UNIVERSITÀ
DEGLI STUDI
DI MILANO

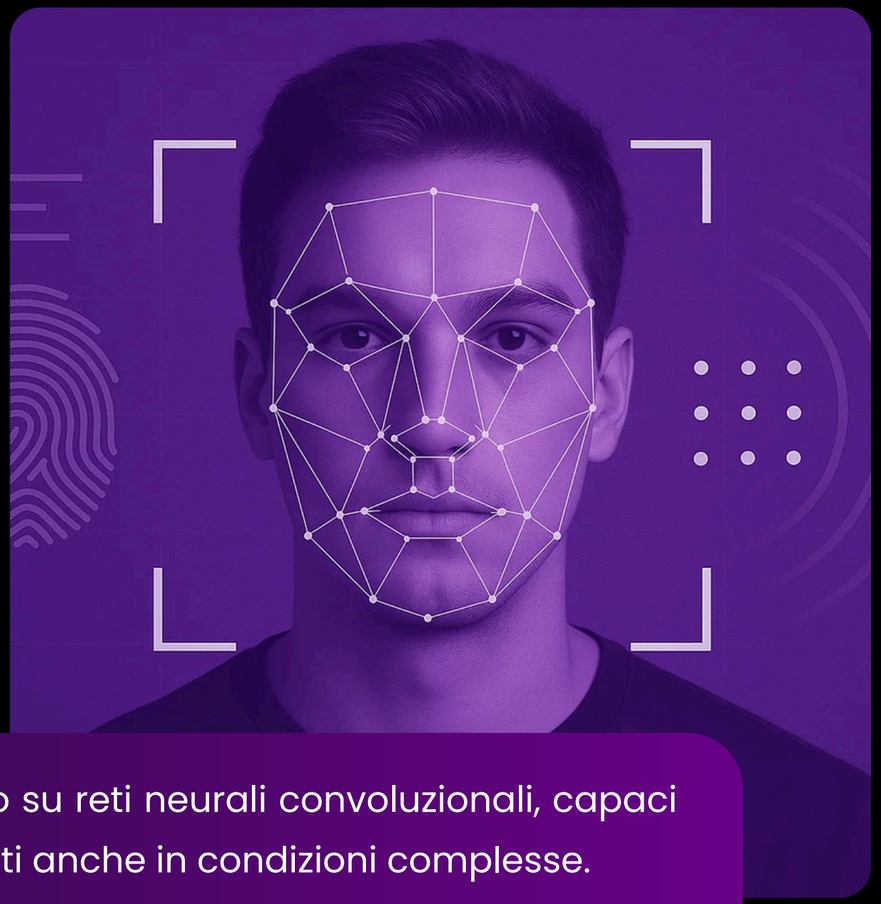
FACE RECOGNITION

PRINCIPI E MODELLI DELLA PERCEZIONE

Hlibey Antonio
Confuorto Stefano
Marinescu Raoul Nicolae

Cosa è?

È una tecnologia biometrica che identifica o verifica l'identità di una persona analizzando i tratti del volto. Utilizza algoritmi di ML/AI per rilevare volti, estrarre caratteristiche uniche e confrontarle con immagini salvate in un database.



Gli algoritmi moderni si basano su reti neurali convoluzionali, capaci di apprendere e distinguere volti anche in condizioni complesse.
È una tecnologia in continua evoluzione, utilizzata in vari ambiti.

Applicazioni e casi d'uso

Il riconoscimento facciale è sempre più presente nella nostra vita quotidiana, combinando tecnologia e sicurezza per semplificare e automatizzare numerose attività

Esempi concreti d'uso:

- Sblocco dei device
- Controllo accessi in aeroporti e stadi
- Videosorveglianza e sicurezza urbana
- Tracciamento dei clienti e analisi comportamentale
- Organizzazione automatica delle foto

Tecnologie avanzate nella face recognition

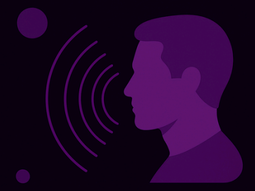
Infrared 3D face recognition

Proiezione di punti a infrarossi e mappa 3D del volto per un riconoscimento sicuro, anche al buio e resistente a foto e maschere.

Anti-Spoofing multimodale + DL

Tecniche anti-spoofing multimodali basate su deep learning combinano dati visivi, termici e di profondità per rilevare tentativi di inganno tramite foto, video o maschere.

Time-of-Flight / LiDAR

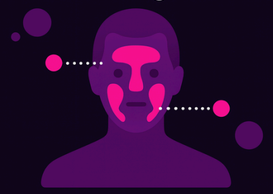


Tecnologia di profondità che misura la distanza con impulsi luminosi, creando modelli facciali dettagliati in 3D ad alta precisione.

Fusion sensoriale (RGB-Depth-Thermal)

Reti neurali profonde analizzano simultaneamente dati RGB, di profondità e termici per distinguere volti reali da foto, video o maschere.

Thermal IR face recognition



Analisi del calore emesso dal viso per identificare tratti fisiologici unici, utile per distinguere anche gemelli o immagini false.

Riconoscimento umano vs macchina

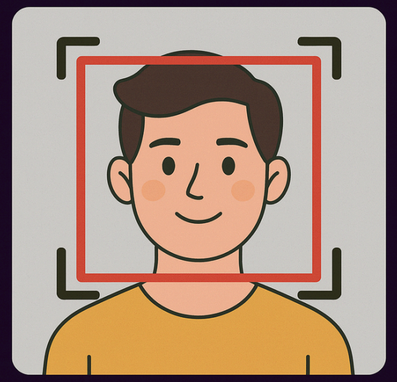
Aspetto	Riconoscimento Umano	Riconoscimento Macchina (CNN)
Metodo	Percezione visiva + euristiche biologiche	Visione computazionale + deep learning
Meccanismo principale	Area fusiforme dei volti (FFA), scorciatoie evolutive	Estrazione feature e embedding numerico
Adattabilità	Alta	Media
Bias	Etnia, genere, familiarità, pareidolia	Dipendenza dal dataset
Robustezza ai cambiamenti	Elevata	Limitata
Riconoscimento volti simili	Bassa	Alta
Flessibilità	Elevata	Bassa

Pipeline

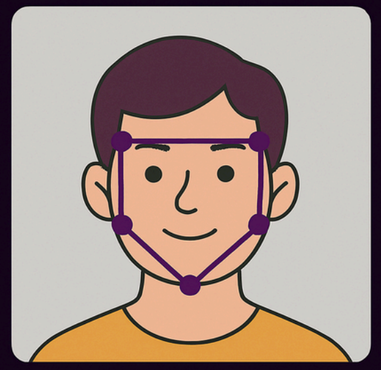
1. Face Detection

Consiste nell'individuazione automatica dei volti, un algoritmo di rilevamento analizza la scena e restituisce le coordinate spaziali che racchiudono ciascun volto individuato. I metodi più moderni si basano su reti neurali convoluzionali.

Face Detection



Face Alignment



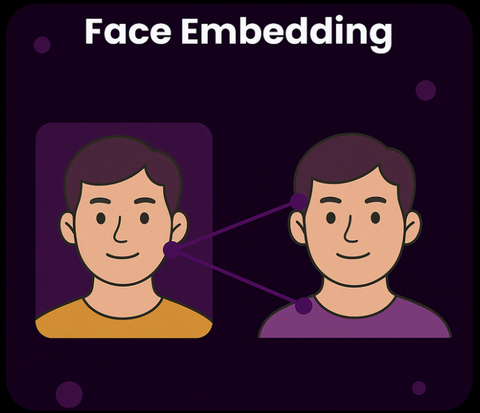
2. Face Alignment

Una volta rilevato, il volto viene normalizzato affinché risulti allineato. L'allineamento si basa sull'individuazione di KP del viso che permettono di correggere eventuali rotazioni, traslazioni o deformazioni prospettiche.

Pipeline

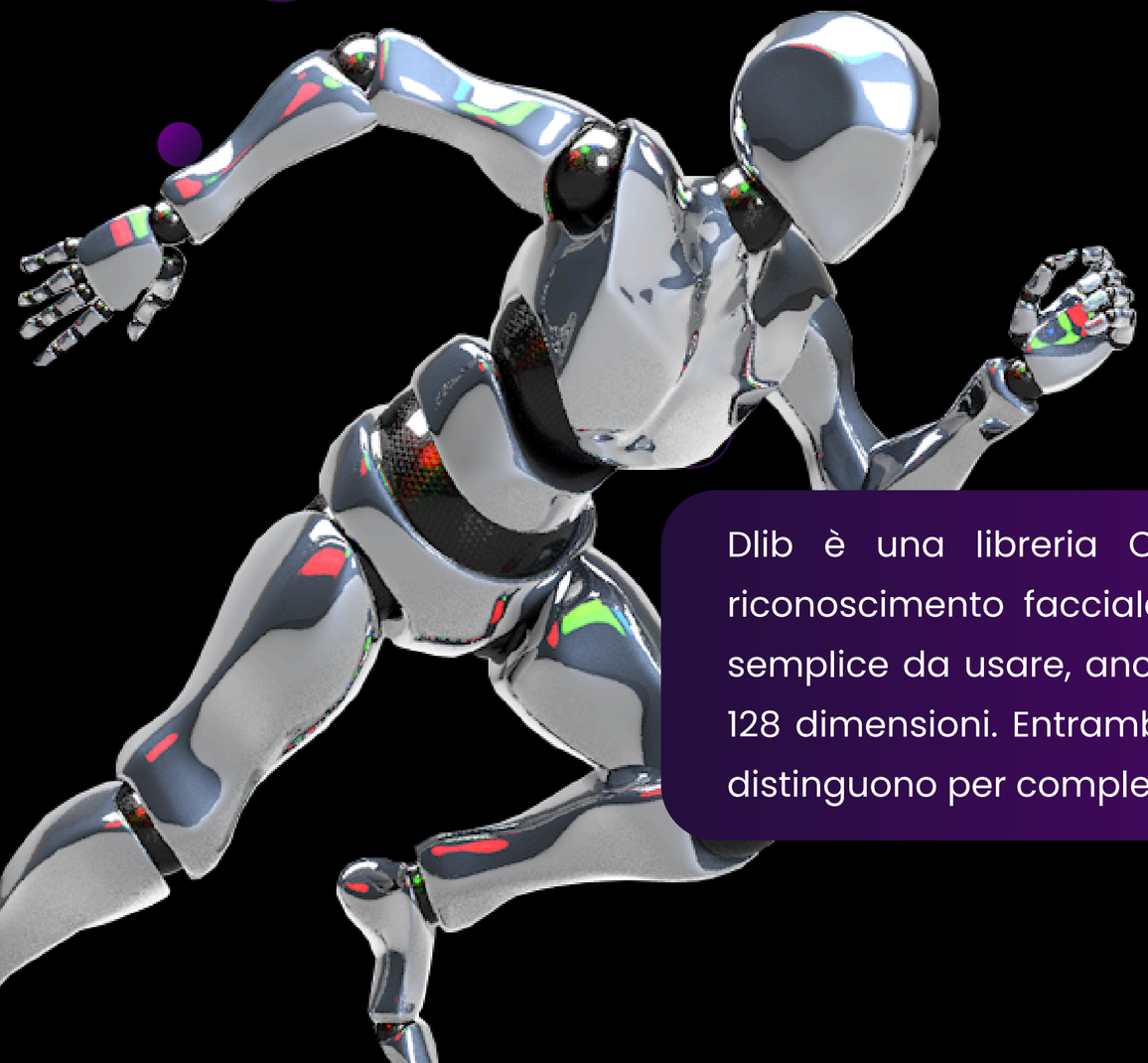
3. Face Embedding

Il volto allineato viene convertito in un vettore numerico, chiamato embedding, che ne rappresenta le caratteristiche più distintive. L'embedding viene calcolato da una rete neurale profonda addestrata per proiettare ogni volto in uno spazio vettoriale dove la distanza tra due punti corrisponde al grado di somiglianza tra i volti.



4. Face Matching

L’embedding estratto viene confrontato con quelli già presenti in un database per identificare il volto o verificarne l’identità. Il confronto avviene calcolando una misura di distanza tra i vettori, come la distanza euclidea o il coseno di similarità. Se la distanza è al di sotto di una certa soglia, si considera che i due volti corrispondano alla stessa persona.



Confronto tra due modelli di embedding

I due principali modelli di embedding sono FaceNet e Dlib, FaceNet è un modello neurale avanzato, preciso e basato su Inception-ResNet e genera embedding da 128 dimensioni. È ideale per applicazioni professionali ed è stato progettato per minimizzare la triplet loss.

Dlib è una libreria C++ con modelli pre-addestrati per il riconoscimento facciale basato su ResNet-34, è più leggero e semplice da usare, anche questo modello genera embedding a 128 dimensioni. Entrambi usano reti neurali convoluzionali, ma si distinguono per complessità, accuratezza e requisiti hardware.

Codice

Preprocessing

```python
# Inizializza alcune variabili
face_locations = []
face_encodings = []
face_names = []

while True:
    # Legge un singolo frame dalla webcam
    ret, frame = video_capture.read()
    if not ret:
        break

    # Ridimensiona il frame a 1/4 per velocizzare l'elaborazione
    small_frame = cv2.resize(frame, (0, 0), fx=0.25, fy=0.25)

    # Converte l'immagine da BGR (OpenCV) a RGB (face_recognition)
    rgb = cv2.cvtColor(small_frame, cv2.COLOR_BGR2RGB)

    # Trova i volti presenti nel frame corrente
    face_locations = face_recognition.face_locations(rgb)
```

Fase di riconoscimento

```python
if face_locations:
    # Calcola gli encoding dei volti rilevati
    face_encodings = face_recognition.face_encodings(rgb, face_locations)

    face_names = []

    for (top, right, bottom, left), face_encoding in zip(face_locations, face_encodings):
        # Verifica se il volto corrisponde a uno di quelli conosciuti
        matches = face_recognition.compare_faces(known_face_encodings, face_encoding)
        name = "Sconosciuto"

        # Calcola la distanza tra il volto rilevato e quelli noti
        face_distances = face_recognition.face_distance(known_face_encodings, face_encoding)
        best_match_index = np.argmin(face_distances)
        confidence = 1 - face_distances[best_match_index]  # Valore tra 0 e 1

        # Se c'è corrispondenza e il livello di confidenza è accettabile
        if matches[best_match_index] and confidence >= 0.1:
            name = f"{known_face_names[best_match_index]} ({int(confidence * 100)}%)"

        # Salva il nome e le coordinate del volto
        face_names.append((name, (top, right, bottom, left)))
```

Esempio risultati ottenuti

Limiti del riconoscimento facciale

Occlusioni

Oggetti come occhiali, mascherine, cappelli o mani possono coprire parti del volto, riducendo l'accuratezza del riconoscimento e aumentando i falsi negativi.

Condizioni di luce

Variazioni di illuminazione influenzano i dati RGB e la qualità delle immagini acquisite. Al buio o in controluce, i sistemi tradizionali possono fallire.

Espressioni facciali

Movimenti del volto (sorrisi, smorfie, emozioni) deformano la geometria facciale, rendendo più difficile il confronto con immagini di riferimento neutre.

Generalizzazione

Difficoltà nel riconoscere volti non presenti nei dati di addestramento, soprattutto in presenza di bias etnici, di età o di genere. I modelli possono non essere equi o robusti.

Bias nei sistemi di riconoscimento

Il bias nei sistemi di riconoscimento facciale si riferisce a errori che portano il modello a prestazioni diseguali tra diversi gruppi demografici. Questo può tradursi in un'accuratezza molto alta per certi gruppi e molto bassa per altri.

La differenza tra quelli umani e algoritmici è che quest’ultimi riflettono i pregiudizi umani nei dati e nei modelli creati. Mentre i bias umani sono correggibili e contestualizzati, quelli algoritmici sono sistematici, replicabili su larga scala e difficili da correggere senza interventi esterni.

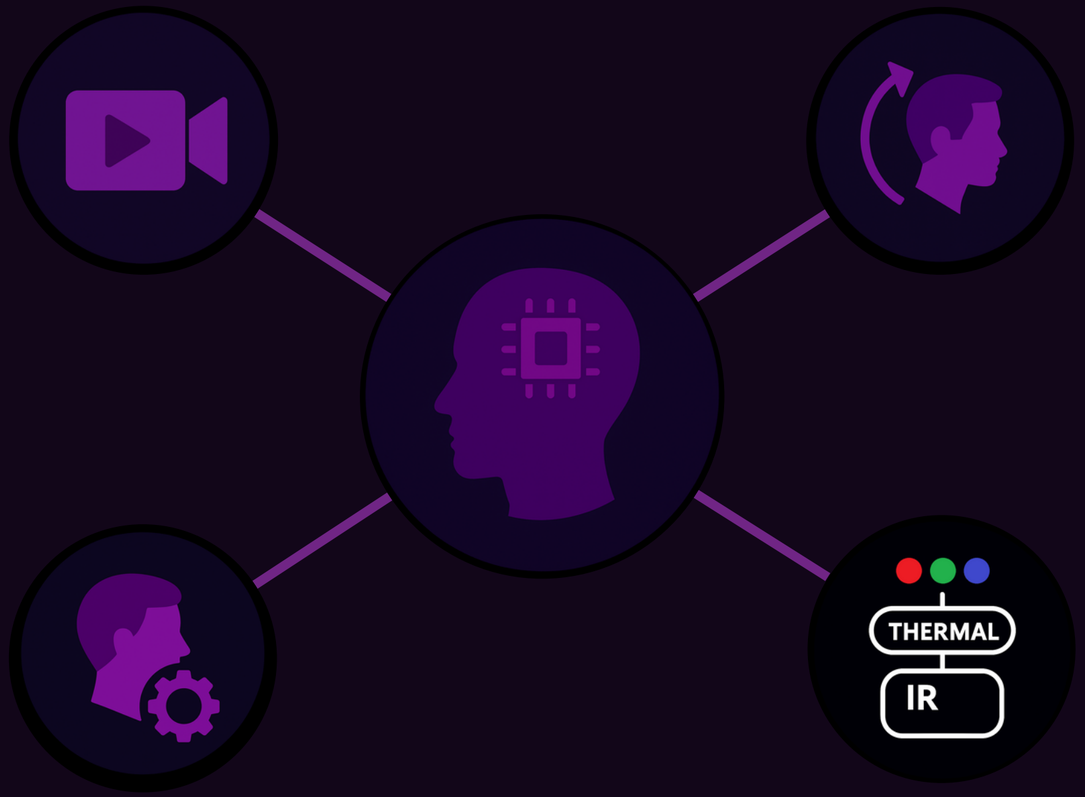
Panoramica dei principali modelli CNN per il riconoscimento facciale

Modello	**Anno**	Architettura	Funzione Loss	Dataset di training	Prestazioni (LFW)
DeepFace	2014	9-layer CNN	Softmax	Sconosciuto (4M img)	~97.35%
VGGFace2	2018	VGG / ResNet	Softmax	VGGFace2 (~3.3M img)	~99.2%
ArcFace	2019	ResNet-100	Additive Angular Margin	MS1MV2 (~5.8M img)	~99.83%
InsightFace	2020	ResNet-100	ArcFace Loss (AAM-Softmax)	MS1MV3, Glint360K	~99.86%

Possibili sviluppi futuri del progetto

Analisi facciale continua su video: tracciamento in tempo reale di identità, movimenti ed espressioni.

Tecniche per allineare il volto in 3D e migliorare il riconoscimento da angolazioni non frontali.



Sistemi in grado di adattarsi all'utente nel tempo, tenendo conto di cambiamenti fisiologici e contesto.

Uso combinato di dati visivi, termici e 3D per maggiore precisione anche in condizioni difficili.