Reconocimiento Facial para Seguridad

Identificación de personas autorizadas y clasificación de género

Jorge Quelas - Data Science II - Coderhouse







Contenido de esta presentación



Intro / Contexto

¿Qué impulsa este trabajo y qué existe en el mercado?



Alcance

¿Qué vamos a hacer?



Objetivo¿Cual es nuestro objetivo?



Dataset Elegido

Labeled Faces in the Wild (LFW)

Reconocimient o de Personas

Este proyecto busca identificar personas autorizadas para circular en un determinado espacio. Se utilizará ML para generar modelos que se conecten a cámaras detecten personas no autorizadas en tiempo real.









Ver video

Página Oficial



Desarrollos en Otras Empresas



también están en desarrollo de nuevos sistemas de seguridad basados en Al, dado al impacto de la tecnología en el sector.

Objetivo

Este projecto busca desarrollar un sistema basado en algoritmos de ML para detectar personal autorizado en áreas de interés usando un modelo de clasificación



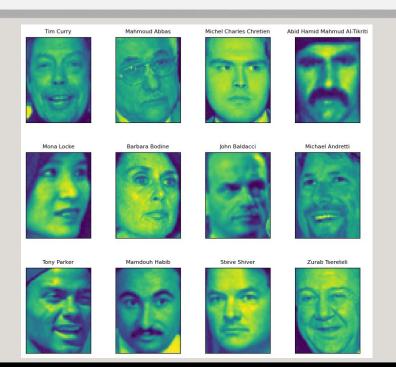


Alcance

- Crear un modelo clasificador de rostros, para determinar personas autorizadas de no autorizadas.
- Clasificar los rostros capturados no solo por identidad, sino también por género.
- Entrenamiento de modelos de reconocimiento facial supervisado.
- Evaluación de métricas de desempeño (accuracy, precision, recall).

Labeled Faces in the Wild





- Para entrenar el modelo se requiere un dataset de caras que simularán al personal autorizado.
- Scikit-Learn ofrece un dataset llamado Labeled Faces in the Wild que posee caras de personalidades famosas sacadas de internet y con etiquetas con sus nombres.
- No posee etiquetas por género.

Labeled Faces in the Wild

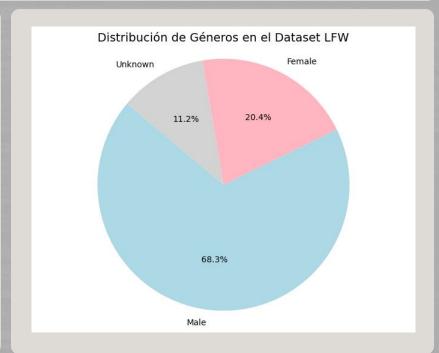
Algunas características del dataset son:

- Número total de imágenes: 13.233 rostros.
- Número de personas: 5.749 individuos únicos.
- Personas con 2 o más imágenes: 1,680 personas.
- Tamaño original de las imágenes: 250 × 250 píxeles (ajustable).
- Fuente de imágenes: Fotografías reales recolectadas de internet.
- Detección de rostros: Utiliza el detector de Viola-Jones para centrar las caras.
- Etiquetas: Cada imagen está etiquetada con el nombre de la persona correspondiente.

Métricas Adicionales del Dataset

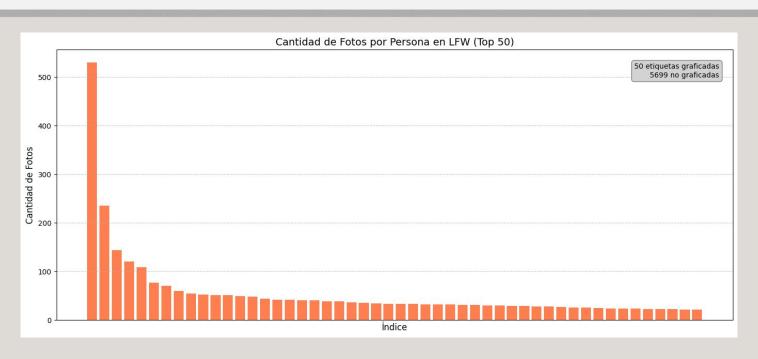


- Al dataset se le debe agregar etiquetas de género ya que no posee inicialmente.
- Se usa Wikipedia como fuente inicial para determinar automáticamente el género.
- Dado que no se logran obtener los géneros de todos los nombres, se corre un segundo proceso donde se estima el género en base al nombre de pila. Si una persona en el dataset se llama "James" y fue identificado inicialmente como de género Masculino, se asume que todos los "James" son de género Masculino.





Cantidad de Fotos por Persona



Métricas Adicionales del Dataset



Métricas sobre cantidades de fotos:

- Media (promedio): 2.30
- Mediana: 1
- Moda: 1
- Mínimo: 1
- Máximo: 530
- Primer cuartil (Q1): 1.0
- Segundo cuartil (Q2): 1.0
- Tercer cuartil (Q3): 2.0

Top 10 personas con más fotos:

- George W Bush (530)
- Colin Powell (236)
- Tony Blair (144)
- Donald Rumsfeld (121)
- Gerhard Schoeder (109)
- Ariel Sharon (77)
- Hugo Chavez (71)
- Junichiro Koizumi (60)
- Jean Chretien (55)
- John Ashcroft (53)



Top 10 de Personas

