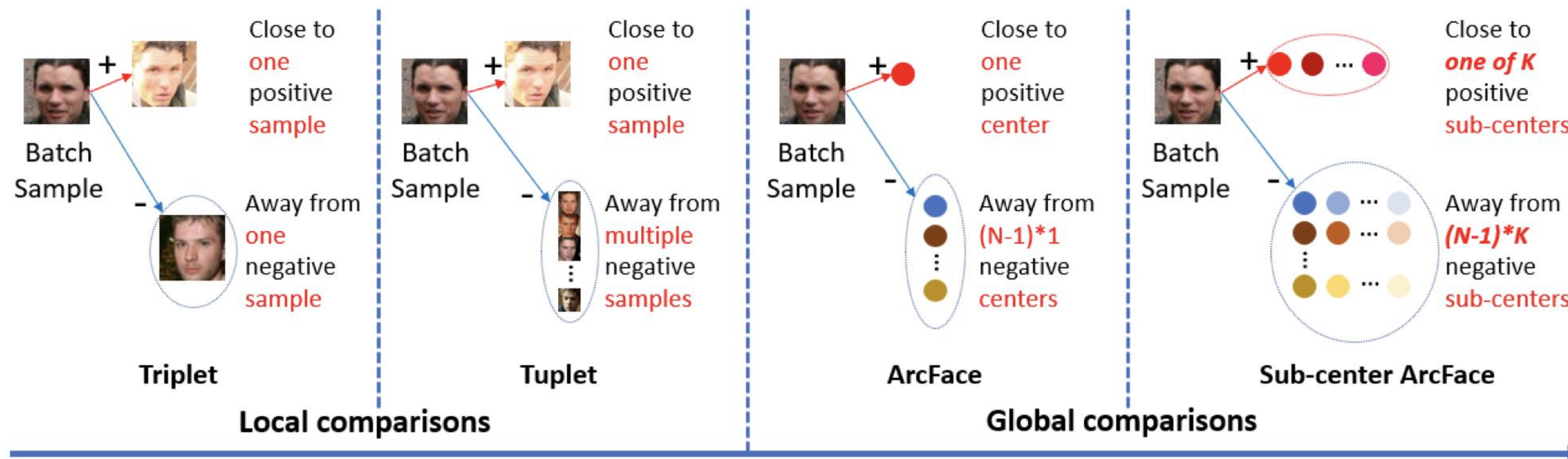
F1



Tendremos que adaptar los parámetros de la red para que:

1. Los vectores asociados a G1 y G2 se parezcan lo más posible
2. Los vectores asociados a G1 y F1 sean lo más diferentes posibles

Training Embeddings



Un aspecto crítico: Las bases de datos

- FERET (1994). Base de datos de imágenes faciales (US Army Research Lab). Conjunto de imágenes de más de 1000 personas en diferentes poses hasta un total de 14.126. (todavía pensadas para clasificar rostros)
- Bases de datos creadas por Universidades (ex. George Mason, UPC-GTAV). Se usan voluntarios (estudiantes, padres, profesores). Se realizan las fotografías con gran calidad (35 mm). Se firman permisos para el uso científico de las imágenes. Se realizan diferentes poses con diferentes iluminaciones en ambientes controlados. El número total de imágenes de cada base de datos es limitado.

Un aspecto crítico: Las bases de datos

UPC GTAV

Face Database



Un aspecto crítico: Las bases de datos

UPC GTAV
Face Database

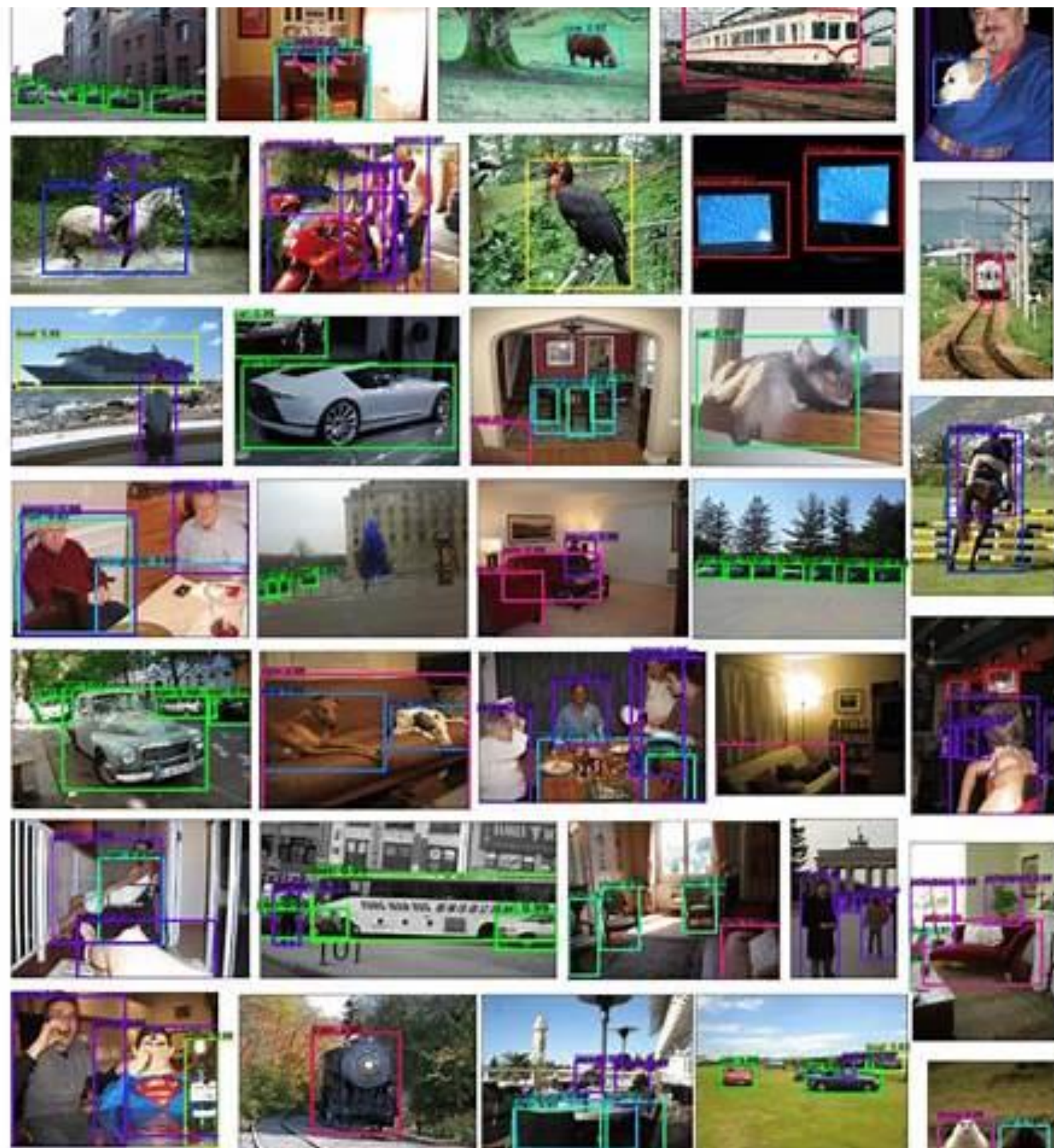


Illumination

Facial Expression

Occlusion

Historia de las Bases de Datos: ImageNet



PASCAL VOC

20 categorias

Person (4000-5000)

Bird, Cat, Cow, Dog (1500-2000),
Horse Sheep

Aeroplane, bicycle, boat, bus, car
(2000-3000), motorbike, train

Bottle, chair, dinning table, potted
plant, sofa, tv/monitor

Historia de las Bases de Datos: ImageNet

Datos Generales ImageNet

Número total de imágenes	> 14M
Número de categorías	aprox 21.000
Número de imágenes etiquetadas con BB	aprox 1M
Idiomas de etiquetas.	inglés, basadas en wordnet
Tamaño del subconjunto ILSVRC	1.2 M imágenes, 1000 clases
Imágenes validación ILSV	50.000
Imágenes de prueba (etiquetas ocultas)	100.000

Historia de las Bases de Datos

- INTERNET, EL RECURSO NATURAL
 - Compartir fotos en redes sociales: millones de selfies, condiciones de luz, posición, pose, profundidad de campo.
 - Se comparten fotos de bebés, familias, fotos antiguas, etc.
 - Facebook (2020) dispone de 350M fotografías nuevas cada día.
 - Twitter (2020). 500M de tweets
 - Las fotografías se etiquetan alegremente, sin cobrar, se dispone de datos etiquetados para entrenar los modelos de IA.
-

Consideraciones sobre los DATOS

Fotos de archivos policiales (publicas)

No quedan claros los cargos, si fueron absueltas .. son simplemente datos. Muchas imágenes tienen moratones, signos de violencia.

Esta bases de datos han pasado de ser un archivo policial a un punto de referencia técnico.

Detección de tipos criminales!!!!!!

Los datos no tienen contexto, historia, naturaleza de la persona

Acaban convirtiendose en un concurso de algoritmos, un challenge

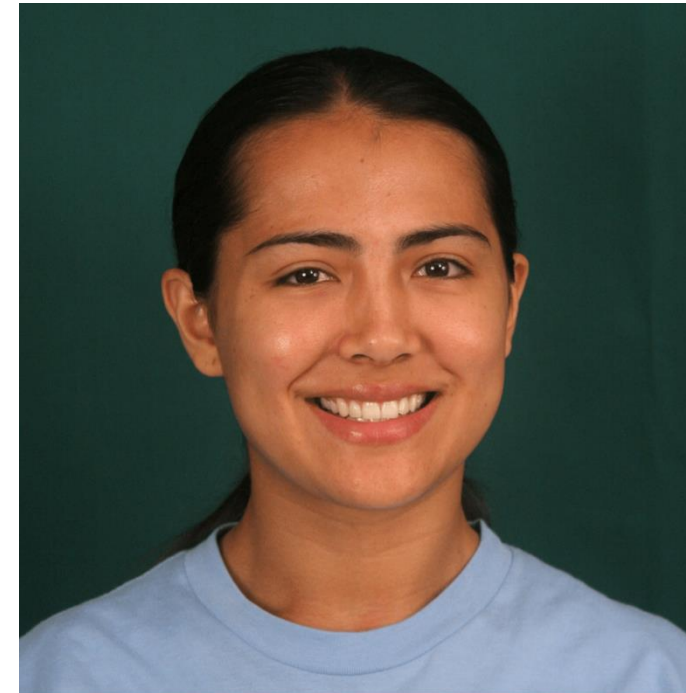


Consideraciones sobre los datos

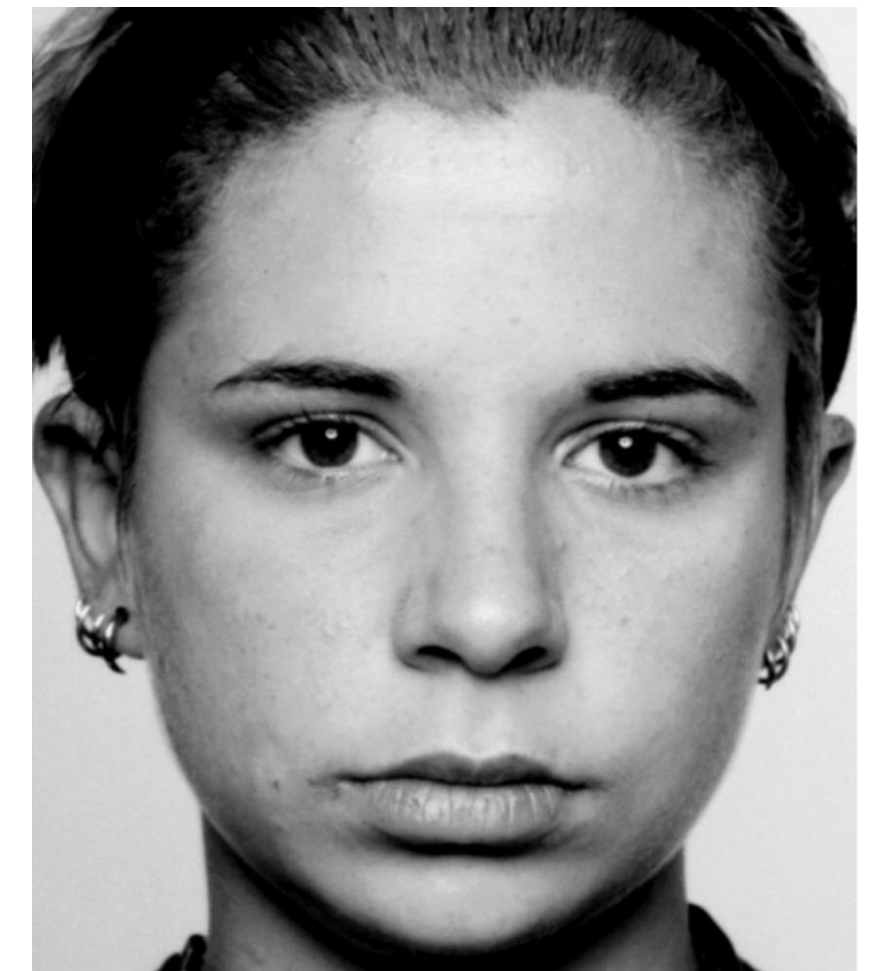
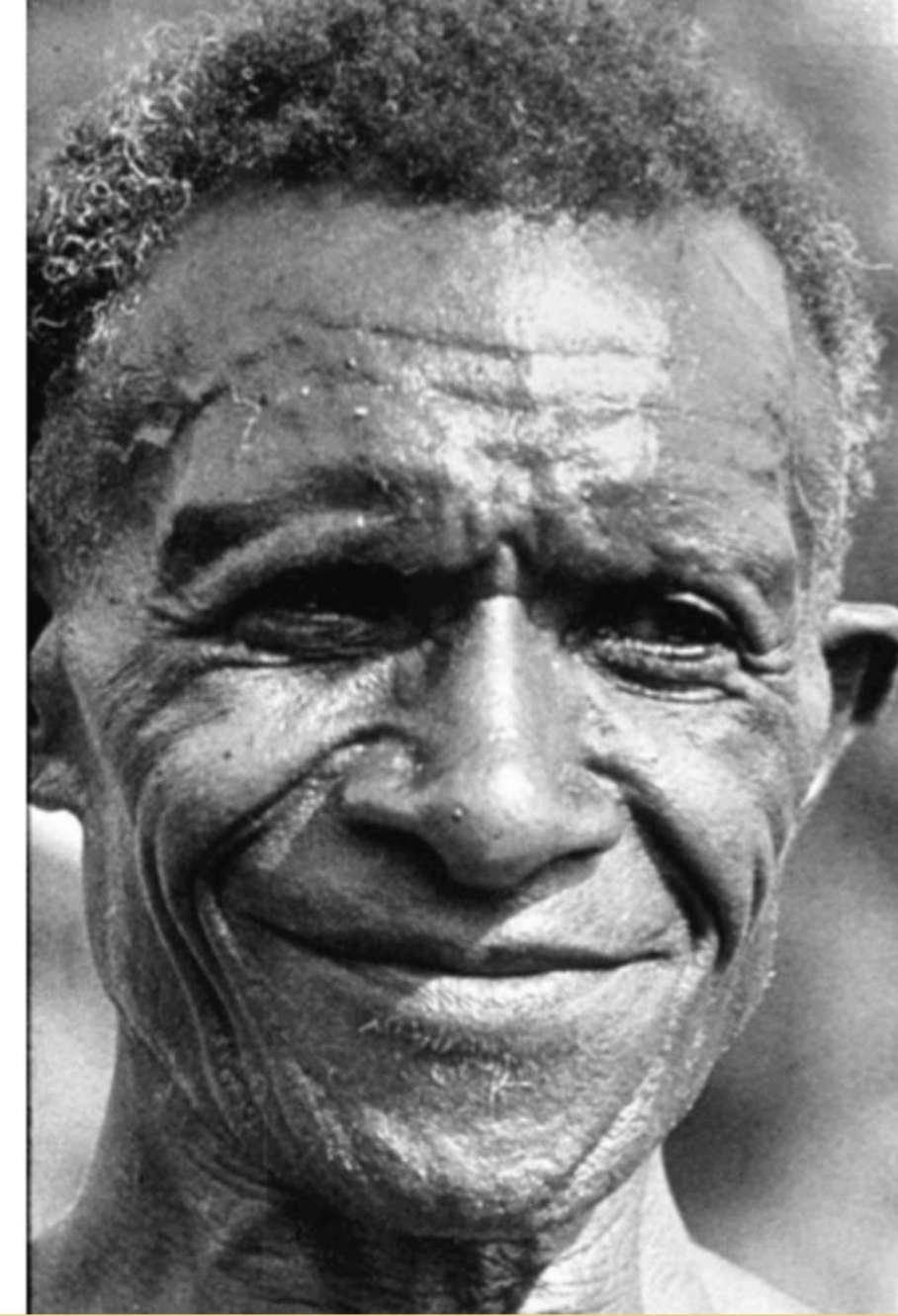
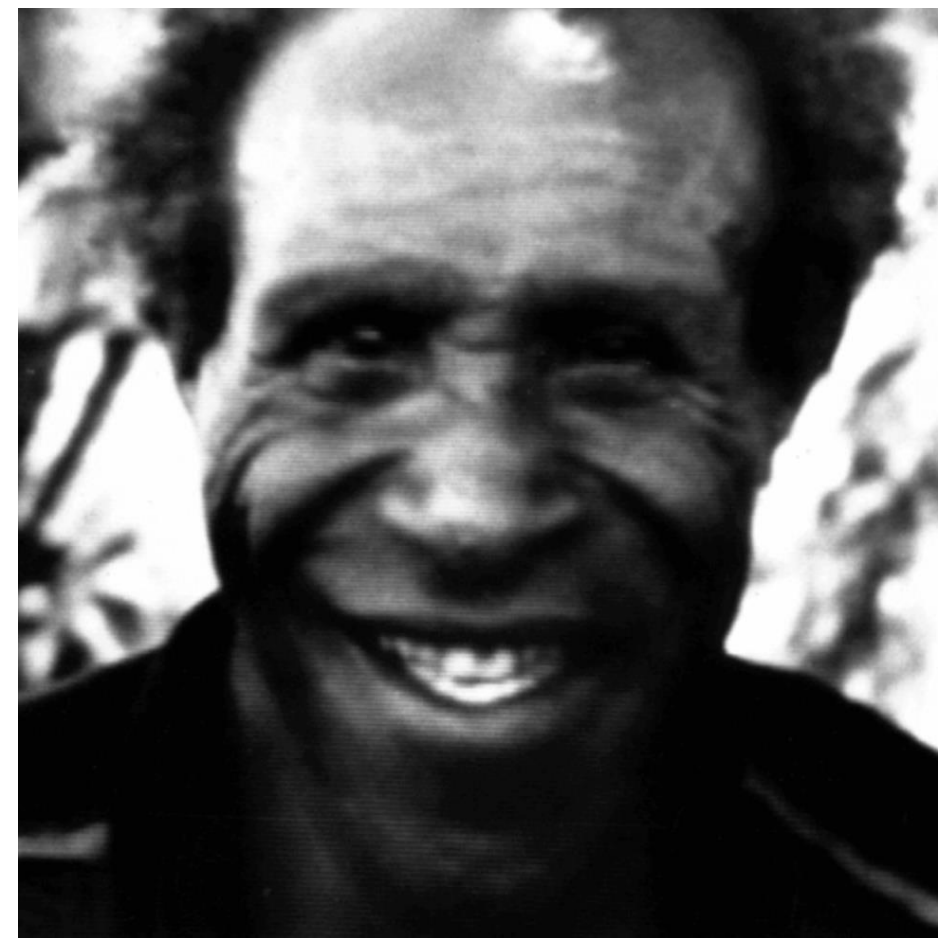
Detección Emociones



Paul Ekman



Happiness



Consideraciones sobre los datos

Detección Emociones



Paul Ekman



Legalidad de la IA

- La legislación europea plantea que los embeddings biométricos pueden significar problemas legales importantes en 3 marcos regulatorios
 - Ley General de Protección de Datos
 - EU Artificial Intelligence Act
 - Law Enforcement Directives (policia)
- Los embeddings se consideran datos biométricos ya que permiten identificar a una persona.
 - Datos específicamente protegidos
 - Su tratamiento está prohibido salvo excepciones que requieren bases legales explícitas: consentimiento explícito, interés público esencial, seguridad pública.

Legalidad de la IA

- El reconocimiento en espacios públicos está prácticamente prohibido. Solo se autoriza para terrorismo o búsqueda de desaparecidos.
- Los embeddings no son anónimos. Con acceso a la base de datos y un modelo similar puede hacerse re-identificación. Se consideran datos pseudoanonimizados.
- Los sistemas de reconocimiento tienen sesgo: funcionan peor con mujeres, personas de piel oscura, grupos demográficos minoritarios. El Act exige el análisis de riesgo, sesgo y documentación técnica

Legalidad de la IA

- Al Act exige proporcionalidad. No es posible usar reconocimiento facial o de huella si existe una alternativa menos intrusiva: tarjeta RFID, Código QR, PIN.
- El problema no es tanto la tecnología de reconocimiento facial sino el riesgo social de vigilancia biométrica.
- Es necesario considerar aspectos como autenticación (1:1) o identificación (1:N)

Legalidad de la IA

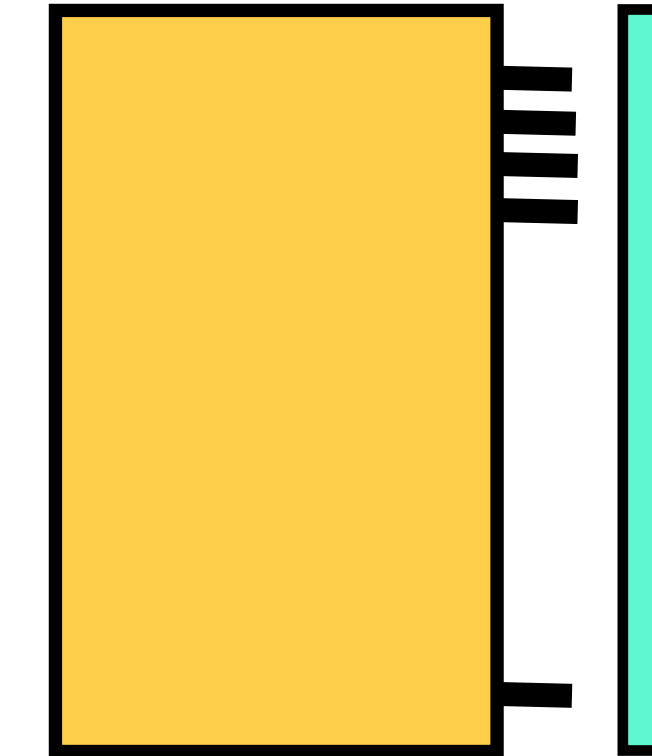
- El RGPD permite tratar datos biométricos si: la persona ha hecho manifiestamente públicos estos datos: actores, presentadores, políticos, personas públicas en medios.
- Existe cierta básica jurídica para que el reconocimiento para la extracción de metadatos públicos pueda realizarse. Se considera más parecido a documentación audiovisual o indexación periodística.
- La base de datos no puede contener personas no públicas (figurantes, público).
- Comparar si es uso 'interno' o 'externo'. Buscar donde aparece Pedro Sanchez vs vender base de datos de apariciones.
- SISTEMA DE INDEXACIÓN AUTOMÁTICA DE IDENTIDADES EN CONTENIDOS AUDIOVISUALES EMITIDOS PÚBLICAMENTE VS **SISTEMA DE RECONOCIMIENTO FACIAL.**

Ejemplo ClearView (prohibido en europa)

- **Proceso:**
 - 1) Foto tomada por policia
 - 2) Subir a ClearView
 - 3) Identidad probable
 - 4) Links a perfiles sociales
- La base de datos se obtuvo con 30.000 millones de imágenes extraídas de redes sociales con scrapping automático de internet.
- Las sanciones fueron de millones de euros y se tuvieron que borrar todos los datos de ciudadanos europeos y no se permitió operar en europa.

Como Construir un Archivo

1. Se supone que tenemos entrenado un sistema que asigna vectores a una cara (REC)

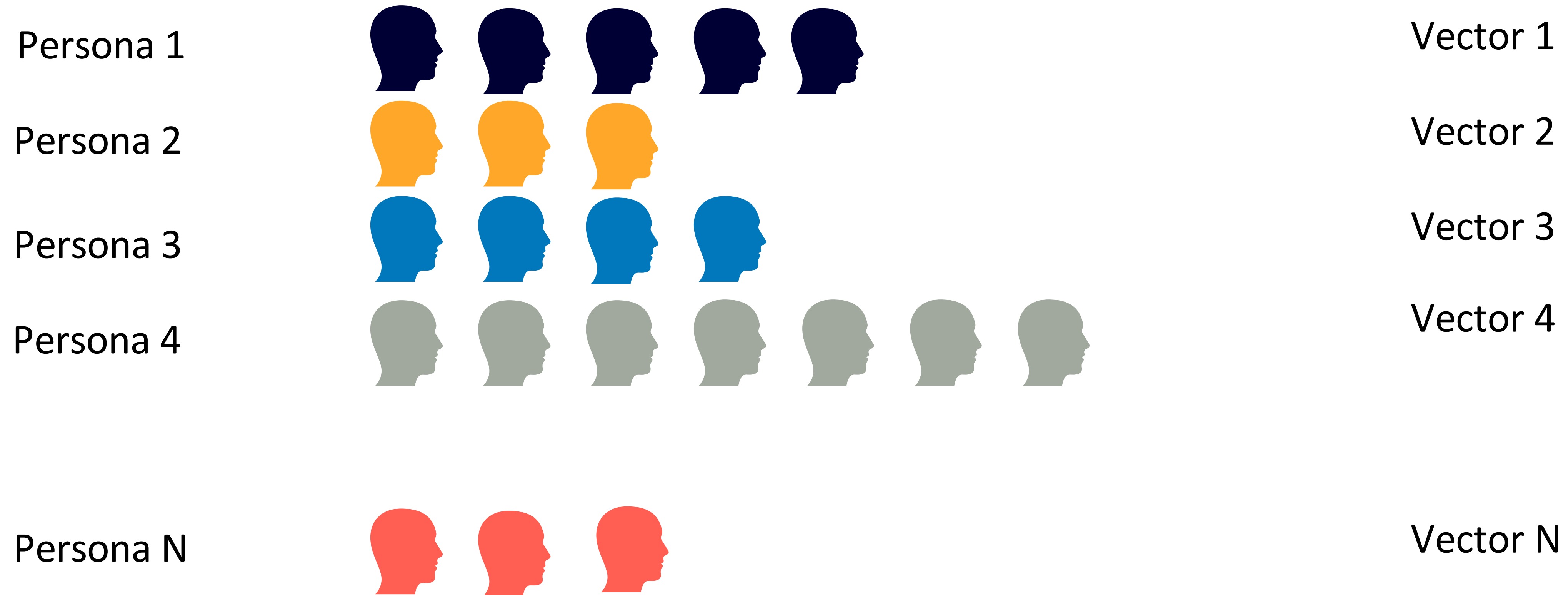


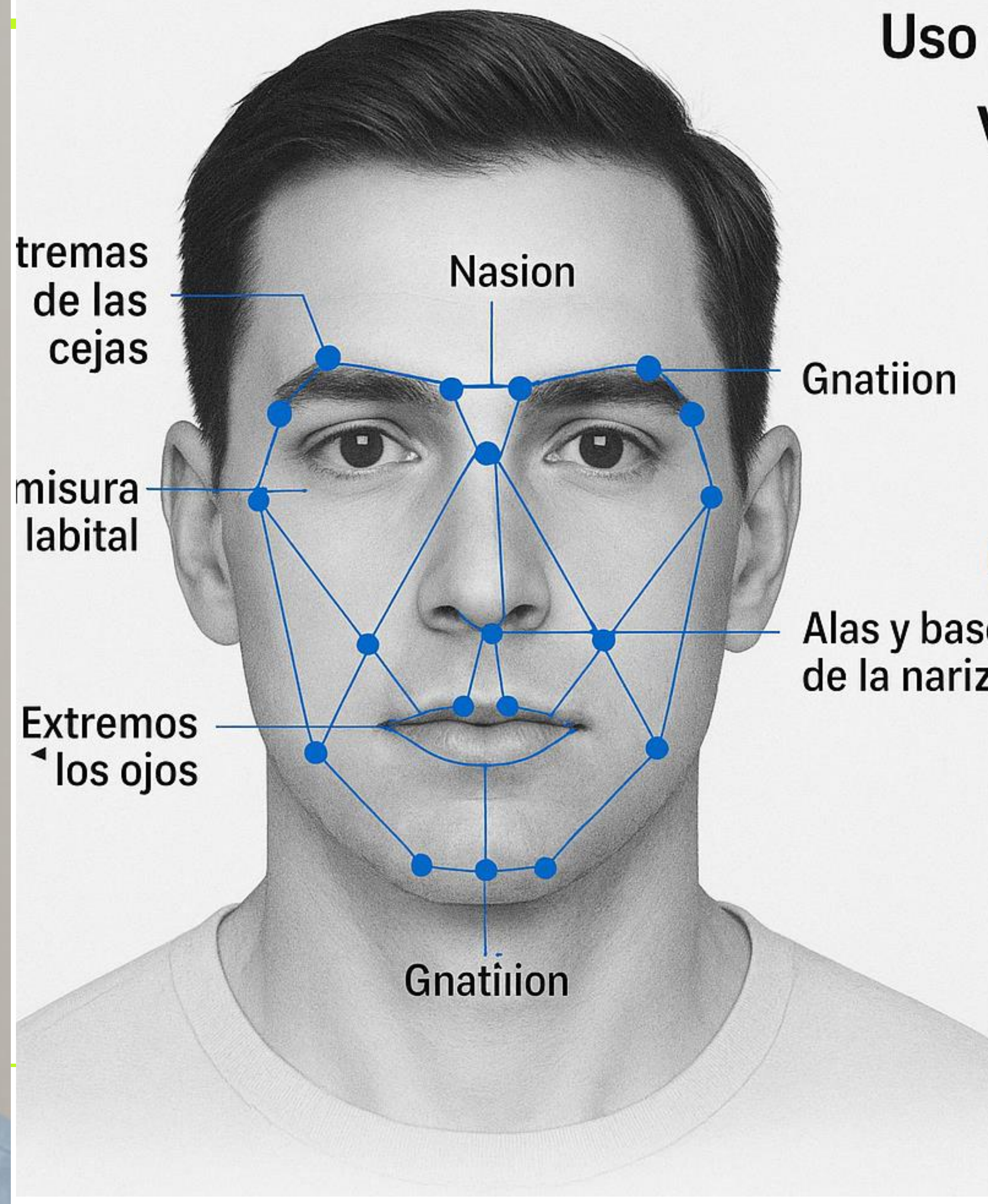
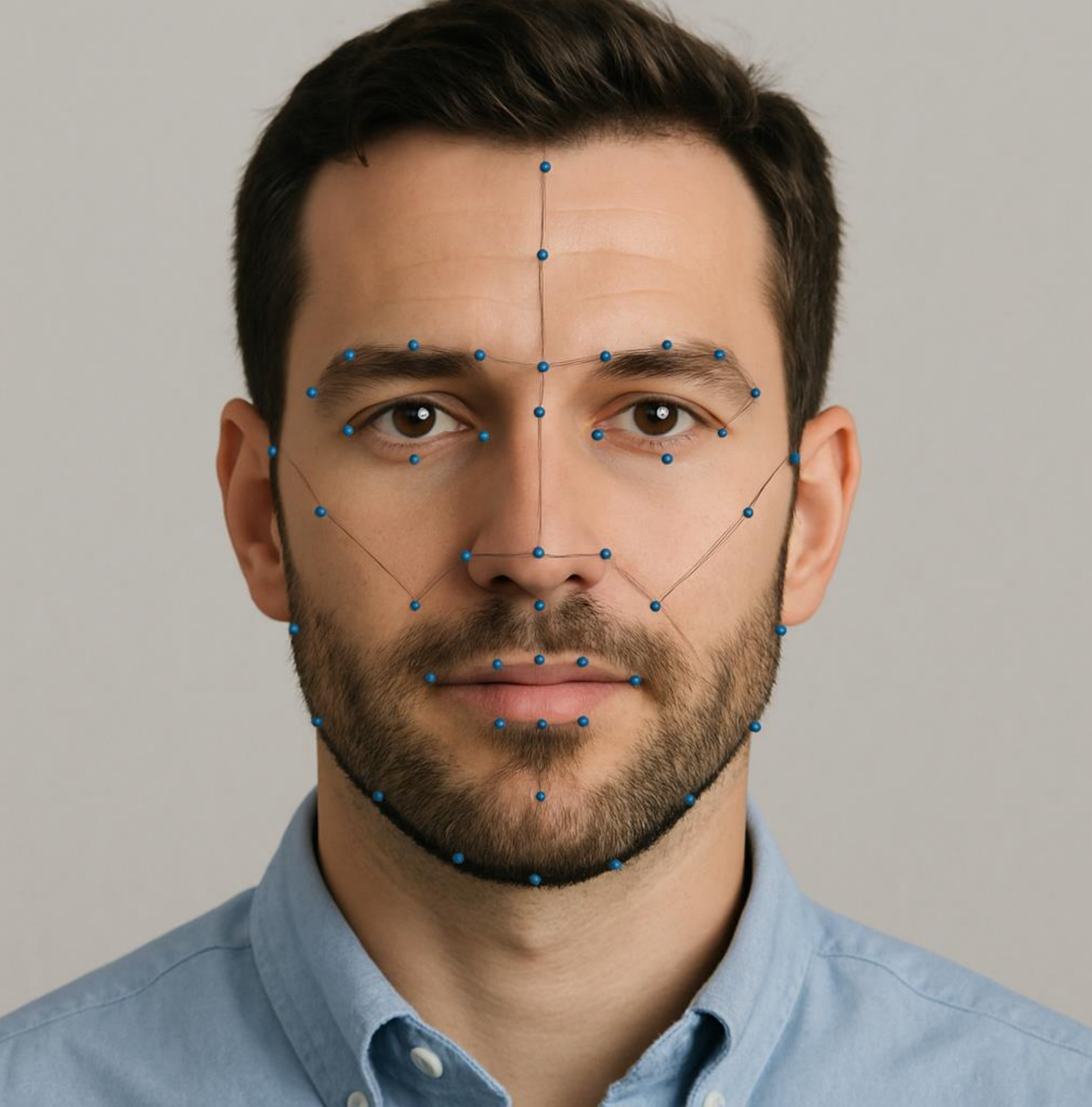
Vector J

Medir distancias entre Vector J y los vectores de nuestra base de datos

Como Construir un Archivo

1. Se supone que tenemos entrenado un sistema que asigna vectores a una cara (REC)







APIs Reconocimiento Facial (Cloud)

- Amazon Rekognition. Celebrities. Facturación por imagen
 - Microsoft Azure. Detección, verificación, agrupación, edad, emoción. Facturación por imagen
 - Google Cloud Vision API. Rostros y landmarks, puntuaciones emoción
 - Clarifai. Celebrities
 - Kairos. Detección, verificación, identificación, liveness. Licencias por volumen
 - Banuba Face API. Para aplicaciones móviles, seguimiento
-

APIs Reconocimiento Facial (On-P)

- Neurotechnology VeriLook
- Cognitec FaceVACs
- Kuxand FaceSDK
- Face ++



Taller Face Recognition: DeepFace

- Librería Python de alto nivel para análisis facial.
 - Permite ejecutar tareas habituales de visión artificial con pocas líneas de código:
 - Detección de caras
 - Extracción de embeddings
 - Reconocimiento facial
 - Verificación de identidad
 - Análisis de atributos: edad, género, etnia aparente
-

DeepFace

- `DeepFace.extract_faces()` - Detectar y recortar caras en una imagen. (Localizar, BB, alinear)
 - `DeepFace.analysis()` - Analizar atributos faciales (edad, género, emoción dominante, etnia, etc.)
 - `DeepFace.verify()` - Comparar 2 caras (2 caras pertenecen a la misma persona?)
 - `DeepFace.find` - Buscar cara en base de datos (Identificación 1:N contra una carpeta)
 - `DeepFace.represent()` - Extraer embeddings faciales (vectores numéricos para representar la cara)
 - `DeepFace.stream()` - Análisis en tiempo real (webcam, análisis en tiempo real continuo)
-

DeepFace

- Selección de métodos:
 - Detección: opencv, retinaface, mtcnn, ssd, dlib,mediapipe, yolov8, centerface, skip
 - Reconocimiento: VGG-Face, FaceNet, OpenFace, DeepFace, DeepID, ArcFace, Dlib, SFace, GhostFaceNet

```
DeepFace.analyze(  
  
    img_path="imagen.jpg",  
  
    actions=["age", "gender", "emotion", "race"],  
  
    detector_backend="opencv",  
  
    enforce_detection=True  
  
)
```

```
DeepFace.analyze(  
  
    img_path="imagen.jpg",  
  
    actions=["age", "gender", "emotion", "race"],  
  
    detector_backend="opencv",  
  
    enforce_detection=True  
  
)
```

DeepFace

Desarrollada por Sefik Ilkin Serengil como herramienta Python de alto nivel para facilitar el uso de modelos de reconocimiento facial y análisis de atributos.

DeepFace no es un único modelo, es una capa de integración con diferentes modelos, que simplifica su uso en diferentes etapas del pipeline facial.

Es importante distinguirla del DeepFace de Facebook (2014) que fue el primer modelo de reconocimiento facial basado en DL.

En la librería actual DeepFace actúa como una capa de extracción que permite cambiar de modelo sin modificar demasiado el código.

Los resultados deben interpretarse con cautela, especialmente la estimación de edad, género, emoción o etnia que deben interpretarse como una inferencia estadística del modelo.

Vamos a hacer un poco de ejercicio



Descarga del taller Face Recognition



<https://bit.ly/4gW8xC6>