ALMA MATER STUDIORUM – UNIVERSITÀ DI BOLOGNA
CAMPUS DI CESENA

DIPARTIMENTO DI INFORMATICA – SCIENZA E INGEGNERIA
Corso di Laurea in Ingegneria e Scienze Informatiche

VERIFICA DELL'IDENTITÀ TRAMITE LA COMBINAZIONE DI FOTOGRAMMI MULTIPLI E METRICHE DI QUALITÀ DEL VOLTO IN CONTESTI DI CONTROLLO ALLE FRONTIERE

Elaborato in
VISIONE ARTIFICIALE

Relatore
Prof. MATTEO FERRARA

Corelatore
Prof.ssa ANNALISA FRANCO

Presentata da
ELISA YAN

II Sessione di Laurea
Anno Accademico 2024 – 2025

"Was mich nicht umbringt, macht mich stärker."

— *Friedrich Nietzsche*

Indice

Introduzione

Negli ultimi anni, il riconoscimento facciale è diventato una tecnologia di riferimento nei sistemi di sicurezza, in particolare nei controlli automatizzati di frontiera. Tuttavia, la diffusione di attacchi ai sistemi biometrici, che includono tecniche di manipolazione delle immagini e di presentazione di dati falsificati, ha messo in luce vulnerabilità significative. Queste minacce possono ingannare i sistemi di riconoscimento, compromettendone l'affidabilità [16].

La ricerca si concentra sulle nuove strategie di rilevazione e mitigazione degli attacchi, spostando l'attenzione dall'analisi di immagini statiche a quella video-based, più rappresentativa delle reali condizioni operative. L'approccio video-based, infatti, consente di sfruttare la variabilità dei fotogrammi per aumentare la robustezza dei processi di verifica, pur introducendo nuove sfide legate alla qualità delle acquisizioni e all'eterogeneità dei dati.

L'obiettivo è di valutare le prestazioni di due modelli di riconoscimento facciale, MagFace [35] e AdaFace [29], applicati a sequenze video acquisite in condizioni realistiche. L'analisi prende in considerazione non solo le metriche tradizionali di errore, come FRR (False Rejection Rate) e FAR (False Acceptance Rate), ma anche l'impatto di diversi criteri di aggregazione degli score (media, massimo, minimo) e di metriche di qualità dei frame (SER-FIQ [50] e OFIQ [36]). I risultati ottenuti permettono di valutare in che misura tali strategie contribuiscano a migliorare l'affidabilità del riconoscimento in presenza di variazioni di posa, occlusioni e condizioni operative reali, oltre a verificarne l'efficacia nel ridurre la probabilità di successo degli attacchi.

La tesi è strutturata come segue: nel Capitolo 1 viene introdotto il problema degli attacchi ai sistemi biometrici e le principali tecniche di generazione. Il Capitolo 2 descrive le metodologie di rilevamento degli attacchi, con particolare attenzione agli approcci basati su singole immagini e su sequenze video. Nel Capitolo 3 viene descritto il processo di acquisizione del database utilizzato per le valutazioni. Il Capitolo 4 illustra la soluzione proposta, i modelli di riconoscimento adottati e i criteri di aggregazione dei fotogrammi. Infine, il Capitolo 5 riporta e discute i risultati sperimentali, confrontando le diverse strategie di valutazione.

Capitolo 1

Face morphing

1.1 Morphing

Il morphing ha origine nella computer grafica e viene concepito come una tecnica per realizzare una transizione fluida e graduale tra due immagini. A partire dagli anni '90, è stato adottato nell'industria dell'animazione per oltre un decennio per creare effetti speciali nei film, come la deformazione progressiva della forma e del colore di un soggetto fino a farlo scivolare su un altro soggetto, creando un effetto di metamorfosi continua [31]. L'effetto si ottiene utilizzando tecniche semplici come il cross-dissolve, che fa svanire un'immagine mentre l'altra appare. Tuttavia, questa tecnica non rappresenta bene la metamorfosi dell'oggetto, poiché produce artefatti come il "ghosting", zone trasparenti doppie, o forme irreali, come riportato in [2]. La qualità del morphing è stata migliorata attraverso tecniche come warping, per allineare le forme degli oggetti nelle due immagini, e il cross-dissolve che combina i colori e le texture (Figura 1.1). Negli anni, questi metodi si sono evoluti fino ad arrivare a modelli generativi, come autoencoder e GAN, con pipeline quasi del tutto automatizzate grazie alla disponibilità di ampi dataset pubblici e di software open source [14].

1.2 Face Morphing Generation

Quando il morphing interessa un volto si parla di face morphing. Questo può avvenire sia nella combinazione di un volto con elementi diversi (es., paesaggi, oggetti), sia, come nel nostro caso, nella fusione di due o più volti differenti, con l'obiettivo di generare un nuovo volto che preservi le caratteristiche biometriche di ciascun soggetto [16]. Il contributo di ogni volto è regolato da un parametro α, con valori compresi tra 0 e 1. Attualmente, la generazione

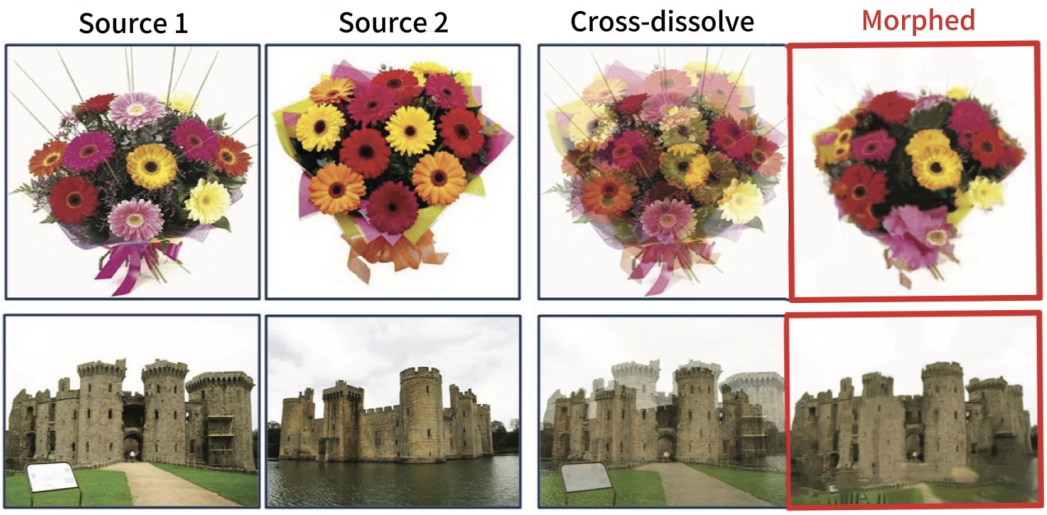

Figura 1.1: Esempio di morphing applicato a immagini generiche. A partire da Source 1 e Source 2, il risultato morphed è ottenuto tramite fusione con cross-dissolve (adattato da [47])

di immagini morphing è diventata un'operazione semplice e a basso costo. Sono infatti disponibili numerose soluzioni open-source, come il plugin GAP per GIMP [24]. Oltre a questi, esistono anche strumenti gratuiti o a pagamento, come FaceMorpher [33] o FantaMorph [1], e applicazioni per dispositivi mobili e servizi online [14].

1.2.1 Tipologie di Face Morphing

Le metodologie di face morphing si suddividono principalmente in due categorie, in base al tipo di approccio utilizzato: landmark based e deep learning based [14]. Di seguito, verranno descritte nel dettaglio entrambe le metodologie, illustrandone i loro processi di funzionamento, i vantaggi e le limitazioni.

Landmark Based Morphing

La prima, nota come *landmark-based* si basa su tre processi principali. Partendo dall'individuazione dei principali landmark facciali quali occhi, naso, bocca e contorno del volto, si vede Figura 1.2. La seconda fase, detta warping, consiste in una deformazione geometrica delle due immagini che consente di allineare in modo coerente tutti i landmark individuati nei due campioni, come illustrato nella Figura 1.2. Infine, nella fase di blending, le intensità dei pixel delle immagini deformate vengono combinate per ottenere l'immagine morphed

Figura 1.2: Esempio di landmarks facciali e i corrispondenti triangoli delaunay (adattato da [14])

finale [44]. Nei paragrafi successivi verranno descritte in modo dettagliato queste fasi, insieme alle tecniche di post elaborazione volte a rimuovere i difetti più evidenti.

Rilevamento e Localizzazione dei Landmark Facciali Questo processo consiste nell'individuare i punti chiave del viso, detti anche landmark o punti caratteristici, a partire da due immagini I_0 e I_1 da unire. Si individuano i rispettivi insiemi di landmark P_0 e P_1, come mostrato in Figura 1.3.

I landmark sono i tratti più evidenti di un volto, come gli angoli degli occhi, la punta del naso, gli angoli della bocca e il contorno del viso [14]. Possono

Figura 1.3: Immagini genitrici I_0 e I_1 (Fonte: [19])

essere suddivisi in due tipologie: i primari, rilevabili anche tramite caratteristiche a basso livello, e i secondari, meno evidenti e la cui individuazione è spesso guidata dalla posizione dei primari [7].

Esistono diversi metodi per il rilevamento dei landmark, tra cui quello manuale, che, se eseguito correttamente, risulta molto accurato ma richiede tempi di esecuzione elevati [44]. Per questo motivo, è stato introdotto il rilevamento automatico, reso possibile da algoritmi come Dlib[1], una libreria open source basata su tecniche di machine learning che utilizza regressori di forma per localizzare automaticamente i punti chiave del volto [41]. L'approccio consiste nel rilevare ogni punto separatamente, utilizzando caratteristiche geometriche [44].

Altri approcci includono gli Active Shape Models (ASM), che utilizzano un modello predefinito adattato al contorno dell'immagine per individuare i punti chiave [34], e gli Elastic Bunch Graph Models (EBGM), che impiegano i Gabor Jets, ovvero insiemi di valori ottenuti applicando dei filtri speciali, come filtri di Gabor, sull'immagine, in diverse direzioni e scale, per evidenziare i dettagli del volto. In questo modo, è possibile identificare con maggiore precisione i punti caratteristici associati ai nodi del grafo [7].

Warping Dopo aver individuato i landmark, dei due volti su cui si vuole applicare il morphing, questi punti vengono sottoposti ad interpolazione (Figura 1.4) per ottenere un nuovo insieme di punti intermedi.

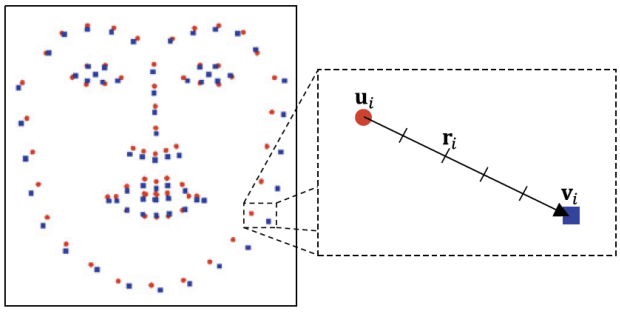

Figura 1.4: Esempio di interpolazione dei landmark tra due insiemi sorgente (P_0) e target (P_1) in funzione del fattore di morphing α (Fonte: [14])

Considerando un singolo landmark, definendo con u_i la sua posizione nell'immagine sorgente e con v_i la posizione del punto corrispondente nell'immagine target, la posizione intermedia si calcola come:

$$r_i = (1 - \alpha) \cdot u_i + \alpha \cdot v_i \tag{1.1}$$

[1]Dlib: `http://dlib.net/`

dove α è il fattore di morphing, che rappresenta la percentuale di presenza del volto sorgente all'interno del volto morphed risultante. L'insieme dei nuovi punti ottenuti, indicato come P_α, verrà utilizzato nelle fasi successive.

Le due immagini vengono deformate geometricamente mediante un processo di triangolazione, nel quale i due insiemi di punti vengono rappresentati con mesh triangolari topologicamente equivalenti, che permettono di allinearli in modo coerente [41]. A questo scopo, viene utilizzata la triangolazione di Delaunay, che consente di massimizzare l'ampiezza degli angoli più piccoli di ogni triangolo, garantendo che i triangoli risultino più regolari e non sovrapposti [34].

Per ogni triangolo in input, viene calcolata una trasformazione affine standard, che mappa i pixel all'interno del triangolo nella corrispondente posizione nella medesima mesh triangolare ottenuta da P_α [32].

Come mostrato in Figura 1.5, un punto p appartenente a un triangolo della mesh dell'immagine target viene riportato al triangolo corrispondente della mesh dell'immagine sorgente tramite una trasformazione affine. In questo modo, ogni pixel viene riallineato in base ai landmark corrispondenti [14].

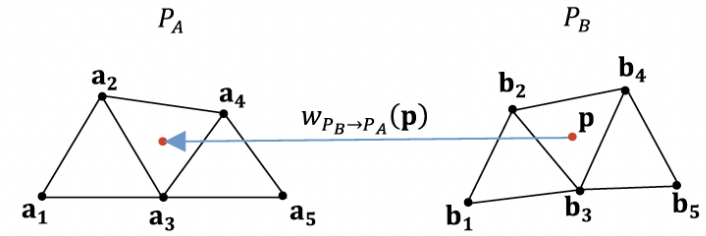

Figura 1.5: Esempio di warping usando il triangular meshes (Fonte: [14])

Formalmente, se un punto p appartiene al triangolo $\Delta b_2 b_3 b_4$ si calcolano le sue coordinate baricentriche $(\lambda_1, \lambda_2, \lambda_3)$ rispetto ai vertici del triangolo. Il punto corrispondente nell'immagine sorgente è dato da:

$$p = \lambda_1 a_2 + \lambda_2 a_3 + \lambda_3 a_4 \tag{1.2}$$

garantendo che la posizione relativa di p all'interno del triangolo venga preservata anche dopo la trasformazione affine [14]. Siano I_s e I_t rispettivamente l'immagine sorgente e l'immagine target, w_{si} e w_{ti} le funzioni di mapping per il triangolo t_i, le immagini risultanti I_{swi} e I_{twi} si ottengono tramite [34]:

- $I_{swi} = w_{si}(I_{si})$
- $I_{twi} = w_{ti}(I_{ti})$

La Figura 1.6 mostra l'immagine risultante dal processo di warping, generata a partire dalle immagini della Figura 1.3.

Figura 1.6: Risultato del warping di un volto

Blending Una volta che le immagini deformate sono state allineate geometricamente, le loro texture vengono fuse per ottenere l'immagine morphed finale. Il metodo più comunemente utilizzato è la fusione lineare: i valori di colore dei pixel vengono calcolati come media ponderata, per poi essere combinati. Anche nella fase di blending è presente un fattore α che regola il contributo delle texture nell'immagine morphed [44].

Quindi il processo di morphing può essere definito come una combinazione di warping e blending, vedi Figura 1.7, ciascuno controllato da un proprio fattore:

$$I_{\alpha_B,\alpha_W}(\mathbf{p}) = (1-\alpha_B)\cdot I_0(w_{P\alpha_W\to P_0}(\mathbf{p})) + \alpha_B \cdot I_1(w_{P\alpha_W\to P_1}(\mathbf{p})) \qquad (1.3)$$

dove:

- p rappresenta un punto nell'immagine morphed
- w rappresenta la funzione di warping

Post-Elaborazione I risultati ottenuti possono presentare artefatti visibili introdotti dal processo di morphing, dovuti principalmente al disallineamento dei landmark o a una densità insufficiente di punti, ma anche all'imprecisione del rilevamento automatico dei punti di riferimento. In particolare, si evidenziano artefatti di tipo ghost, ovvero zone semitrasparenti o simili a ombre, facilmente percepibili a occhio umano, come mostrato in Figura 1.9. Anche in

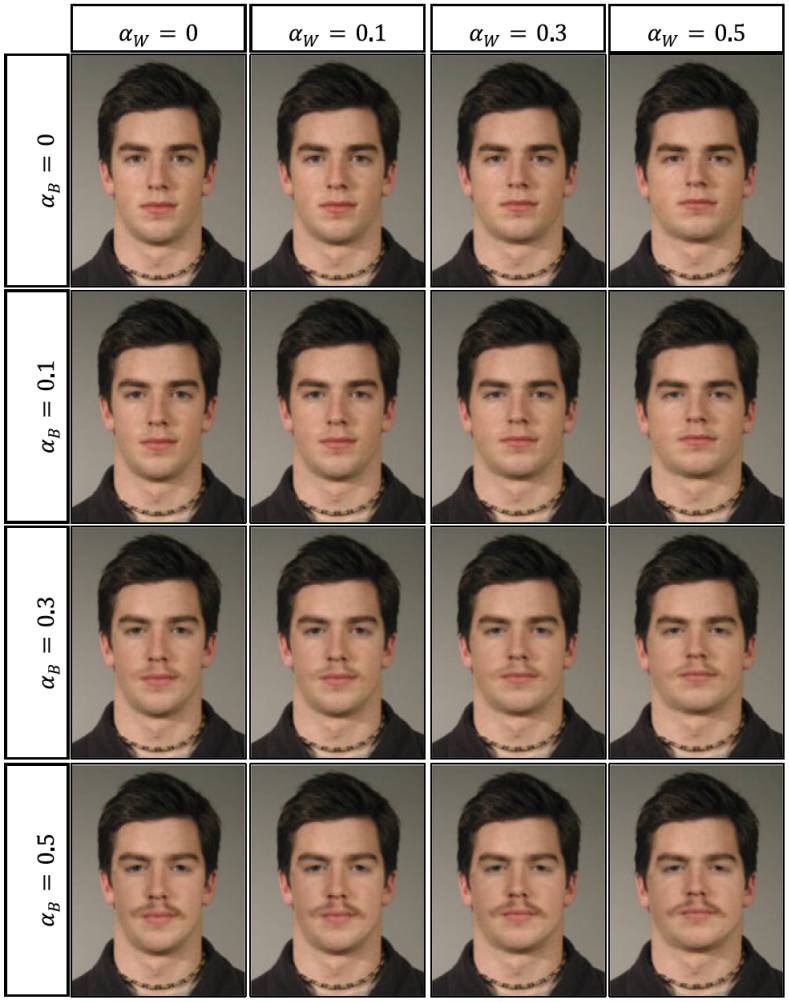

Figura 1.7: Esempio degli effetti combinati del fattore di warping α_W (colonne) e del fattore di blending α_B (righe) sui soggetti della Figura 1.3, con il relativo risultato di face morphing (Fonte: [19])

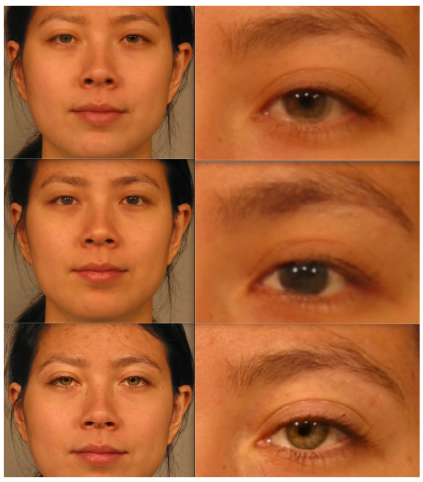

Figura 1.8: Immagine morphed senza rimozione degli artefatti ghost (prima riga, con dettaglio a destra); le righe successive mostrano le versioni ritoccate con diversi metodi (Fonte: [3])

prossimità di occhi, sopracciglia, bocca, naso e mento possono comparire difetti minori dovuti a un numero insufficiente o a un'identificazione non corretta dei landmark, vedi Figura 1.8. Oltre a questi, il processo di morphing può generare gradienti di colore innaturali, bordi netti o interruzioni, che possono derivare da differenze tra le immagini sorgente o dall'uso di metodi di interpolazione non adeguati. Infine, la perdita di contrasto e la sfocatura generale possono emergere come effetto della mediazione dei pixel e dei relativi valori cromatici durante il blending [14].

Per ridurre questi artefatti e rendere l'immagine più realistica, si ricorre a diversi passaggi di elaborazione [14, 44]:

- **sostituzione automatica dello sfondo**: durante il morphing la fusione interessa principalmente la regione del volto, mentre lo sfondo rimane quello originale delle due immagini. Poiché gli sfondi possono differire per colore, luminosità o texture, ciò può generare artefatti visivi evidenti. La sostituzione automatica prevede di utilizzare lo sfondo di una sola immagine, solitamente quella con fattore di morphing α maggiore (più vicina al target), in modo da garantire continuità attorno al volto, come mostrato nella Figura 1.9;

- **histogram matching**: dopo il blending, le differenze nella tonalità della pelle tra le due immagini possono generare zone irregolari. L'histogram matching allinea la distribuzione dei valori di colore (es., intensità, tonalità) tra le due immagini, uniformando l'aspetto della pelle, vedi Figura 1.10;

Figura 1.9: Immagine morphed con artefatti ghost attorno al volto (a); versione corretta con sostituzione automatica dello sfondo (b) (Fonte [14])

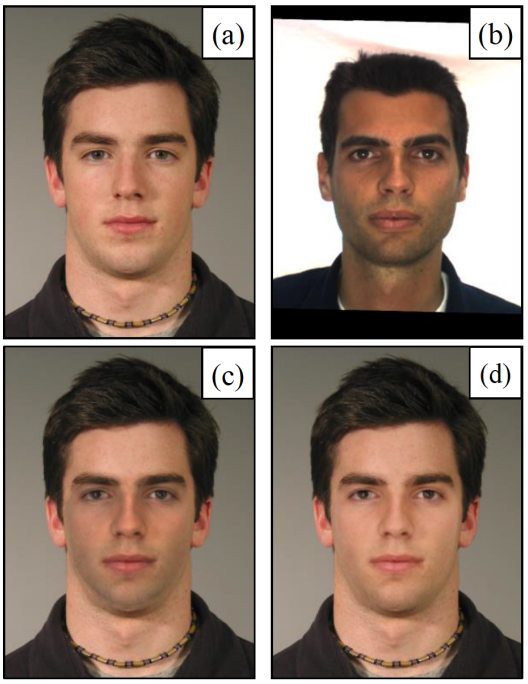

Figura 1.10: Esempio di histogram matching: (a) immagine sorgente, (b) immagine target, (c) immagine morphed con differenze di tonalità evidenti, (d) risultato dopo l'histogram matching con tonalità della pelle uniformate (Fonte: [19])

Figura 1.11: Artefatti di morphing nella regione oculare, come bordi doppi e riflessi multipli nell'iride (a). La stessa area dopo la rimozione manuale degli artefatti (b) (Fonte: [14])

- **smoothing**: la triangolazione può generare piccole discontinuità tra i triangoli. Per correggerle, si applica lo smoothing, un filtro mediano 2×2 sostituisce ogni pixel con il valore mediano dei suoi pixel vicini, ottenendo una transizione più fluida;

- **sharpening**: dopo smoothing e blending, il volto può apparire troppo sfocato, con perdita di dettagli nei contorni (es., bocca, occhi, capelli). Lo sharpening aumenta il contrasto locale ai bordi, restituendo nitidezza e profondità.

- **equalizzazione dell'istogramma**, anche dopo histogram matching, l'immagine morphed può risultare poco contrastata o troppo scura/chiara, soprattutto se le foto originali hanno illuminazioni diverse. L'equalizzazione dell'istogramma ridistribuisce i livelli di intensità, migliorando luminosità e contrasto globale.

Per i difetti impercettibili, non risolvibili automaticamente, il modo più accurato per risolvere è un ritocco manuale con strumenti di image editing, vedi Figura 1.11. Questo tipo di intervento richiede esperienza e tempo, ma riesce ad aumentare considerevolmente le possibilità di successo di un attacco di morphing [14].

Deep Learning Based Morphing

Per superare i limiti dei metodi landmark based, recentemente sono stati proposti metodi in grado di generare un volto morphed basati su tecniche di *deep learning*, in particolare le Generative Adversarial Network (GAN) e Morphing through Identity Prior driven GAN (MIPGAN) [14]. Gli approcci GAN-based sfruttano le reti neurali, una tipologia di modelli di apprendimento

automatico in grado di estrarre le caratteristiche dai dati per sintetizzare le immagini morphed. Questi approcci non richiedono né landmark né allineamento, ma operano nello spazio latente, un tipo di rappresentazione dei dati che cattura le caratteristiche e i modelli più importanti in una forma astratta e strutturata. Ciò consente di generare immagini di alta qualità e risoluzione e di alleggerire il carico di lavoro umano [41]. Gli approcci di morphing basati su GAN [9] riescono a produrre immagini pulite di alta qualità, in grado di ingannare alcuni FRS, ma non permettono di controllare in maniera precisa la somiglianza dell'immagine ottenuta rispetto ai due soggetti originali e potrebbero non ingannare facilmente un esperto umano. Sono comunque necessari ulteriori miglioramenti per poter competere con il landmark based morphing.

Negli studi esistenti, sono stati fatti pochi tentativi di usare le GAN per generare immagini morphed. Il primo è stato MorGAN [10], che produce immagini molto piccole 64×64 pixel, poi migliorate fino a 120×120 pixel. Tuttavia, la qualità rimane bassa e, soprattutto, le immagini ottenute non sono conformi agli standard ICAO, pertanto non sono utilizzabili in scenari pratici come documenti d'identità o controlli di frontiera.

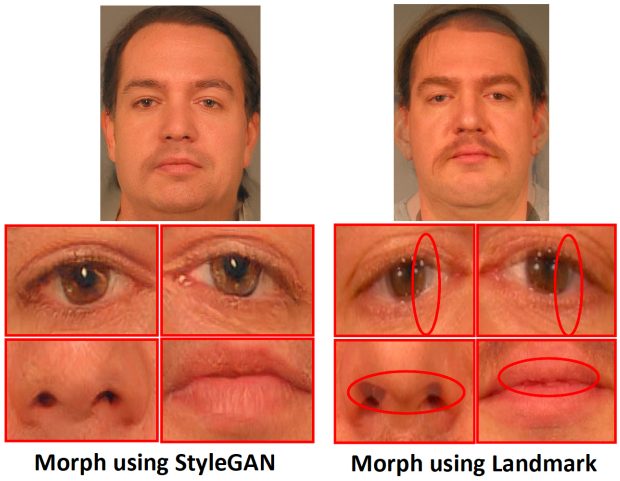

Figura 1.12: Confronto tra morph generati con StyleGAN e morph generati con metodo landmark-based (Fonte: [55])

L'idea è che, per rappresentare una minaccia reale, le immagini morphed devono avere un'alta qualità: devono sembrare realistiche a un esperto umano che controlla un documento e, allo stesso tempo, devono riuscire a ingannare i sistemi di riconoscimento automatici. Per risolvere questo problema di qualità, gli autori di [55] hanno proposto un nuovo approccio basato su StyleGAN, una delle GAN più potenti. Con StyleGAN è possibile generare morph ad alta risoluzione 1024×1024 pixel, molto più realistici e con meno artefatti visivi. Con

questo metodo, i più tipici artefatti intorno agli occhi vengono eliminati automaticamente (Figura 1.12). Quindi, rispetto alle immagini morphed generate con il metodo landmark-based, che spesso presentano artefatti e richiedono un lavoro manuale intenso per essere pulite, i morph ottenuti con StyleGAN hanno una qualità elevata e sono potenzialmente conformi agli standard ICAO.

1.3 Face Morphing Attack

Nell'ultima decade gli attacchi di morphing hanno rappresentato una grave minaccia per i sistemi di Face Recognition System (FRS), in quanto sfruttano la tolleranza del sistema alle variazioni dello stesso soggetto, e possono essere impiegati in scenari critici come la gestione dell'identità, i controlli alle frontiere, e il rilascio di visti, ecc [38]. Questa tipologia di attacco è stata rilevata per la prima volta nel contesto degli Electronic Machine Readable Travel Document (eMRTD) [27], che memorizzano le caratteristiche biometriche per l'identificazione automatizzata, ed è oggi considerata una delle principali minacce alla sicurezza dei sistemi di Automated Border Control (ABC). Infatti, la verifica dell'identità presso i varchi ABC si basa su algoritmi di riconoscimento facciale e sull'ispezione visiva da parte dell'operatore [16].

(a) Subject 1 (b) Morphed (c) Subject 2

Figura 1.13: Esempio di face morphing, dove l'immagine (b) rappresenta il volto morphed ottenuto combinando i soggetti (a) e (c) (Fonte: [4])

Nella Figura 1.13 è riportato un esempio di face morphing tra due soggetti: se l'immagine risultante è sufficientemente simile ai volti dei soggetti originali, allora può essere inserita nell'eMRTD, consentendo a due persone di condividere lo stesso documento. Di conseguenza, negli ABC questa immagine è in grado di ingannare l'esperto umano senza destare sospetti e di ingannare anche il sistema di riconoscimento per la verifica automatica dell'identità. A questo punto, il volto modificato può essere abbinato con successo a entrambi i soggetti [14]. Questo scenario è rappresentato dalla Figura 1.14.

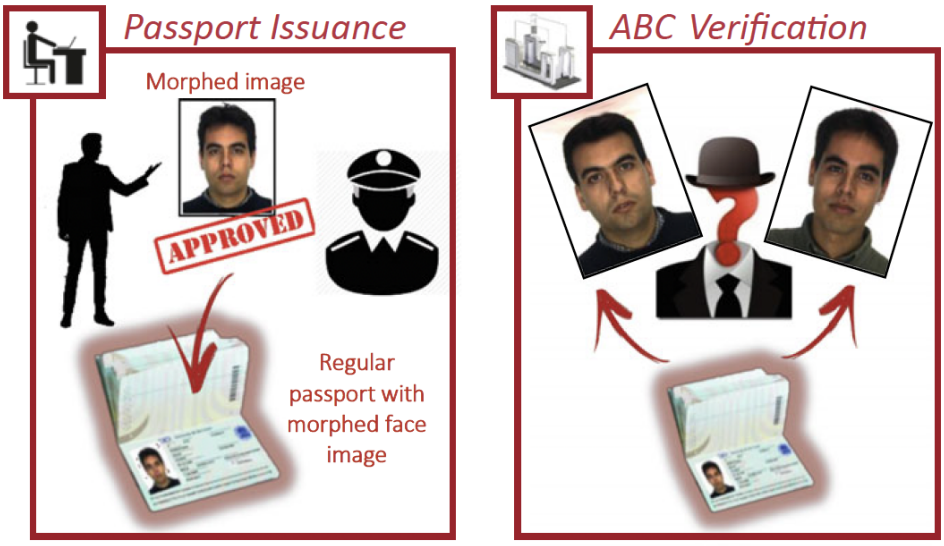

Figura 1.14: Scenario di attacco morphing su eMRTD (Fonte: [16])

Capitolo 2

Face morphing detection

Numerosi studi hanno evidenziato che i sistemi di riconoscimento facciale sono vulnerabili agli attacchi di face morphing, anche in scenari operativi avanzati. Il problema è aggravato dal fatto che negli ultimi anni i documenti d'identità tradizionali sono stati sostituiti da documenti elettronici. Inoltre, le foto contenute nei documenti elettronici devono rispettare rigorosi standard di qualità, come stabilito dalla norma ISO/IEC 19794-5 [28], che segue le linee guida proposte dall'ICAO. Questi standard prevedono che il soggetto abbia una posa neutra, un'illuminazione corretta, un'espressione naturale e che non indossi accessori che coprano il volto, ecc. [21]. Oltre ai requisiti qualitativi, nei varchi aeroportuali (ABC) si seguono le linee guida [23] di Frontex (European Border and Coast Guard Agency) [22], secondo le quali, in tali sistemi operanti in modalità verifica, l'algoritmo di riconoscimento deve garantire un False Accept Rate (FAR) non superiore a 0,1% e un False Rejection Rate (FRR) inferiore al 5% [16].

Gli attacchi di morphing possono essere realizzati in due modalità principali: immagini digitali e immagini stampate e successivamente scansionate (P&S) [38]. La prima modalità prevede che il cittadino possa caricare direttamente una fotografia in formato digitale tramite piattaforme online ufficiali. Questa opzione è adottata in alcuni paesi, come il Regno Unito per il rinnovo del passaporto [25] o la Nuova Zelanda per le richieste di visto [12]. La foto viene quindi inviata come file digitale, tipicamente nel formato JPEG, come si può vedere nella Figura 2.1. Questo processo riduce il tempo e gli errori generati durante la scansione o la stampa, ma presenta comunque dei rischi, in quanto il fatto che il cittadino possa caricare un file digitale può portare a manipolazioni intenzionali. Prima dell'upload, l'attaccante può modificare la foto con software di morphing, basati su landmark o GAN, come descritti nel capitolo precedente. Successivamente, l'immagine morphed viene inviata come se fosse una foto genuina e, non essendoci la fase di stampa e scansione,

Figura 2.1: Esempi di immagini facciali sottoposte a diverse operazioni di data augmentation: (a) originale, (b) stampa e scansione, (c) compressione JPEG, (d) combinazione dei due processi (Fonte: [4])

l'immagine mantiene tutti i dettagli digitali del morph, che possono facilmente ingannare i sistemi di riconoscimento facciale. Invece, in molti altri Paesi, la procedura per l'emissione di documenti elettronici richiede al cittadino di consegnare una fotografia stampata su supporto cartaceo. In questo scenario, un attaccante può prima manipolare digitalmente l'immagine e successivamente stamparla, ottenendo così una foto apparentemente regolare che viene presentata all'autorità competente.

In questi due scenari, oltre alla verifica di conformità agli standard, si potrebbe chiedere un ulteriore controllo per accertarsi che la foto non sia stata alterata [21, 38]. Se l'immagine supera i requisiti, allora verrà scansionata per essere salvata nel chip del documento. Durante la fase di stampa e scansione (P&S) la qualità dell'immagine degrada notevolmente, nascondendo dei piccoli dettagli e artefatti digitali che rendono più difficile l'individuazione dell'immagine morphed sia per il sistema automatizzato sia per il personale addetto, vedi Figura 2.1; di fatto, le immagini P&S restano quelle più difficili da gestire [43, 20].

Questa evidenza ha portato la comunità scientifica a sviluppare metodi di Morphing Attack Detection (MAD) [38] con l'obiettivo di mitigare la vulnerabilità dei sistemi biometrici. Questi sistemi hanno l'obiettivo di distinguere tra immagini genuine e immagini morphed, riducendo il rischio che un documento elettronico contenente una fotografia morphed venga accettato da un FRS o da un operatore umano durante i controlli. Negli ultimi anni sono state proposte numerose soluzioni MAD, che differiscono per approccio, complessità e ambito di applicazione. Tradizionalmente questi approcci si suddividono in due categorie:

- Single-image-Based MAD (S-MAD), si basa sull'analisi di una singo-

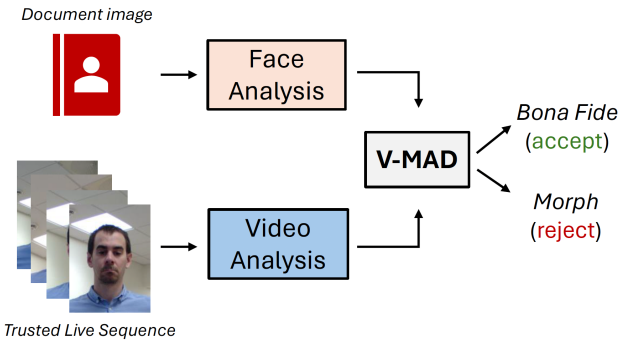

Figura 2.2: Schema del processo di Video-based Morphing Attack Detection (V-MAD): l'immagine del documento e la sequenza video live vengono analizzate e confrontate per classificare l'input (Fonte: [5])

la immagine facciale, tipicamente quella memorizzata nel documento elettronico, senza necessità di confronto con altri campioni;

- Differential-Image-Based MAD (D-MAD), sfrutta il confronto tra l'immagine sospetta e un'immagine fidata, acquisita direttamente in un contesto controllato come un varco ABC.

Tuttavia, nei contesti reali, come i varchi ABC presenti negli aeroporti internazionali, i sistemi FRS commerciali acquisiscono spesso flussi video e non solo immagini singole, perché una sequenza di fotogrammi, acquisita in diverse pose o condizioni di illuminazione, consente di selezionare le immagini più adatte e quindi di verificare con maggiore accuratezza l'identità del soggetto. Allo stesso tempo, però, ciò rappresenta anche un vantaggio per i sistemi MAD che possono sfruttare la sequenza per rendere il rilevamento più robusto, scartando i fotogrammi di bassa qualità dovuti a un'illuminazione non uniforme o a pose non frontali. Per affrontare questa sfida, è stato proposto di estendere il MAD al contesto video, dando origine al Video-based MAD (V-MAD) [5]. L'obiettivo è di sfruttare le sequenze video per aumentare precisione e robustezza, vedi Figura 2.2, adattando gli algoritmi MAD a scenari reali come i controlli aeroportuali. Nelle sezioni successive verrà presentata un'analisi dettagliata di queste tre tipologie.

2.1 Single Image-Based MAD

L'obiettivo dell'S-MAD è di rilevare un attacco di face morphing basandosi su una singola immagine presentata all'algoritmo [52]. Gli algoritmi S-MAD [38] generalmente si basano sul training di un classificatore che distingue

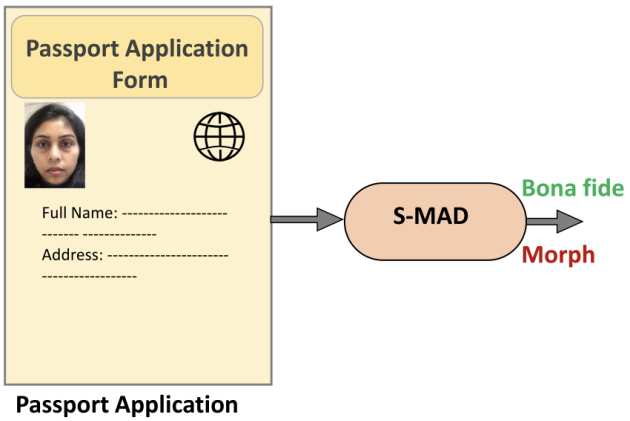

Figura 2.3: Esempio di applicazione del S-MAD nello scenario di domanda per il passaporto (Fonte: [52])

tra immagini autentiche e immagini morphed. Un esempio è riportato in Figura 2.3, relativo alla fase di richiesta del passaporto, in cui la foto consegnata dal cittadino viene sottoposta ad analisi per accertarne l'autenticità e rilevare eventuali morph. Nei primi lavori di ricerca, si utilizzavano i metodi più classici [37] per l'analisi della texture, e le principali feature utilizzate erano:

- Binarized Statistical Image Features (BSIF), è un descrittore di texture. Funziona creando delle maschere statistiche che analizzano piccoli blocchi dell'immagine e trasformano le informazioni in sequenze binarie (0 e 1). Serve per catturare dettagli fini e ricorrenti della pelle o del volto;

- Local Binary Patterns (LBP), confronta ogni pixel con i suoi vicini e segnala con 1 i pixel più chiari e con 0 quelli più scuri. Questo crea una sorta di impronta binaria della regione. È utile per rilevare differenze nei pattern di pelle, rughe, pori, ecc.

I lavori BSIF e LBP riguardavano solo le texture in bianco e nero o in spazi di colore semplici. Tuttavia, un morph può lasciare tracce anche nei colori, come nelle piccole differenze tra i canali RGB, o nei residui nei canali HSV [53]. Inoltre, alcuni artefatti sono più visibili a diverse scale: in una scala fine si possono notare sfocature o difetti nei bordi, mentre in una scala più grande, invece, emergono incoerenze di tonalità o blending [54]. Per questo motivo, alcuni ricercatori hanno combinato l'analisi dei vari spazi di colore e a più scale, per catturare meglio le tracce invisibili del morphing [40].

Inoltre, alcuni Paesi non accettano le immagini digitali, ma richiedono la foto stampata, che poi viene scansionata dall'ufficio passaporti. Questo processo di stampa e scansione degrada l'immagine, pertanto le tecniche basate

su texture classiche, LBP e BSIF diventano meno efficaci. In questo caso, vengono utilizzate reti profonde pre-addestrate, VGG19 [48] e AlexNet [30], spesso impiegate per estrarre feature più robuste e capaci di cogliere tracce più sottili anche dopo la degradazione da P&S , come dimostrato in [39]. Ricavate le caratteristiche di texture, queste vengono date in input a un classificatore, spesso una SVM preaddestrata [43] per riconoscere se l'immagine è genuina o morphed.

2.2 Differential Image-Based MAD

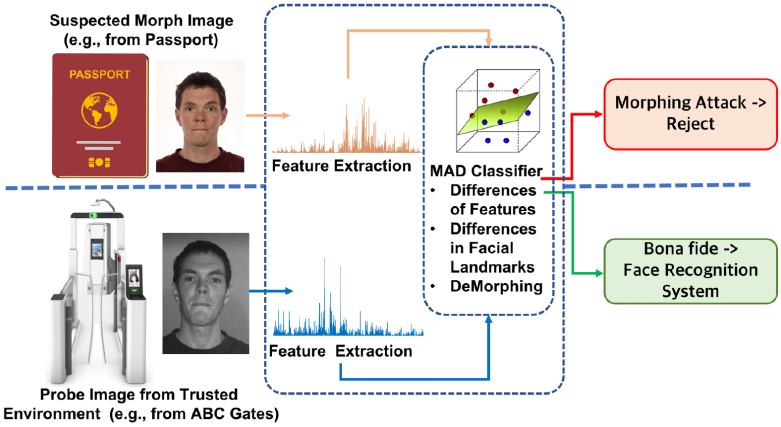

Figura 2.4: Schema di un sistema D-MAD: confronto tra immagine sospetta del documento e immagine acquisita in ambiente controllato per rilevare la presenza di morphing (Fonte: [38])

L'obiettivo del D-MAD è quello di stabilire se un'immagine sospetta è stata manipolata o se è genuina, sfruttando però una seconda immagine di riferimento acquisita in un ambiente affidabile [52]. Nel caso dei controlli di frontiera ABC, l'immagine sospetta (generalmente quella presente nel passaporto) viene confrontata con un'immagine acquisita dal vivo; il processo è illustrato nella Figura 2.4. Il principio generale è il seguente: si prende l'immagine sospetta I_s e la si confronta con un'immagine di riferimento I_t, acquisita in un ambiente fidato. In seguito, si misura la differenza tra le due immagini; questa operazione può essere effettuata direttamente in uno spazio immagine o nello spazio delle feature. Se la differenza ottenuta è bassa, il sistema tende ad accettare l'immagine come autentica; se invece è alta può indicare che l'immagine sospetta è un morph [38]. I metodi D-MAD si possono suddividere in due grandi categorie: la feature difference-based D-MAD [45] e il demorphing [17].

Feature difference Il Feature Difference-Based D-MAD [52] si basa sul confronto delle feature estratte da un'immagine sospetta e da un'immagine di riferimento. Per rilevare eventuali manipolazioni, la differenza viene calcolata nello spazio delle feature, che possono rappresentare diversi aspetti del volto. I primi studi si sono concentrati sulle texture e sui gradienti [8], in quanto le texture descrivono la distribuzione dei pixel e i pattern superficiali della pelle, come rughe, pori e micro-dettagli. I gradienti, invece, rappresentano le variazioni di intensità luminosa tra i pixel adiacenti e mettono in evidenza i bordi e i contrasti. Le immagini morphed, a causa delle operazioni di blending, spesso presentano transizioni innaturali lungo i bordi facciali, che possono essere rilevate tramite questa analisi.

I volti manipolati con morphing spesso possono presentare anche anomalie geometriche rispetto a quelli autentici. Un ulteriore approccio riguarda i landmark facciali [42], perché durante il processo di morphing questi punti possono subire piccoli spostamenti e deformazioni, che risultano individuabili se confrontati con quelli rilevati sull'immagine fidata. Con l'avvento delle reti neurali profonde, sono state introdotte le deep features [45], rappresentazioni ad alto livello estratte da reti neurali profonde. Queste hanno prestazioni migliori nel raccogliere differenze più sottili e complesse rispetto ai metodi classici, rendendo l'analisi più robusta anche in scenari difficili.

Infine, per stabilire se l'immagine è autentica o modificata viene calcolata la differenza tra i vettori di feature. Da notare che la maggior parte dei lavori riguarda casi con immagini digitali. Solo recentemente sono stati studiati anche i casi di immagini P&S, con risultati peggiori rispetto alle immagini digitali [45].

Face Demorphing Il Face Demorphing è un'altra soluzione proposta contro il morphing nei passaporti elettronici [17]. L'idea di base è di pensare che il morph è come una combinazione di due volti:

$$M = A + C$$

dove A rappresenta il complice, e C il criminale. Al momento della verifica si ha l'immagine morphed M e una nuova acquisizione live $\tilde{C}$, con cui si può tentare di riottenere il soggetto complice andando a rimuovere dall'immagine morphed la nuova immagine acquisita:

$$D = M - \tilde{C}$$

Se M è un morph, allora D assomiglierà molto di più al complice, invece se M non è morphed, allora D rimarrà sostanzialmente uguale all'identità legittima. Nella Figura 2.5, il soggetto viene sottoposto al normale sistema di

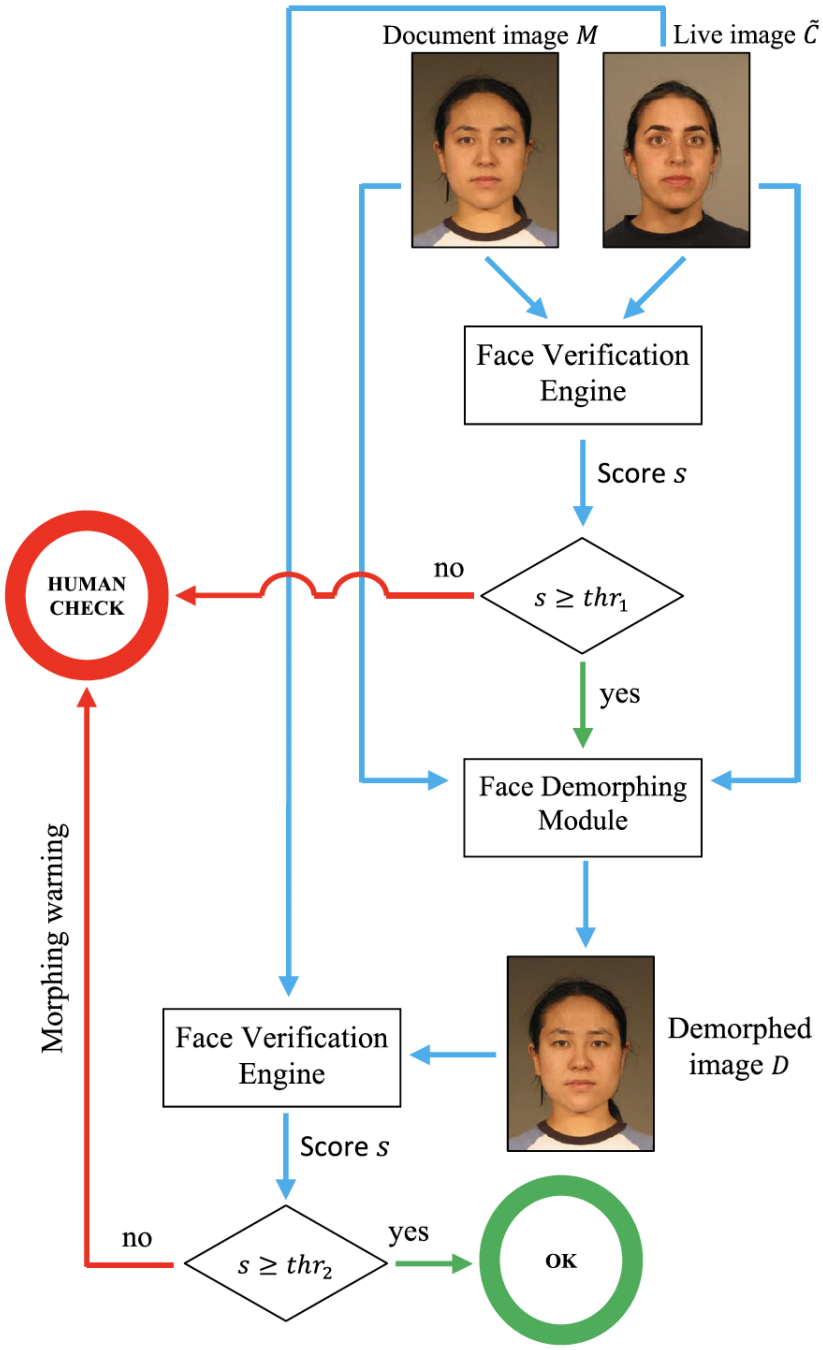

Figura 2.5: Flusso operativo del metodo di Face Demorphing in un sistema ABC (Fonte: [17])

riconoscimento facciale; se supera il controllo, viene eseguito il demorphing. Se invece emerge un mismatch, il sistema genera un avviso e invia il caso a un ufficiale umano per ulteriori controlli. Per funzionare, il sistema deve perciò mantenere basso sia il Morphing Acceptance Rate sia il numero di falsi allarmi.

In teoria, il face demorphing dovrebbe riuscire a rilevare un'immagine morphed per capire chi c'è dietro, ma nella pratica questo processo non è preciso. Questo perché nel gate non abbiamo la stessa foto utilizzata per creare il morph, ma una foto scattata al momento al gate. Inoltre, la persona in questa foto può avere una posa leggermente diversa, una luce differente o un'espressione diversa. Queste piccole differenze rendono più difficile il processo di demorphing, rischiando di generare falsi allarmi [18].

2.3 Video-Based MAD

L'idea alla base dello scenario V-MAD è simile allo scenario D-MAD, in quanto entrambi presuppongono il confronto tra un'immagine sospetta e un altro dato affidabile (una singola immagine nel D-MAD e un video in V-MAD) [49]. Nel caso del D-MAD, la foto memorizzata nel documento elettronico viene confrontata con una singola immagine acquisita live al gate, producendo un punteggio che esprime la probabilità che l'immagine del documento sia un morph. Il V-MAD estende questo concetto allo scenario video: invece di un'unica immagine, il sistema riceve in input una sequenza di fotogrammi catturati durante il passaggio del soggetto al varco.

Quindi l'idea di V-MAD è di analizzare l'intera sequenza di fotogrammi $F = (f_1, f_2, ..., f_n)$ e confrontarla con la foto del documento d, producendo come output un unico punteggio, che indica la probabilità che l'immagine del documento sia un morph. Poiché non esistono ancora metodi V-MAD consolidati, nelle prime soluzioni è stato proposto di adattare i metodi D-MAD al nuovo scenario [5], come rappresentato nella Figura 2.6. Un sistema D-MAD calcola un punteggio $D(d, f_i)$, probabilità che l'immagine del documento d sia morphed, confrontando con un singolo frame f_i. Nel V-MAD si ripete questo processo per tutti i frame della sequenza ottenendo una serie di punteggi:

$$S(d, F) = (D(d, f_i), i = 1, ...n)$$

Attraverso una funzione di aggregazione ϕ possiamo trasformare questi punteggi multipli in un unico risultato finale:

$$V(d, F) = \phi(S(d, F)) \tag{2.1}$$

Si deve tenere conto che i frame acquisiti possono non avere tutti la stessa qualità. In questo caso, se il V-MAD desse lo stesso peso a tutti i frame,

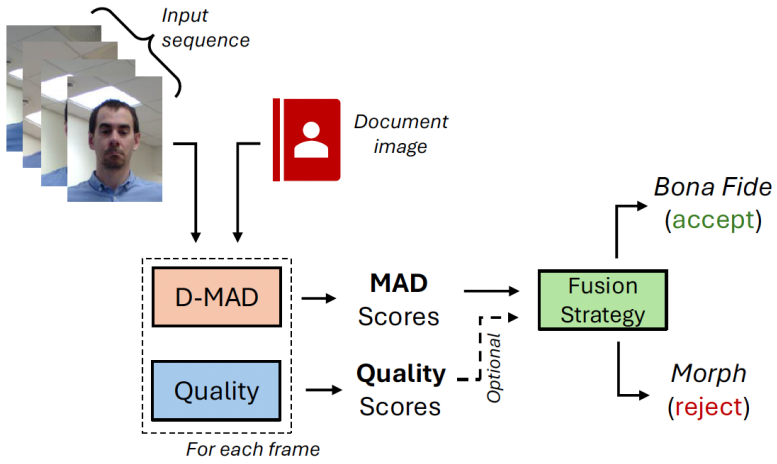

Figura 2.6: (Fonte: [5])

quelli di qualità inferiore rischierebbero di abbassare l'accuratezza del sistema [5]. Per questo motivo, vengono utilizzati gli algoritmi di Face Image Quality Assessment (FIQA) [46], come:

- MagFace [35], produce un embedding per misurare la qualità di una faccia: la misura è data dalla magnitudine dell'embedding, quindi più è grande, più la qualità del volto presente nell'immagine è elevata.
- CR-FIQA [6], lavora sulla posizione della faccia nello spazio delle feature angolari. Pertanto, valuta la posizione del volto rispetto al suo centro e ai centri delle altre identità: più il volto è distinguibile, migliore è la qualità.
- SER-FIQ [50], l'idea è che una foto di alta qualità produce embedding stabili, quindi simili anche se il modello cambia leggermente. Per verificarlo, misura quanto gli embedding restano stabili con variazioni del modello: più sono stabili, migliore è la qualità.
- OFIQ [36], stima la qualità del volto mappando i parametri interni della rete di riconoscimento (es., feature maps, attivazioni, gradienti) in un punteggio. Più il volto contiene caratteristiche discriminative ben rappresentate nei parametri della rete, più alto sarà il punteggio di qualità.

In questo modo, quando si calcola il punteggio finale del V-MAD, i frame di qualità elevata avranno un peso maggiore, mentre quelli con pose o illuminazione sfavorevoli vengono attenuati o scartati.

Capitolo 3

Database

Uno degli aspetti più rilevanti nello studio degli attacchi di morphing e nello sviluppo di contromisure efficaci è la disponibilità di database adeguati e realistici. La maggior parte dei dataset pubblici contiene immagini singole acquisite in condizioni controllate che non rispecchiano pienamente gli scenari reali di controllo alle frontiere. Per questo motivo, uno degli obiettivi della tesi è stato quello di acquisire un dataset[1] contenente, per ciascun soggetto, almeno una foto conforme agli standard International Civil Aviation Organization (ICAO)[2], come quelle usate nei documenti ufficiali quali i passaporti e più sequenze video, raccolte in condizioni differenti.

Le acquisizioni sono state studiate per simulare in maniera realistica lo scenario di un controllo automatizzato alle frontiere: in particolare, riproducono il passaggio di un soggetto davanti a un gate ABC, registrando sequenze video che riflettono la variabilità delle pose, delle condizioni di illuminazione, della direzione dello sguardo e della qualità dei frame acquisiti.

Sono stati coinvolti 65 soggetti, per ciascuno di essi sono state raccolte 5-6 immagini statiche in posa standard, successivamente selezionate tramite strumenti di valutazione della qualità, e 6 sequenze video, contenenti in media 845 frame per sequenza. Le immagini statiche sono state acquisite mediante uno smartphone Nokia Lumia 930[3], con una risoluzione di 3024×5376 pixel, salvate in formati standard PNG o JPG, al fine di garantire fotografie di alta qualità. Le sequenze video sono state registrate con una telecamera Intel RealSense D435i[4], con risoluzione di 1280×720 pixel e un frame rate di 30 fps e salvate sia in formato MP4 sia come sequenze di frame.

[1] Per motivi di privacy, il dataset raccolto non viene reso disponibile né illustrato in questa tesi

[2] ICAO: `https://www.icao.int/publications/doc-series/doc-9303`

[3] Lumia 930: `https://techcrunch.com/2014/04/03/nokia-lumia-930-hands-on/`

[4] Intel: `https://www.intel.com/content/www/us/en/products/sku/190004/intel-realsense-depth-camera-d435i/specifications.html`

3.1 Acquisizione di Immagini ICAO

Durante l'acquisizione è stata prestata particolare attenzione alla qualità visiva e alla coerenza delle inquadrature, così da simulare foto archiviate nei chip degli eMRTD per essere conformi alle specifiche ICAO/ISO [28]. Ogni soggetto è stato posizionato frontalmente alla telecamera davanti a una parete bianca con sfondo uniforme, e per ciascuno sono state acquisite più immagini in diverse condizioni: con e senza flash e variando la distanza della fotocamera. Se il soggetto indossava gli occhiali, l'intero processo è stato ripetuto con e senza di essi, per aumentare la varietà delle condizioni di acquisizione.

Una volta acquisite tutte le immagini, è stata verificata la loro conformità agli standard ICAO, che rappresentano i requisiti internazionali per le foto memorizzate nei passaporti elettronici. Per eseguire questa verifica è stato utilizzato il tool BioLab-ICAO Check Tool [13], sviluppato all'interno del Biometric System Laboratory dell'Università di Bologna.

Il BioLab-ICAO Check Tool valuta 30 requisiti ICAO, tra cui la posizione degli occhi, la presenza di accessori (es., occhiali, cappelli), la qualità dell'illuminazione e la rotazione del volto. I dettagli dei vari criteri per valutare la conformità alle normative ICAO sono riportati nella Tabella 3.1. Questo strumento fornisce una valutazione da 0 a 100 per ogni requisito. Un requisito è considerato superato se supera il punteggio minimo richiesto. Se l'immagine soddisfa tutti i requisiti, viene riconosciuta conforme allo standard ICAO. L'obiettivo dell'uso di questo strumento è quello di selezionare in modo oggettivo le immagini che rispettano i requisiti richiesti dallo standard ICAO, garantendo così un livello adeguato di qualità del dataset per le fasi successive.

Per alcuni soggetti non è stato possibile ottenere un'immagine che soddisfacesse tutti i requisiti previsti. In questi casi, è stata selezionata l'immagine che superava il maggior numero possibile di requisiti. Successivamente, le immagini selezionate tramite il tool sono state rinominate in modo sistematico sulla base di diverse informazioni associate a ciascun file. Ogni immagine è stata etichettata con un nome che codifica:

- L'identificativo del soggetto (001, 002, 003, ...)
- Il genere (GM per maschio, GF per femmina)
- L'etnia del soggetto, come ad esempio EEA per soggetti di origine europea ed americana o EAS per soggetti est-asiatici
- L'età del soggetto, espressa in anni (A25 per 25 anni, A00 per età sconosciuta)
- La presenza o meno di occhiali (T10 se presenti, T00 se assenti)

Tabella 3.1: Test definiti per la valutazione dei sistemi di verifica della conformità allo standard ISO/IEC 19794-5 (Fonte: [15])

N°	**Description of the test**	**EER Threshold**
Feature extraction accuracy tests		
1	Eye Location Accuracy	
2	Face Location Accuracy (other points)	
Geometric tests (Full Frontal Image Format)		
3	Eye Distance (min 90 pixels)	
4	Vertical Position ($0.3B \leq M_y \leq 0.5B$[1])	
5	Horizontal Position ($0.45A \leq M_x \leq 0.55A$)	
6	Head Image Width Ratio ($0.5A \leq CC \leq 0.75A$)	
7	Head Image Height Ratio ($0.6B \leq DD \leq 0.9B$[2])	
Photographic and pose-specific tests		
8	Blurred	4
9	Looking Away	64
10	Ink Marked/Created	99
11	Unnatural Skin Tone	81
12	Too Dark/Light	70
13	Washed Out	56
14	Pixelation	10
15	Hair Across Eyes	75
16	Eyes Closed	100
17	Varied Background	99
18	Roll/Pitch/Yaw Greater than 8°	100
19	Flash Reflection on Skin	77
20	Red Eyes	39
21	Shadows Behind Head	96
22	Shadows Across Face	86
23	Dark Tinted Lenses	28
24	Flash Reflection on Lenses	43
25	Frames too Heavy	33
26	Frame Covering Eyes	35
27	Hat/Cap	62
28	Veil over Face	66
29	Mouth Open	100
30	Presence of Other Faces or Toys too Close to Face	86

[1] $0.3B \leq M_y \leq 0.6B$ for children under the age of 11 years.
[2] $0.5B \leq DD \leq 0.9B$ for children under the age of 11 years.

- La tipologia di acquisizione, come immagini ICAO (LI), o video da gate (LV)
- Il dispositivo utilizzato, ad esempio D05 per Nokia Lumia 930 o D04 per Intel RealSense D435i
- La sessione di acquisizione (S01, S02, ...)
- La fotocamera specifica impiegata, se ve ne erano più di una (C1, C2, ...)
- Il numero progressivo del file all'interno della sessione (I001, I002, ...)

3.2 Acquisizione di Sequenze Video

Per ogni soggetto sono state acquisite sei sequenze video studiate per simulare le condizioni tipiche di un controllo automatizzato ai varchi aeroportuali. Le sequenze sono state acquisite in due ambienti distinti, appositamente scelti per introdurre variabilità in termini di luminosità, sfondo e presenza di ombre. Il primo ambiente è caratterizzato da una combinazione di luce naturale, proveniente da una finestra laterale, e illuminazione artificiale. Il secondo ambiente è illuminato solo artificialmente e con zone circostanti complessivamente più scure. In ciascun ambiente sono state raccolte tre sequenze, per un totale di sei, così da riprodurre scenari con diverse condizioni di illuminazione. Le tre sequenze acquisite in ciascun ambiente hanno le seguenti caratteristiche:

- Sequenza frontale, il soggetto guarda dritto verso la telecamera, mantenendo la posa frontale mentre di muove verso l'obiettivo;
- Sequenza con rotazione del capo, il soggetto ruota leggermente la testa a sinistra o a destra, senza fissare direttamente la telecamera, introducendo variazioni di posa e di direzione dello sguardo;
- Sequenza con accessori, il soggetto indossa un accessorio, come un cappello o una sciarpa e ruota leggermente la testa, combinando occlusioni parziali e variazioni di posa.

3.3 Struttura

Terminata la fase di acquisizione, il dataset è stato organizzato come mostrato nella Figura 3.1. Per ciascun soggetto è stata selezionata la foto ICAO con la migliore qualità. L'intero dataset ha un peso totale di circa 103 GB.

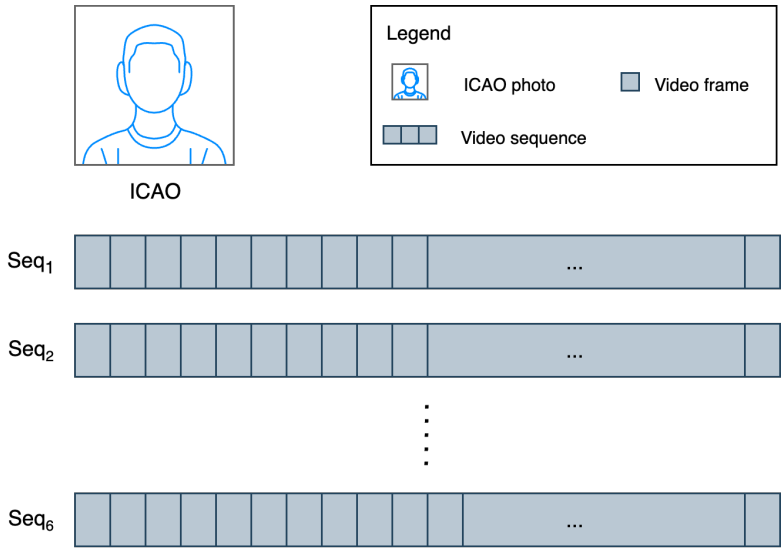

Figura 3.1: Organizzazione dei dati per soggetto: una foto ICAO e sei sequenze video

La struttura delle cartelle del dataset, vedi Figura 3.2, è organizzata in maniera gerarchica. Per ciascun soggetto (es. ID001, ...,ID065) sono disponibili:

- una cartella con le immagini ICAO-compliant;
- una cartella con immagini frontali non ICAO;
- una cartella contenente le sequenze video, suddivise a loro volta in sequenze relative ai diversi ambienti (sequence_01, sequence_02) e, per ciascuna sequenza, nei diversi tipi di posa (frontal_gaze, looking_around, looking_around_with_occlusion).

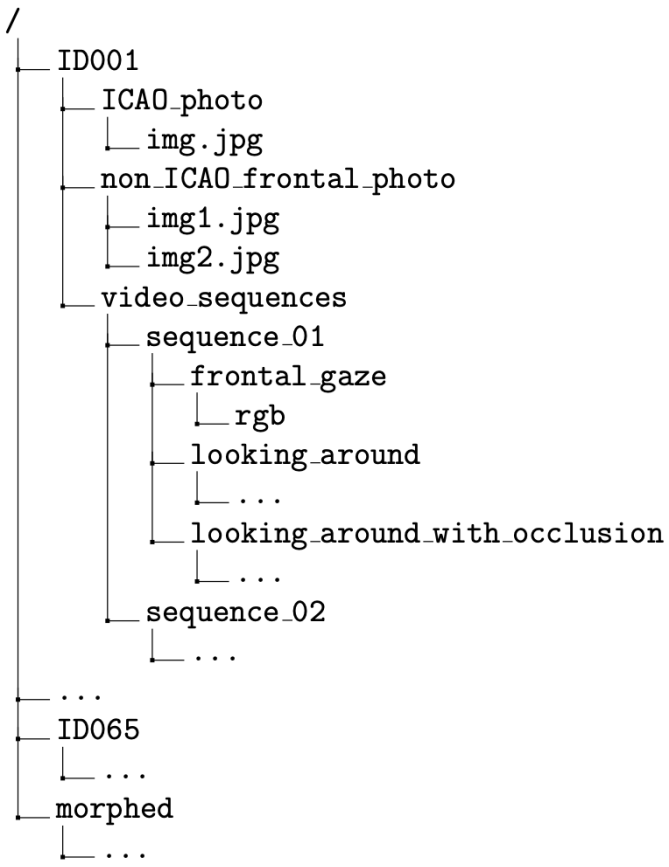

Figura 3.2: Struttura ad albero del dataset

Capitolo 4

Soluzione proposta

Nei sistemi di controllo automatico delle frontiere (ABC), basati su tecniche di riconoscimento facciale, è fondamentale garantire un equilibrio tra sicurezza ed efficienza operativa. Da un lato, il sistema deve essere sufficientemente robusto da evitare che un passeggero legittimo venga respinto ingiustamente, circostanza che si traduce in un aumento del False Rejection Rate (FRR) con conseguenti disagi e ritardi. Dall'altro, è altrettanto importante prevenire che un volto morphed venga erroneamente accettato, fenomeno che incrementa il False Acceptance Rate (FAR) e che rappresenta una minaccia diretta alla sicurezza del sistema. In tale contesto, l'analisi presentata si propone di valutare le prestazioni di due modelli di riconoscimento facciale di ultima generazione, MagFace e AdaFace, applicati a sequenze video acquisite in condizioni realistiche. L'obiettivo è stimarne l'affidabilità sia nei confronti di variabilità naturali (es., posa, illuminazione e qualità visiva), sia in presenza di attacchi di morphing, misurando in modo sistematico i valori di FRR e FAR.

In questo capitolo vengono innanzitutto presentati i modelli e le tecnologie adottate, con la descrizione delle loro caratteristiche e del loro ruolo nel processo di riconoscimento facciale. In seguito, viene illustrata la metodologia sviluppata per affrontare il problema e vengono definite le procedure per la stima delle metriche di prestazione FRR e FAR, calcolate sia a livello di singolo fotogramma sia mediante criteri di aggregazione dell'intera sequenza.

4.1 Modelli e Tecnologie di Riferimento

In questa sezione vengono descritte le tecnologie su cui si basa la soluzione proposta. Si presentano i modelli di riconoscimento facciale MagFace e AdaFace, che introducono meccanismi adattivi per integrare la qualità dell'immagine nella rappresentazione degli embedding, e le metriche di qualità SER-FIQ e OFIQ, utilizzate per selezionare i frame più affidabili. Questi strumenti co-

stituiscono il riferimento metodologico per le procedure di valutazione delle prestazioni illustrate nei capitoli successivi.

4.1.1 MagFace

Uno dei limiti delle precedenti loss function basate sulla similarità coseno, come ArcFace [11], è l'impiego di un margine fisso m. In ArcFace gli embedding dei volti vengono confrontati in similarità coseno, e il margine serve ad aumentare la separazione tra soggetti diversi. Tuttavia, essendo costante e uguale per tutti i confronti, questo approccio risulta efficace in scenari controllati, ma mostra dei limiti in condizioni non controllate, dove la variabilità dovute a pose differenti, illuminazione o qualità delle immagini può rendere instabile la struttura degli embedding dello stesso soggetto.

Per affrontare questo problema, MagFace [35] propone un nuovo schema in cui il modulo del vettore delle feature diventa parte integrante della rappresentazione facciale. Dato un embedding f_i, il suo modulo è definito come:

$$a_i = ||\mathrm{f}_i|| \tag{4.1}$$

A differenza dei metodi tradizionali che normalizzano gli embedding, MagFace mantiene l'informazione contenuta nel modulo, trattandola come un indicatore implicito di qualità. In questo modo, la direzione del vettore rappresenta l'identità del soggetto, mentre il modulo riflette il livello di qualità del campione.

Per garantire queste proprietà, MagFace introduce due funzioni ausiliarie:

- il margine angolare dipendente dal modulo $m(a_i)$, che restringe la regione ammissibile per i campioni di alta qualità, con modulo elevato, concentrandoli attorno al centro di classe w, ovvero il vettore di peso associato all'identità corretta;
- il regolizzatore $g(a_i)$, progettato come funzione convessa decrescente del modulo, rafforza gli embedding di alta qualità e li avvicina al centro w, vedi Figura 4.1.

La MagFace loss estende la ArcFace loss integrando $m(a_i)$ e $g(a_i)$:

$$L_{\mathrm{Mag}} = \frac{1}{N}\sum_{i=1}^{N}\left(-\log \frac{e^{s\cos(\theta_{y_i}+m(a_i))}}{e^{s\cos(\theta_{y_i}+m(a_i))}+\sum_{j\neq y_i} e^{s\cos(\theta_j)}} \cdot \lambda g(a_i)\right) \tag{4.2}$$

dove:

- θ_{y_i} è l'angolo tra l'embedding f_i e il centro della classe y_i,

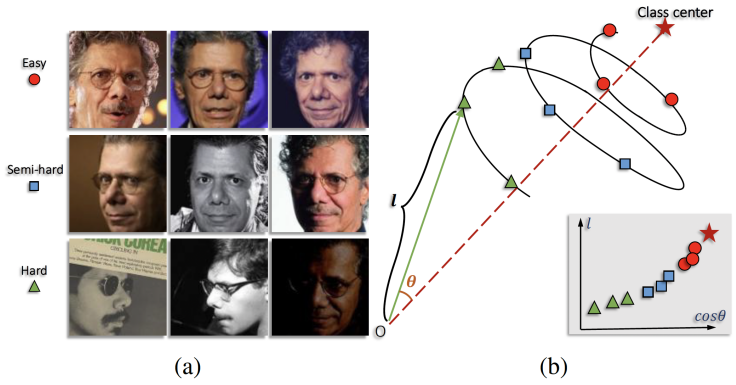

Figura 4.1: Esempi di volti con diversa difficoltà (a) e rappresentazione nello spazio delle feature (b). Nei campioni facili, gli embedding hanno moduli maggiori e si concentrano vicino al centro della classe, mentre quelli difficili hanno moduli ridotti e si avvicinano all'origine (Fonte: [35])

- $m(a_i)$ è il margine adattivo legato al modulo,
- $g(a_i)$ è il regolarizzatore,
- λ bilancia il peso tra classificazione e regolarizzazione.

Questa definizione consente di ottenere la proprietà di convergenza per $a_i \in [l_a, u_a]$, per la quale la funzione di perdita risulta strettamente convessa e ammette una soluzione ottimale unica, garantendo stabilità e rapidità nell'addestramento. L'altra proprietà è quella di monotonicità, che garantisce l'aumento del modulo ottimale a_i^* con la qualità del campione, riflettendone il grado di difficoltà e fungendo da indicatore della qualità del volto.

4.1.2 AdaFace

Spesso, nelle immagini contenenti un volto quando la qualità è troppo bassa, l'identità del soggetto diventa irriconoscibile, e possono creare problemi nel training, perché il modello, non trovando tratti facciali, impara a sfruttare indizi irrilevanti, come il colore dei vestiti, la risoluzione, lo sfondo [51], vedi Figura 4.2. È quindi necessario assegnare un peso diverso ai campioni in base alla loro difficoltà e alla qualità dell'immagine. Nel caso di immagini ad alta qualità è opportuno enfatizzare i campioni difficili (ad esempio pose particolari o leggere occlusioni), mentre nelle immagini di bassa qualità i campioni troppo complessi da riconoscere vanno ignorati, altrimenti il modello rischia di addestrarsi sul rumore [26].

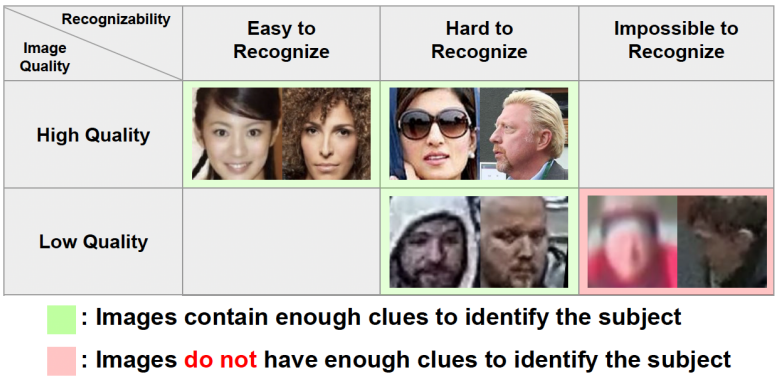

Figura 4.2: Esempi di immagini facciali caratterizzate da diversa qualità e grado di riconoscibilità (Fonte: [29])

È stato dimostrato che la norma del vettore di feature può essere utilizzata come valido proxy della qualità dell'immagine. Partendo da questa osservazione, è stata introdotta una loss function adattiva, AdaFace [29], che modifica dinamicamente il margine angolare in funzione della qualità del campione. In questo modo i campioni difficili vengono enfatizzati soltanto quando la qualità è elevata, mentre quelli troppo difficili, tipici delle immagini a bassa qualità e spesso non riconoscibili, vengono ridimensionati nel loro contributo all'addestramento. L'intero processo avviene senza la necessità di moduli aggiuntivi per stimare la qualità, poiché la qualità viene ricavata implicitamente dalla feature norm.

La progettazione di AdaFace si fonda sulla combinazione di due osservazioni chiave. In primo luogo, la norma del vettore di feature $||z_i||$ può essere utilizzata come indicatore della qualità dell'immagine, se con valori elevati si riferisce a immagini di buona qualità, mentre i valori bassi riflettono campioni degradati o difficili da riconoscere. In secondo luogo, l'impiego di differenti funzioni di margine consente di enfatizzare in maniera mirata campioni di diversa difficoltà.

Normalizzazione della feature norm La norma $||z_i||$ è intrinsecamente dipendente dal modello e può assumere distribuzioni differenti. Per rendere tale misura confrontabile tra campioni e batch, essa viene normalizzata:

$$\widehat{\|z_i\|} = \left\lfloor \frac{\|z_i\| - \mu_z}{\sigma_z / h} \right\rfloor_{-1}^{1} \tag{4.3}$$

dove μ_z e σ_z rappresentano rispettivamente la media e la deviazione standard stabilizzate tramite una exponential moving average (EMA). L'operatore di

clipping limita i valori a $[-1, 1]$, mentre h (tipicamente 0.33) regola la concentrazione. Il gradiente viene bloccato per evitare che la rete manipoli direttamente la norma, garantendo così un indicatore stabile e affidabile della qualità dell'embedding.

Funzione di margine adattivo L'obiettivo della loss adattiva è di enfatizzare i campioni difficili quando l'immagine è di elevata qualità, e di ridurre l'impatto dei campioni troppo difficili quando la qualità è bassa. Per raggiungere questo scopo, la loss introduce due termini adattivi, g_{angle} e g_{add}, che modulano rispettivamente il margine angolare e quello additivo:

$$f(\theta_j, m)\text{AdaFace} = \begin{cases} s\big(\cos(\theta j + g_{\text{angle}}) - g_{\text{add}}\big), & j = y_i \\ s\cos\theta_j, & j \neq y_i \end{cases} \tag{4.4}$$

con

$$g_{\text{angle}} = -m \cdot \widehat{\|z_i\|}, \quad g_{\text{add}} = m \cdot \widehat{\|z_i\|} + m \tag{4.5}$$

La loss di AdaFace è una formula che contiene al suo interno sia ArcFace che CosFace [56] come casi speciali:

- se il valore normalizzato della norma $\widehat{\|z_i\|} = -1$, la loss si comporta esattamente come ArcFace
- se $\widehat{\|z_i\|} = 0$, comporta come CosFace
- se $\widehat{\|z_i\|} = 1$, produce un amrgine diverso, un nuovo caso non coperto dai precedenti modelli.

4.1.3 SER-FIQ

La qualità delle immagini è un fattore cruciale che influenza le prestazioni dei sistemi di riconoscimento facciale. Le metodologie tradizionali per la valutazione della qualità presentano tuttavia diversi limiti: alcune si basano su etichette derivate da punteggi di confronto tra immagini, rendendo la qualità relativa e dipendente da fattori esterni; altre, invece, utilizzano etichette ottenute dalla percezione umana, introducendo inevitabilmente un grado di soggettività. In entrambi i casi, la misura della qualità non risulta pienamente coerente con le esigenze di un sistema automatico di riconoscimento.

Per superare tali limiti, è stato proposto Stochastic Embedding Robustness for Face Image Quality (SER-FIQ) [50], un approccio non supervisionato per valutare la qualità delle immagini facciali, che non richiede etichette di addestramento. L'idea è di misurare la robustezza degli embedding prodotti

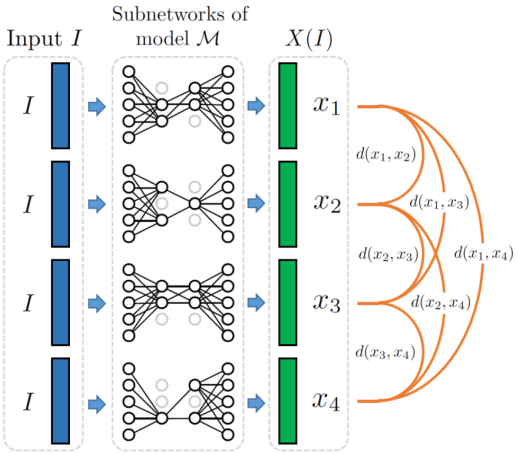

Figura 4.3: Processo del metodo SER-FIQ: l'immagine facciale di input viene elaborata da diverse sottoreti del modello di riconoscimento (generate tramite dropout), producendo più embedding. La qualità dell'immagine è stimata in base alla variabilità tra questi embedding: minore è la variazione, maggiore è la qualità (Fonte: [50])

dal modello: vengono generati m embedding stocastici a partire dalla stessa immagine, sfruttando il meccanismo di dropout, e ottenendo così un insieme:

$$X(I) = \{x_s\}_{s=1}^{m} \tag{4.6}$$

dove m è il numero di forward pass stocastici e x_s sono gli embedding che provengono da una subnetworks casuale del modello $\mathcal{M}$. Se la variazione tra gli embedding è ridotta, l'immagine è considerata di alta qualità; al contrario, una forte instabilità tra i vettori indica un'immagine di bassa qualità (Figura 4.3). La qualità di un'immagine I è definita come:

$$q(X(I)) = \sigma\left(-\frac{2}{m^2}\sum_{i<j} d(x_i, x_j)\right), \tag{4.7}$$

dove:

- $d(x_i, x_j)$ è la distanza euclidea tra due embedding,
- la somma è calcolata su tutte le coppie (i, j),
- $\sigma(\cdot)$ è la funzione sigmoide che normalizza il punteggio in $[0, 1]$.

4.1.4 OFIQ

Negli ultimi anni, grazie all'introduzione di tecniche basate su deep learning, gli errori nel riconoscimento facciale si sono ridotti in maniera significativa. Però, persistono difficoltà legate alle condizioni di acquisizione, al livello della collaborazione del soggetto e della qualità tecnica dell'immagine. Per affrontare queste criticità risulta fondamentale avere algoritmi in grado di stimare la qualità delle immagini facciali, assegnando un Quality Score (QS) a ciascuna immagine. Questo punteggio consente di escludere le immagini che non superano una soglia minima di qualità e, se necessario, di richiedere una nuova acquisizione.

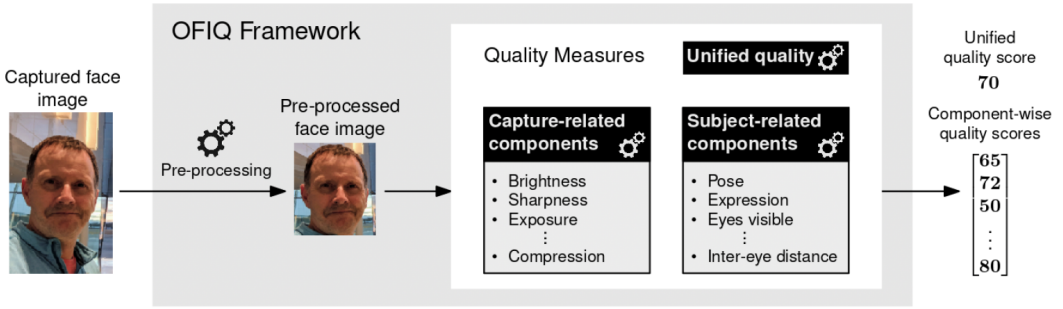

Figura 4.4: Schema del framework OFIQ: dall'immagine facciale acquisita vengono eseguite operazioni di pre-processing e successivamente calcolati l'unified quality score e i vari componenti di qualità, normalizzati nell'intervallo [0–100] (Fonte: [36])

A tale scopo è stato sviluppato Open Source Facial Image Quality (OFIQ) [36] che integra diversi algoritmi, vedi Figura 4.4, tra cui:

- un algoritmo per il calcolo di un punteggio unificato di qualità, che predice l'utilità complessiva dell'immagine ai fini del riconoscimento;

- un insieme di algoritmi per la valutazione di singoli componenti di qualità, che verificano la conformità a requisiti specifici definiti nello standard ISO/IEC 39794-5:2019. Tali componenti possono essere legati all'ambiente di acquisizione (capture-related, come illuminazione o nitidezza) oppure a caratteristiche del volto stesso (subject-related, come posa, occlusioni o espressione), vedi Tabella 4.1.

I valori ottenuti possono essere interpretati e analizzati secondo tre prospettive principali: il punteggio unificato di qualità, i componenti legati al processo di acquisizione e quelli legati al soggetto.

Capture-related Quality Components	**Subject-related Quality Components**
Background uniformity	Single Face Present
Illumination uniformity	Eyes open
Moments of the luminance distribution (Brightness, Variance)	Mouth closed
Over-exposure prevention	Eyes visible
Under-exposure prevention	Mouth occlusion prevention
Dynamic range	Face occlusion prevention
Sharpness	Inter-eye distance
No compression artifacts	Head size
Natural colour	Crop of the face (leftward, rightward, upward, downward)
	Head pose (yaw, pitch, roll)
	Expression neutrality
	No head coverings

Tabella 4.1: Elenco dei componenti di qualità legati all'acquisizione e al soggetto in OFIQ (Fonte: [36])

Unified Quality Score L'Unified quality score può essere ottenuto seguendo due approcci differenti: da un lato è possibile aggregare i punteggi delle singole componenti di qualità, dall'altro si può ricorrere a un algoritmo end-to-end, tipicamente basato su tecniche di deep learning. Dai risultati sperimentali di OFIQ con i due approcci, si osserva che i metodi basati su reti neurali profonde forniscono le prestazioni migliori. Un esempio rilevante è MagFace, che rappresenta un caso concreto di come un algoritmo di deep learning possa fornire direttamente, insieme alle feature per il riconoscimento, anche un indice di qualità complessiva dell'immagine. Nel corso delle valutazioni, MagFace ha dimostrato un impatto significativo sulle prestazioni: scartando il 10% delle immagini di bassa qualità, riducendo il tasso di false non-match dal 10% al 6%.

Capture-related Quality Components Tra i componenti di qualità legati all'acquisizione, quelli con l'impatto maggiore sulle prestazioni del riconoscimento facciale sono *sharpness* e *no compression artifacts*. Eliminando il 10% delle immagini con punteggi peggiori per questi parametri, l'indicatore False Non-Match Error (FNMR) si è ridotto dal 10% all'8%. Altri componenti, come *illumination uniformity*, *natural colour* e *background uniformity* hanno mostrato un'influenza solo marginale.

Subject-related Quality Components Per quanto riguarda i componenti legati al soggetto, molti derivano dalla stima dei landmark facciali o dalla segmentazione del volto, con un'accuratezza superiore al 95%. Tra i fattori più rilevanti si conferma l'effetto delle occlusioni (occhi, bocca e volto), la cui esclusione ha portato a una riduzione dell'errore dal 10% all'8%. Risultati simili si sono osservati per la componente *eyes open*, con una riduzione di circa il 2%. Anche la distanza interpupillare (*inter-eye distance*) influisce in maniera sensibile: valori troppo bassi o troppo elevati peggiorano la qualità della rappresentazione, e l'eliminazione del 10% delle immagini peggiori ha comportato un miglioramento dal 10% all'8%.

Per quanto riguarda il *head pose*, le tre componenti principali mostrano un impatto differente: la rotazione laterale (*yaw*) è la più critica, con riduzioni dell'errore fino al 4%, mentre le variazioni di *pitch* e *roll* influiscono in misura minore (1–2%). Al contrario, componenti come *mouth closed*, *no head coverings* ed *expression neutrality* hanno mostrato un'influenza trascurabile ($\leq 0.5\%$) sulle prestazioni dei sistemi di riconoscimento.

4.2 Verifica dell'identità

Come punto di partenza, l'analisi si è concentrata sulla valutazione delle prestazioni di due modelli di face recognition, MagFace e AdaFace, applicati ai singoli frame del dataset. Ogni frame è stato trattato come un'immagine indipendente, al fine di stimare il comportamento dei modelli in condizioni variabili di posa, illuminazione e qualità visiva.

I modelli sono stati utilizzati per estrarre gli embedding a partire dalle immagini facciali, successivamente impiegati per calcolare la similarità tra la foto ICAO di riferimento di ciascun soggetto e i frame delle relative sequenze video. La misura di similarità adottata è la *cosine similarity*, definita come:

$$\text{sim}(x, y) = \frac{x \cdot y}{\|x\| \, \|y\|} \tag{4.8}$$

dove x e y rappresentano i vettori di embedding confrontati. Il valore ottenuto dalla misura di *cosine similarity* varia nell'intervallo $[-1, 1]$ e rappresenta il grado di somiglianza tra i due vettori di embedding x e y. In particolare:

- un risultato vicino a 1 indica che i vettori sono quasi paralleli, quindi l'immagine di input è altamente simile alla foto ICAO di riferimento;
- un risultato vicino a 0 indica una bassa correlazione, ovvero che i due embedding condividono poche caratteristiche discriminanti;
- valori negativi denotano una forte dissimilarità.

Per garantire un confronto equo tra i due modelli di face recognition, è stata adottata la soglia di riferimento $FAR = 0,1\%$, valore indicato dalle linee guida FRONTEX [23] per i gate ABC utilizzabili negli aeroporti. Tale soglia corrisponde a un tasso massimo di una falsa accettazione ogni 1000 tentativi, consente in un contesto applicativo come quello aeroportuale di assicurare che al più un passeggero su mille venga accettato erroneamente dal sistema.

4.2.1 Analisi a livello di frame

Per ogni sequenza video, l'analisi parte dal confronto a livello di singolo frame. Ogni frame i viene confrontato con la foto ICAO del soggetto mediante l'estrazione degli embedding MagFace e AdaFace, producendo uno score di similarità sim_i, vedi Figura 4.5. La decisione è presa verificando se lo score supera la soglia operativa τ:

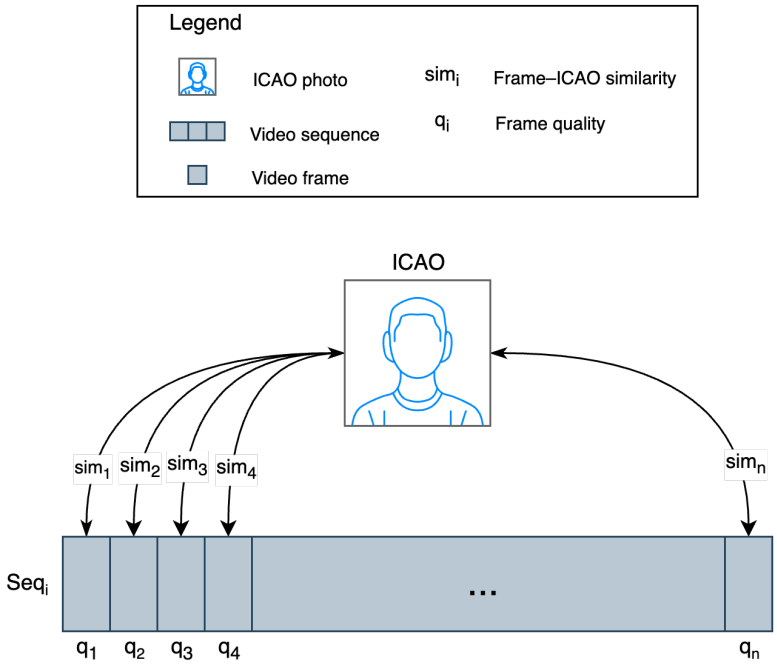

Figura 4.5: Schema del processo di valutazione: ogni frame è confrontato con la foto ICAO tramite embedding (MagFace e AdaFace), producendo le similarità sim_i. Tali valori, aggregati o combinati con le metriche di qualità (SER-FIQ e OFIQ), rappresentano l'intera sequenza e sono confrontati con la soglia operativa.

$$\hat{y}_i = \begin{cases} 1 & \text{se } \text{sim}_i \geq \tau \\ 0 & \text{se } \text{sim}_i < \tau \end{cases} \tag{4.9}$$

dove τ rappresenta la soglia operativa, fissata a

$$\tau = 0.3501 \quad \text{per MagFace}$$

$$\tau = 0.27497 \quad \text{per AdaFace}$$

4.2.2 Criteri di Aggregazione delle Sequenze

Per estendere l'analisi dal singolo frame all'intera sequenza video, sono stati adottati diversi criteri di aggregazione delle similarità calcolate. In particolare, data una sequenza con n frame e i relativi score di similarità sim_i, sono stati considerati i seguenti criteri di aggregazione:

- Valore massimo: selezione dello score di similarità più alto tra tutti i frame della sequenza.
$$s_{\max} = \max_{1 \leq i \leq n} \text{sim}_i$$
- Valore minimo: selezione dello score di similarità più basso.
$$s_{\min} = \min_{1 \leq i \leq n} \text{sim}_i$$
- Media: calcolo della media degli score di similarità di tutti i frame.
$$s_{\text{avg}} = \frac{1}{n} \sum_{i=1}^{n} \text{sim}_i$$
- Metriche di qualità: selezione dei frame con qualità migliore, stimata tramite SER-FIQ e OFIQ. Sono stati considerati i primi 1, 3 e 5 frame di qualità più alta. Su cui è stata poi condotta l'analisi di similarità e, nei casi in cui il numero di frame fosse maggiore di uno, i valori ottenuti sono stati aggregati tramite media aritmetica.

I valori così ottenuti sono stati successivamente confrontati con le soglie operative definite per ciascun modello. In particolare, dato un valore di similarità aggregato s:

$$\hat{y} = \begin{cases} 1 & \text{se } s \geq \tau \\ 0 & \text{se } s < \tau \end{cases} \tag{4.10}$$

In questo modo, una sequenza viene accettata ($\hat{y} = 1$) se il valore aggregato di similarità supera la soglia del modello, altrimenti viene respinta ($\hat{y} = 0$).

Capitolo 5

Risultati sperimentali

In questo capitolo vengono riportati e discussi i risultati sperimentali ottenuti nell'analisi comparativa dei modelli AdaFace e MagFace. L'obiettivo è valutare prestazioni dei sistemi di riconoscimento facciale in scenari realistici, caratterizzati da variabilità nelle condizioni di acquisizione e dalla presenza di attacchi di morphing.

Dopo aver introdotto le metriche di valutazione adottate, viene analizzato i risultati ottenuti a livello di singolo frame e di sequenza video, applicando diversi criteri di aggregazione e strategie basate sulla qualità dei frame. Per entrambe le metriche, i risultati vengono presentati sia globalmente sia per singola sequenza, con l'obiettivo di cogliere l'impatto delle condizioni di acquisizione sulle prestazioni dei modelli.

5.1 Metriche di Valutazione

In questa sezione vengono definiti gli indicatori di performance utilizzati per la valutazione dei modelli, ovvero FRR (False Rejection Rate) e FAR (False Acceptance Rate).

5.1.1 Valutazione del FRR

L'FRR misura la probabilità che il sistema respinga erroneamente un campione genuino. Valori elevati indicano difficoltà di usabilità, poiché soggetti legittimi rischiano di essere rifiutati. Per il calcolo dell'FRR, ogni frame o aggregato di sequenza viene confrontato con la corrispondente immagine ICAO. Se lo score di similarità risulta inferiore alla soglia operativa τ, il sistema respinge il campione. Qualora tale campione appartenga in realtà allo stesso soggetto, il rifiuto viene classificato come false rejection. Formalmente, l'FRR è espresso come:

$$FRR = \frac{\text{falsi rifiuti}}{\text{genuini totali}} = \frac{\sum_{i=1}^{N} \mathbf{1} \cdot \{\text{sim}_i < \tau\}}{N_{\text{genuini}}} \tag{5.1}$$

5.1.2 Valutazione del FAR

Il FAR rappresenta la probabilità che il sistema accetti erroneamente un'immagine morphed come se fosse genuina. In pratica, dato uno score di similarità $\text{sim}(M, \text{f}_i)$ tra un'immagine morphed M e un frame f_i della sequenza video, se tale valore supera la soglia operativa τ, il campione viene accettato dal sistema e, qualora si tratti di un morph, l'accettazione è considerata un falso positivo. Formalmente, il FAR è definito come:

$$FAR = \frac{\text{accettazioni errate}}{\text{attacchi totali}} = \frac{\sum_{i=1}^{N} \mathbf{1} \cdot \{\text{sim}(M_i, \text{f}_i) \geq \tau\}}{N_{\text{attacchi}}} \tag{5.2}$$

5.2 Risultati sul FRR

In questa sezione vengono confrontati i risultati del FRR ottenuti con tutti i criteri analizzati (frame singolo, aggregazioni e metriche di qualità). L'obiettivo è fornire una visione complessiva delle prestazioni di AdaFace e MagFace, mettendo in luce le differenze, i punti di forza e i limiti dei due modelli nelle diverse condizioni di acquisizione.

Per comprendere meglio i risultati, nelle tabelle riportate, la riga GLOBAL indica il valore di riferimento calcolato considerando l'intero insieme dei frame confrontati con le foto ICAO, senza distinzione di sequenza. Le altre righe riportano invece i valori specifici per ciascuna tipologia di sequenza video: *frontal_gaze*, *with_occlusion* e *looking_around*. Ciascuna tipologia è stata acquisita in due sessioni sperimentali distinte: *sequence_01*, con illuminazione naturale e artificiale, e *sequence_02*, con sola illuminazione artificiale, generalmente più critica.

5.2.1 Risultati a livello di frame

Nella Tabella 5.1 sono riportati i valori di FRR calcolati a livello di frame per ciascuna sequenza del dataset, confrontando le prestazioni di AdaFace e MagFace. I risultati evidenziano alcune differenze significative:

- AdaFace mantiene un FRR complessivo molto contenuto (4.3%), con valori particolarmente bassi nelle condizioni di *frontal gaze* (0.66% e 3.47%), che rappresentano lo scenario più favorevole. Anche in presenza di occlusioni o variazioni di posa, pur registrando un incremento dell'FRR,

Sequence	AdaFace (%)	MagFace (%)
GLOBAL	4.30	64.08
sequence_01_frontal_gaze	0.66	34.85
sequence_01_with_occlusion	5.57	77.47
sequence_01_looking_around	3.70	74.49
sequence_02_frontal_gaze	3.47	43.58
sequence_02_with_occlusion	11.23	82.16
sequence_02_looking_around	7.78	79.54

Tabella 5.1: Valori di FRR calcolati a livello di frame per ciascuna sequenza, confrontando AdaFace e MagFace (in percentuale).

le percentuali rimangono sotto il 12%. Quindi, nella maggior parte dei casi AdaFace soddisfa i requisiti indicati dalle linee guida FRONTEX (FRR $\leq 5\%$ con soglia fissata a FAR $= 0.1\%$), dimostrando una buona affidabilità in contesti applicativi controllati.

- MagFace, al contrario, presenta un FRR estremamente elevato (64.1%), con performance insufficienti già in condizioni semplici di *frontal gaze* (34.9% e 43.6%) e valori critici nelle sequenze con occlusioni e movimento, dove l'FRR supera spesso il 75–80%. Tali risultati dimostrano che MagFace non soddisfa mai le specifiche FRONTEX, evidenziando valori sistematicamente troppo elevati.

Si osserva inoltre che la *sequence_02* produce sistematicamente valori di FRR più elevati rispetto alla *sequence_01* per entrambi i modelli. Tale differenza può essere ricondotta alle condizioni di acquisizione: le sequenze del secondo set sono state registrate in presenza di sola illuminazione artificiale, mentre le sequenze del primo set presentano una combinazione di luce naturale e artificiale. La differenza può essere ricondotta principalmente alla qualità dell'acquisizione: le *sequence_01* risultano probabilmente caratterizzate da una migliore messa a fuoco e da un livello di rumore inferiore rispetto alle *sequence_02*.

5.2.2 Risultati con criteri di aggregazione

Media degli score

Nella Tabella 5.2 sono riportati i valori ottenuti applicando come criterio di aggregazione la media degli score di similarità a livello di sequenza video, espressi in percentuale. Si osserva che AdaFace mantiene una notevole robustezza anche in condizioni di acquisizione non ottimali: i valori più elevati si

registrano nelle sequenze con occlusioni e variazioni di posa della *sequence_02*, rispettivamente pari a 1.59% e 0.79%, mentre nelle altre configurazioni, incluse le condizioni *frontal_gaze*, l'FRR risulta pari a zero. Questo conferma che l'utilizzo della media come criterio di aggregazione rende AdaFace un modello affidabile e stabile per applicazioni reali.

Sequence	**AdaFace (%)**	**MagFace (%)**
GLOBAL	0.26	69.79
sequence_01_frontal_gaze	0.00	16.92
sequence_01_with_occlusion	0.00	93.75
sequence_01_looking_around	0.00	91.41
sequence_02_frontal_gaze	0.00	33.85
sequence_02_with_occlusion	1.59	95.24
sequence_02_looking_around	0.79	93.65

Tabella 5.2: Confronto dei valori di FRR ottenuti con aggregazione media degli score, per AdaFace e MagFace (in percentuale)

Al contrario, MagFace dimostra molto più sensibile alle variazioni di posa, alle occlusioni e alle condizioni di illuminazione, registrando un valore globale del 69.79%. È tuttavia interessante notare che, considerando l'aggregazione media, i risultati per le sequenze *frontal_gaze* sono migliorati rispetto all'analisi di livello del frame, Tabella 5.1: in *sequence_01* l'FRR passa dal 34.85% al 16.92%, mentre in *sequence_02* dal 43.58% al 33.85%, con una riduzione di circa dieci punti percentuali. Ciò dimostra che, sebbene il modello resti poco adatto in scenari critici, l'uso della media consente di mitigare parzialmente l'effetto di frame degradati o non rappresentativi.

Valore massimo degli score

I risultati riportati in Tabella 5.3 hanno adottato come criterio di aggregazione il valore massimo degli score di similarità, sia AdaFace sia MagFace ottengono un FRR_max pari a 0% in tutte le sequenze.

Questo risultato è spiegabile con la natura stessa dell'operatore massimo: è infatti sufficiente che almeno un frame della sequenza superi la soglia τ affinché l'intera sequenza venga accettata come genuina. Nel caso di AdaFace, tale comportamento conferma quanto già osservato con il criterio della media, poiché anche in condizioni difficili almeno un frame risulta sufficiente a garantire l'accettazione corretta dei campioni genuini. Per MagFace, invece, l'adozione del massimo tende a mascherare le debolezze emerse con altri criteri di aggregazione: la presenza di un singolo frame positivo consente alla sequen-

Sequence	AdaFace (%)	MagFace (%)
GLOBAL	0.00	0.00
sequence_01_frontal_gaze	0.00	0.00
sequence_01_with_occlusion	0.00	0.00
sequence_01_looking_around	0.00	0.00
sequence_02_frontal_gaze	0.00	0.00
sequence_02_with_occlusion	0.00	0.00
sequence_02_looking_around	0.00	0.00

Tabella 5.3: Confronto dei valori di FRR ottenuti con aggregazione massima degli score di similarità, per AdaFace e MagFace (in percentuale)

za di superare il controllo, eliminando i falsi rifiuti ma riducendo al contempo il potere discriminativo della valutazione.

In termini applicativi, l'impiego del criterio basato sul valore massimo comporta, da un lato, l'azzeramento dei falsi rifiuti, ma dall'altro introduce un aumento potenziale del tasso di false accettazioni. Questo accade perché il sistema, assumendo un approccio eccessivamente permissivo, prende decisione sull'unico frame con similarità più elevata, trascurando la qualità complessiva della sequenza.

Valore minimo degli score

I valori riportati in Tabella 5.4 evidenziano i risultati dell'FRR quando viene adottato come criterio di aggregazione il valore minimo degli score di similarità. Si tratta di un approccio particolarmente restrittivo, poiché considera rappresentativo dell'intera sequenza il frame con la prestazione peggiore.

Sequence	AdaFace (%)	MagFace (%)
GLOBAL	73.44	100.00
sequence_01_frontal_gaze	43.08	100.00
sequence_01_with_occlusion	82.81	100.00
sequence_01_looking_around	69.53	100.00
sequence_02_frontal_gaze	78.46	100.00
sequence_02_with_occlusion	96.83	100.00
sequence_02_looking_around	90.48	100.00

Tabella 5.4: Confronto dei valori di FRR ottenuti con aggregazione minima degli score di similarità, per AdaFace e MagFace (in percentuale)

Per AdaFace, i risultati si attestano su valori complessivamente molto elevati. Nonostante la robustezza del modello, la presenza di singoli frame degradati

all'interno della sequenza compromette l'intero processo di riconoscimento, innalzando il tasso di falsi rifiuti. La situazione è ancora più critica per MagFace, che registra un FRR pari al 100% in tutte le condizioni analizzate. Ciò implica che, indipendentemente dalla sequenza considerata, è sempre presente almeno un frame al di sotto della soglia di accettazione, con il conseguente rifiuto totale di tutti i campioni genuini.

Criteri basati sulla qualità

Per valutare in che misura la qualità delle immagini influisca sulle prestazioni dei modelli, sono stati applicati criteri basati su SER-FIQ e OFIQ, selezionando i migliori frame (Q1, Q3, Q5) di ciascuna sequenza per analizzarne l'effetto sull'FRR.

SER-FIQ Dai risultati basati su SER-FIQ emerge un quadro netto.

Per AdaFace, vedi Tabella 5.5, l'FRR è sempre pari a 0% in tutte le sequenze e per tutte le configurazioni Q1, Q3 e Q5: la selezione dei frame di qualità più alta è sufficiente, già con Q1, a garantire l'accettazione dei campioni genuini, e l'aumento a Q3 e Q5 non porta benefici ulteriori, il che è indice di stabilità dei frame e di embedding robusti anche in condizioni non ideali.

Sequence	**Q1 (%)**	**Q3 (%)**	**Q5 (%)**
sequence_01_frontal_gaze	0.00	0.00	0.00
sequence_01_with_occlusion	0.00	0.00	0.00
sequence_01_looking_around	0.00	0.00	0.00
sequence_02_frontal_gaze	0.00	0.00	0.00
sequence_02_with_occlusion	0.00	0.00	0.00
sequence_02_looking_around	0.00	0.00	0.00

Tabella 5.5: Valori di FRR con criteri SER-FIQ (AdaFace), selezionando i primi 1, 3 e 5 frame di qualità.

Per MagFace, vedi Tabella 5.6, si osserva un miglioramento significativo rispetto ai risultati precedenti: nelle condizioni più semplici (*frontal_gaze*), l'FRR scende fino allo 0% con Q3 e Q5, mentre negli scenari più critici (*with_occlusion, looking_around*) rimane comunque non trascurabile ($\approx$ 14–25%), pur riducendosi con l'aumentare di Q. Anche con l'impiego di SER-FIQ, MagFace continua a mostrare criticità nelle condizioni più complesse.

Sequence	Q1 (%)	Q3 (%)	Q5 (%)
sequence_01_frontal_gaze	3.08	0.00	0.00
sequence_01_with_occlusion	21.88	18.75	20.31
sequence_01_looking_around	7.81	4.69	7.81
sequence_02_frontal_gaze	7.69	4.62	4.62
sequence_02_with_occlusion	19.05	17.46	14.29
sequence_02_looking_around	25.40	22.22	17.46

Tabella 5.6: Valori di FRR con criteri SER-FIQ (MagFace), selezionando i primi 1, 3 e 5 frame di qualità.

OFIQ L'impiego della metrica OFIQ consente di valutare l'impatto della qualità delle immagini sulle prestazioni dei modelli di riconoscimento. Oltre al punteggio unificato, *UnifiedQualityScore*, sono stati considerati anche alcuni parametri specifici, ritenuti tra i più influenti sul riconoscimento facciale, come *Sharpness* e *HeadPoseYaw*. L'obiettivo è verificare se la selezione dei frame in base alla qualità, e in particolare dei migliori 1, 3 o 5 frame di ciascuna sequenza, possa ridurre FRR e migliorare l'affidabilità complessiva dei sistemi analizzati.

- **Unified Quality Score:** I risultati riportati nelle Tabelle 5.7 e 5.8 mostrano i risultati dell'FRR quando si applica un filtro basato su OFIQ, selezionando rispettivamente i migliori 1, 3 e 5 frame di ciascuna sequenza video.

Sequence	Q1 (%)	Q3 (%)	Q5 (%)
sequence_01_frontal_gaze	0.00	0.00	0.00
sequence_01_with_occlusion	0.00	0.00	0.00
sequence_01_looking_around	1.92	1.72	1.64
sequence_02_frontal_gaze	0.00	0.00	0.00
sequence_02_with_occlusion	2.50	3.92	1.92
sequence_02_looking_around	0.00	0.00	0.00

Tabella 5.7: Valori di FRR (%) con AdaFace applicando il criterio *UnifiedQualityScore*

Per AdaFace, i valori risultano prossimi a zero nella maggior parte delle condizioni. Solo in scenari più complessi, come le sequenze *looking_around* e *with_occlusion* della *sequence_02*, si registrano FRR lievemente superiori (fino a circa il 3.9%), che restano comunque contenuti e trascurabili in prospettiva applicativa.

Sequence	Q1 (%)	Q3 (%)	Q5 (%)
sequence_01_frontal_gaze	4.62	7.69	7.69
sequence_01_with_occlusion	32.81	35.94	37.50
sequence_01_looking_around	32.81	29.69	31.25
sequence_02_frontal_gaze	10.77	10.77	9.23
sequence_02_with_occlusion	50.82	54.84	57.14
sequence_02_looking_around	44.44	46.03	49.21

Tabella 5.8: Valori di FRR (%) con MagFace applicando il criterio *Unified-QualityScore*

La situazione di MagFace è più complessa: pur mantenendo valori di FRR elevati nelle condizioni più difficili, si osservano miglioramenti rispetto ai risultati senza filtro di qualità. Nelle sequenze più semplici, come *frontal_gaze*, l'FRR scende a circa 9–10%, mentre anche in presenza di occlusioni o variazioni di posa, l'uso di OFIQ consente comunque una riduzione parziale degli errori.

- **Sharpness:** I risultati ottenuti con il criterio *Sharpness* (Tabelle 5.9 e 5.10) evidenziano differenze significative tra i due modelli.

Sequence	Q1 (%)	Q3 (%)	Q5 (%)
sequence_01_frontal_gaze	0.00	0.00	0.00
sequence_01_with_occlusion	0.00	0.00	0.00
sequence_01_looking_around	0.00	1.75	1.69
sequence_02_frontal_gaze	1.64	0.00	0.00
sequence_02_with_occlusion	0.00	4.88	2.17
sequence_02_looking_around	3.92	5.26	3.45

Tabella 5.9: Valori di FRR (%) con AdaFace considerando il criterio *Sharpness*

Sequence	Q1 (%)	Q3 (%)	Q5 (%)
sequence_01_frontal_gaze	6.15	6.15	7.69
sequence_01_with_occlusion	46.88	48.44	51.56
sequence_01_looking_around	43.75	42.19	43.75
sequence_02_frontal_gaze	13.85	12.31	12.31
sequence_02_with_occlusion	62.30	66.13	66.67
sequence_02_looking_around	58.06	63.49	58.73

Tabella 5.10: Valori di FRR (%) con MagFace considerando il criterio *Sharpness*

AdaFace mostra valori di FRR molto bassi, con variazioni quasi nulle all'aumentare dei frame selezionati (Q3, Q5). MagFace, invece, registra valori sensibilmente più elevati, soprattutto in presenza di occlusioni e variazioni di posa, con FRR che superano il 40–60%. In alcuni casi, l'inclusione di più frame comporta addirittura un incremento degli errori, segnalando una certa instabilità del modello.

- **Head Pose Yaw:** Con il criterio *HeadPoseYaw* (Tabelle 5.11 e 5.12) si osserva un andamento analogo. AdaFace mantiene i valori di FRR vicini allo zero nella maggior parte delle condizioni, con lievi incrementi nelle sequenze più critiche. Per MagFace, invece, i valori rimangono elevati e, ancora una volta, l'effetto della selezione multipla dei frame non è uniforme: a volte riduce l'errore, ma in altri casi lo aumenta.

Sequence	**Q1 (%)**	**Q3 (%)**	**Q5 (%)**
sequence_01_frontal_gaze	0.00	0.00	0.00
sequence_01_with_occlusion	0.00	0.00	0.00
sequence_01_looking_around	0.00	1.61	1.61
sequence_02_frontal_gaze	1.69	0.00	0.00
sequence_02_with_occlusion	0.00	2.00	0.00
sequence_02_looking_around	1.89	3.51	3.39

Tabella 5.11: Valori di FRR (%) con AdaFace considerando il criterio *HeadPoseYaw*

Sequence	**Q1 (%)**	**Q3 (%)**	**Q5 (%)**
sequence_01_frontal_gaze	3.13	6.15	7.69
sequence_01_with_occlusion	26.56	25.00	28.13
sequence_01_looking_around	25.40	21.88	21.88
sequence_02_frontal_gaze	6.25	7.69	9.23
sequence_02_with_occlusion	42.62	41.94	44.44
sequence_02_looking_around	30.16	34.92	36.51

Tabella 5.12: Valori di FRR (%) con MagFace considerando il criterio *HeadPoseYaw*

In generale, i dati suggeriscono che le metriche di qualità possano ridurre parzialmente gli errori, ma non sono sufficienti a compensare le difficoltà di MagFace in condizioni non controllate.

5.2.3 Confronto complessivo sul FRR

Dalla Figura 5.1 è possibile osservare l'andamento dell'FRR per ciascuna sequenza video, confrontando i diversi criteri di aggregazione adottati.

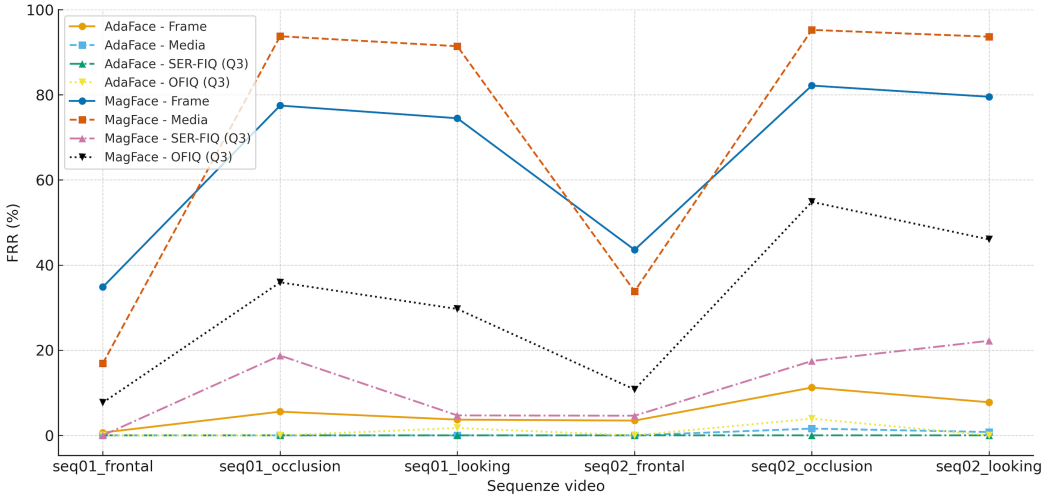

Figura 5.1: Confronto dei valori di FRR (%) per AdaFace e MagFace su diverse sequenze video, utilizzando differenti criteri di aggregazione e filtri di qualità (Frame, Media, SER-FIQ Q3, OFIQ Q3)

Per quanto riguarda AdaFace, i risultati mostrano valori di FRR prossimi allo zero in quasi tutte le condizioni, indipendentemente dal tipo di sequenza. In scenari semplici, come le sequenze *frontal*, sia *sequence_01* sia *sequence_02*, la combinazione di criteri di aggregazione basati sulla media o sulle metriche di qualità (SER-FIQ e OFIQ, entrambi con Q3) permette di mantenere un tasso di falsi rifiuti nullo o trascurabile. Anche nelle condizioni più critiche, caratterizzate da occlusioni parziali o variazioni di posa, l'applicazione dei criteri di qualità (SER-FIQ o OFIQ) garantisce un FRR quasi nullo, confermando la stabilità del modello.

La situazione di MagFace appare più complessa: l'utilizzo diretto dei frame o della media degli score comporta valori molto elevati di FRR, spesso superiori al 70–90%, anche in condizioni favorevoli. Un miglioramento significativo si osserva adottando il criterio SER-FIQ (Q3), che riduce l'FRR fino a valori compresi tra lo 0% e il 22%, a seconda della sequenza. Anche OFIQ (Q3) produce benefici, soprattutto nelle sequenze frontali, dove l'FRR si riduce a circa il 9–11%, ma nelle sequenze in presenza di occlusioni o pose variabili i valori rimangono alti ($\approx$ 30–50%).

5.3 Risultati sul FAR

In questa sezione vengono presentati i risultati relativi al FAR ottenuti nei diversi scenari sperimentali per verificare la robustezza dei due modelli rispetto agli attacchi di face morphing. L'analisi considera sia i valori calcolati a livello di singolo frame, sia quelli derivanti dall'applicazione di criteri di aggregazione e metriche di qualità. L'obiettivo è fornire una panoramica completa delle prestazioni dei modelli AdaFace e MagFace, mettendone in luce le differenze, i punti di forza e le criticità.

5.3.1 Risultati a livello di frame

Dalla Tabella 5.13 si osserva che, a livello globale, MagFace presenta un FAR più contenuto (13.32%) rispetto ad AdaFace (45.92%), evidenziando una maggiore capacità di limitare le false accettazioni sull'intero dataset. Anche in condizioni di acquisizione favorevoli, come nel caso *frontal_gaze*, AdaFace tende comunque a generare un numero elevato di accettazioni errate. Nelle sequenze più complesse, caratterizzate da movimenti del volto, variazioni di posa e presenza di occlusioni, MagFace conferma la propria robustezza, mantenendo valori di FAR ridotti, a differenza di AdaFace che mostra una significativa perdita di affidabilità. Va tuttavia sottolineato che, per entrambi i modelli, i valori restano di gran lunga superiori alla soglia dello 0.1% indicata dalle linee guida FRONTEX, risultando quindi non compatibili con i requisiti operativi dei sistemi aeroportuali.

Sequence	**AdaFace (%)**	**MagFace (%)**
GLOBAL	45.92	13.32
sequence_01_frontal_gaze	57.85	28.58
sequence_01_with_occlusion	45.61	7.21
sequence_01_looking_around	47.38	8.73
sequence_02_frontal_gaze	44.36	21.72
sequence_02_with_occlusion	37.86	4.79
sequence_02_looking_around	37.91	5.63

Tabella 5.13: Valori di FAR (%) calcolati a livello di sequenza per AdaFace e MagFace senza filtri.

5.3.2 Risultati con criteri di aggregazione

Media degli score

Dalla Tabella 5.14 emerge che l'adozione del criterio della media mette in evidenza la differenza tra i due modelli. A livello globale, MagFace riduce il FAR al 3.65%, rispetto al 40.36% di AdaFace. Anche nelle sequenze più critiche, AdaFace mantiene valori elevati e poco stabili, mentre MagFace in diversi casi azzera completamente le false accettazioni, confermando una maggiore affidabilità.

Sequence	**AdaFace (%)**	**MagFace (%)**
GLOBAL	40.36	3.65
sequence_01_frontal_gaze	61.54	13.85
sequence_01_with_occlusion	45.31	0.00
sequence_01_looking_around	45.31	0.00
sequence_02_frontal_gaze	38.46	7.69
sequence_02_with_occlusion	20.63	0.00
sequence_02_looking_around	25.40	0.00

Tabella 5.14: Valori di FAR (%) calcolati con aggregazione media degli score, per AdaFace e MagFace.

Dal punto di vista tecnico, la media rappresenta un criterio valido, in quanto riduce l'influenza dei singoli frame non conformi, fornendo una stima più stabile della similarità complessiva. A livello applicativo, questo criterio risulta più adatto a scenari reali, dove il sistema deve valutare sequenze video e non singoli frame isolati.

Valore massimo degli score

Dalla Tabella 5.15 si può notare che l'uso del criterio del valore massimo porta a valori di FAR estremamente elevati per entrambi i modelli. La differenza rimane significativa soprattutto nelle sequenze con occlusioni o variazioni di posa, dove MagFace riduce il FAR di circa 10–20 punti percentuali rispetto al concorrente. Dal punto di vista applicativo, questo criterio risulta poco adatto agli scenari reali, soprattutto in contesti ad alta sicurezza come i varchi aeroportuali, in cui non è accettabile che un singolo frame comprometta l'intero processo di verifica. Sebbene MagFace mostri una maggiore robustezza rispetto ad AdaFace, i valori assoluti di FAR rimangono troppo elevati per garantire l'affidabilità operativa.

Sequence	AdaFace (%)	MagFace (%)
GLOBAL	97.40	85.94
sequence_01_frontal_gaze	98.46	93.85
sequence_01_with_occlusion	95.31	84.38
sequence_01_looking_around	97.66	86.72
sequence_02_frontal_gaze	96.92	90.77
sequence_02_with_occlusion	95.24	80.95
sequence_02_looking_around	96.83	78.57

Tabella 5.15: Valori di FAR (%) calcolati con aggregazione massima degli score di similarità, per AdaFace e MagFace.

Valore minimo degli score

I risultati riportati nella Tabella 5.16 sono calcolati con il criterio del valore minimo, che porta i FAR dei due modelli a zero. In particolare, MagFace ottiene valori nulli in tutte le sequenze, mentre AdaFace presenta deviazioni minime (fino all'1.56% nel caso di occlusioni). Questo risultato riflette la presenza, in quasi tutte le sequenze, di almeno un frame altamente discriminante, anche per i tentativi impostori.

Sequence	AdaFace (%)	MagFace (%)
GLOBAL	0.26	0.00
sequence_01_frontal_gaze	0.00	0.00
sequence_01_with_occlusion	1.56	0.00
sequence_01_looking_around	0.78	0.00
sequence_02_frontal_gaze	0.00	0.00
sequence_02_with_occlusion	0.00	0.00
sequence_02_looking_around	0.00	0.00

Tabella 5.16: Valori minimi di FAR (%) calcolati per AdaFace e MagFace.

Dal punto di vista tecnico, il criterio del valore minimo adotta una politica conservativa che riduce al minimo le false accettazioni, ma aumenta il rischio di falsi rifiuti, in quanto basta un frame degradato per invalidare l'intera sequenza. A livello applicativo, il criterio del minimo non è adatto come regola decisionale finale, ma rappresenta un utile indicatore del limite inferiore del sistema e uno strumento per la selezione dei frame più affidabili.

Criteri basati sulla qualità

In questa sezione vengono analizzati i criteri di selezione dei frame basati su misure di qualità, con l'obiettivo di valutare se e in che misura possano ridurre il tasso di false accettazioni nei modelli considerati.

SER-FIQ La selezione dei frame con SER-FIQ evidenzia una differenza tra i due modelli. Come mostrato nella Tabella 5.17, AdaFace mantiene valori di FAR elevati in tutte le sequenze, anche quando vengono selezionati i frame di qualità migliore.

Al contrario, come mostrato nella Tabella 5.18, MagFace beneficia maggiormente di questo criterio, in quanto il FAR si riduce sensibilmente, soprattutto in presenza di occlusioni e variazioni di posa. Ciò conferma che l'uso di metriche di qualità come SER-FIQ può migliorare la robustezza della verifica video, risultando efficace con MagFace, mentre l'impatto su AdaFace rimane limitato.

Sequence	**Q1 (%)**	**Q3 (%)**	**Q5 (%)**
sequence_01_frontal_gaze	84.13	81.54	81.54
sequence_01_with_occlusion	75.93	77.42	75.81
sequence_01_looking_around	72.13	74.60	77.78
sequence_02_frontal_gaze	65.08	70.77	75.38
sequence_02_with_occlusion	65.57	70.97	69.84
sequence_02_looking_around	57.14	63.49	66.67

Tabella 5.17: Valori di FAR (%) per AdaFace con selezione dei frame tramite SER-FIQ (Q1, Q3, Q5).

Sequence	**Q1 (%)**	**Q3 (%)**	**Q5 (%)**
sequence_01_frontal_gaze	63.08	61.54	53.85
sequence_01_with_occlusion	34.38	37.50	32.81
sequence_01_looking_around	39.06	34.38	35.94
sequence_02_frontal_gaze	44.62	46.15	43.08
sequence_02_with_occlusion	31.75	26.98	28.57
sequence_02_looking_around	36.51	31.75	31.75

Tabella 5.18: Valori di FAR (%) per MagFace con selezione dei frame tramite SER-FIQ (Q1, Q3, Q5).

OFIQ In questo paragrafo vengono analizzati i risultati ottenuti applicando diversi criteri di selezione dei frame basati su misure di qualità fornite da

OFIQ. In particolare, sono state considerate tre varianti: *Unified Quality Score*, *CompressionArtifacts* e *InterEyeDistance*, in quanto questi fattori sono ritenuti tra i più influenti sulla qualità delle immagini.

- **Unified Quality Score:** La selezione dei frame tramite OFIQ conferma quanto osservato con SER-FIQ. Come mostrato nella Tabella 5.19, AdaFace mantiene valori di FAR elevati (70–85%), con riduzioni marginali anche scegliendo i frame di qualità più alta. Al contrario, i risultati della Tabella 5.20 di MagFace sembrano produrre effetti positivi, con un calo sensibile del FAR, specialmente in presenza di occlusioni e variazioni di posa, che si stabilizza intorno al $20 - 30\%$.

Sequence	**Q1 (%)**	**Q3 (%)**	**Q5 (%)**
sequence_01_frontal_gaze	83.05	83.61	80.65
sequence_01_with_occlusion	84.62	84.00	82.69
sequence_01_looking_around	80.77	74.14	72.13
sequence_02_frontal_gaze	75.81	70.31	70.31
sequence_02_with_occlusion	80.00	72.55	71.15
sequence_02_looking_around	71.15	67.80	60.00

Tabella 5.19: Valori di FAR (%) per AdaFace con selezione dei frame tramite il criterio *UnifiedQualityScore* (Q1, Q3, Q5)

Sequence	**Q1 (%)**	**Q3 (%)**	**Q5 (%)**
sequence_01_frontal_gaze	53.85	55.38	55.38
sequence_01_with_occlusion	37.50	34.38	34.38
sequence_01_looking_around	28.13	31.25	29.69
sequence_02_frontal_gaze	58.46	55.38	55.38
sequence_02_with_occlusion	19.67	19.35	19.05
sequence_02_looking_around	23.81	19.05	17.46

Tabella 5.20: Valori di FAR (%) per MagFace con selezione dei frame tramite il criterio *UnifiedQualityScore* (Q1, Q3, Q5)

- **Compression Artifacts:** Dalla Tabella 5.21 emerge che, con il criterio *CompressionArtifacts*, AdaFace mantiene valori di FAR elevati (55–77%) con riduzioni limitate anche selezionando i frame più affidabili. Al contrario, la Tabella 5.22 mostra che MagFace trae vantaggio significativo dalla selezione. Il FAR diminuisce significativamente fino a valori inferiori al 10% nelle condizioni più critiche. Questo risultato evidenzia che, mentre AdaFace rimane vulnerabile agli artefatti di compressione,

MagFace riesce a sfruttare efficacemente tale metrica di qualità, rilevandosi più adatto agli scenari applicativi reali in cui i video sono spesso sottoposti a compressione.

Sequence	**Q1 (%)**	**Q3 (%)**	**Q5 (%)**
sequence_01_frontal_gaze	72.22	77.05	74.19
sequence_01_with_occlusion	65.63	69.05	68.00
sequence_01_looking_around	70.73	60.38	70.49
sequence_02_frontal_gaze	59.26	56.45	57.14
sequence_02_with_occlusion	63.64	55.56	57.45
sequence_02_looking_around	63.41	52.83	48.28

Tabella 5.21: Valori di FAR (%) per AdaFace con selezione dei frame tramite il criterio *CompressionArtifacts* (Q1, Q3, Q5)

Sequence	**Q1 (%)**	**Q3 (%)**	**Q5 (%)**
sequence_01_frontal_gaze	42.19	41.54	33.85
sequence_01_with_occlusion	21.88	14.06	7.81
sequence_01_looking_around	23.44	10.94	10.94
sequence_02_frontal_gaze	33.85	29.23	24.62
sequence_02_with_occlusion	11.48	6.45	3.17
sequence_02_looking_around	14.52	6.35	7.94

Tabella 5.22: Valori di FAR (%) per MagFace con selezione dei frame tramite il criterio *CompressionArtifacts* (Q1, Q3, Q5)

- **Inter Eye Distance:** Dalla Tabella 5.23 si nota che AdaFace sembra ottenere valori di FAR più contenuti rispetto ad altre metriche, in alcuni casi anche inferiori al 20%. Questo potrebbe indicare un miglioramento relativo, pur rimanendo meno stabile di MagFace. Come riportato nella Tabella 5.24, MagFace mantiene valori nulli in tutte le sequenze, suggerendo una maggiore robustezza. A livello applicativo, l'utilità del criterio *InterEyeDistance* sembra comunque limitata e potrebbe essere considerato più come un supporto per altre metriche piuttosto che come uno strumento autonomo di selezione.

Sequence	Q1 (%)	Q3 (%)	Q5 (%)
sequence_01_frontal_gaze	0.00	23.81	28.85
sequence_01_with_occlusion	8.33	21.21	32.61
sequence_01_looking_around	7.14	19.05	22.45
sequence_02_frontal_gaze	7.14	14.63	15.09
sequence_02_with_occlusion	18.75	8.11	16.28
sequence_02_looking_around	8.33	7.89	13.33

Tabella 5.23: Valori di FAR (%) per AdaFace con selezione dei frame tramite il criterio *InterEyeDistance* (Q1, Q3, Q5)

Sequence	Q1 (%)	Q3 (%)	Q5 (%)
sequence_01_frontal_gaze	0.00	0.00	0.00
sequence_01_with_occlusion	0.00	0.00	0.00
sequence_01_looking_around	0.00	0.00	0.00
sequence_02_frontal_gaze	0.00	0.00	0.00
sequence_02_with_occlusion	0.00	0.00	0.00
sequence_02_looking_around	0.00	0.00	0.00

Tabella 5.24: Valori di FAR (%) per MagFace con selezione dei frame tramite il criterio *InterEyeDistance* (Q1, Q3, Q5)

5.3.3 Confronto complessivo sul FAR

Dalla Figura 5.2 si nota chiaramente la differenza tra i due modelli. Per AdaFace, le metriche di qualità non portano miglioramenti significativi: il modello resta vulnerabile e instabile. L'uso della media degli score riduce in parte il problema, ma i risultati non sono abbastanza solidi per gli scenari in cui è neccessaria un'alta sicurezza. Per MagFace, invece, la media degli score è il metodo più efficace, in quanto assicura prestazioni affidabili anche in condizioni difficili come l'occlusione o i cambi di posa. Le metriche di qualità possono comunque essere utili, ma non raggiungono i benefici garantiti dall'aggregazione statistica.

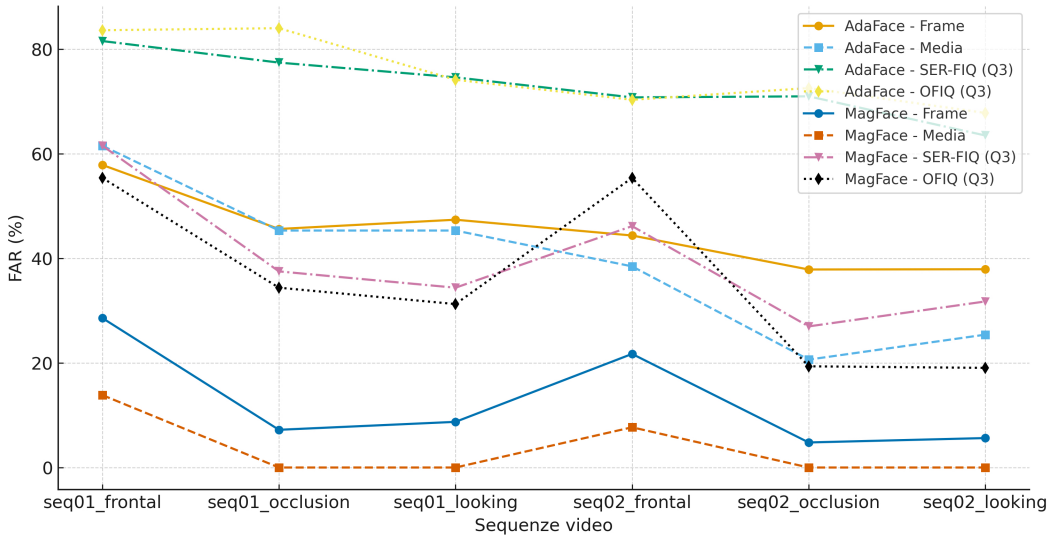

Figura 5.2: Andamento del FAR (%) per AdaFace e MagFace sulle diverse sequenze video, considerando i criteri di aggregazione Frame, Media, SER-FIQ (Q3) e OFIQ Unified (Q3)

Conclusioni e sviluppi futuri

Il lavoro di ricerca ha messo in luce differenze sostanziali tra i due modelli di riconoscimento facciale analizzati, ciascuno caratterizzato da specifici punti di forza e limiti. AdaFace ha mostrato una maggiore stabilità nei confronti genuini, con valori di FRR contenuti anche senza l'applicazione di criteri di aggregazione, soprattutto nelle sequenze acquisite in condizioni favorevoli, prive di variazioni di posa o occlusioni. In questo contesto, l'impiego della metrica di qualità SER-FIQ ha permesso di raggiungere risultati ottimali, riducendo il FRR a valori prossimi allo zero. Tuttavia, in presenza di immagini morphed, AdaFace ha evidenziato una marcata vulnerabilità: sebbene l'aggregazione tramite la media degli score abbia consentito una riduzione parziale del FAR, i valori raggiunti non sono sufficienti a garantire livelli di sicurezza adeguati in scenari applicativi reali.

MagFace, al contrario, non ha sempre mostrato prestazioni eccellenti in termini assoluti, ma ha registrato un comportamento più equilibrato e, soprattutto, una maggiore capacità di contenere le false accettazioni. Già senza criteri di aggregazione, i valori di FAR sono rimasti sensibilmente inferiori a quelli di AdaFace, con un range compreso tra il 5% e il 28%. L'applicazione della media degli score ha ulteriormente migliorato la robustezza del modello, azzerando il FAR in diverse sequenze. Per quanto riguarda il FRR, MagFace ha mostrato un miglioramento significativo grazie all'uso della metrica SER-FIQ, che ha permesso di ridurre gli errori di rifiuto a circa il $9 - 10\%$ nelle sequenze frontali, evidenziando un effetto positivo, seppur non risolutivo.

Nel complesso, i risultati confermano che l'introduzione di criteri di aggregazione e di metriche di qualità può incrementare la resistenza dei sistemi di riconoscimento agli attacchi di morphing in scenari video-based. Le strategie che hanno mostrato gli effetti più significativi sono state l'aggregazione tramite media, capace di stabilizzare le prestazioni complessive, e l'impiego di SER-FIQ, che ha prodotto un impatto positivo su entrambi i modelli. Restano tuttavia alcune criticità: AdaFace continua a mostrare una marcata vulnerabilità in termini di FAR, mentre MagFace, pur risultando più affidabile, non raggiunge prestazioni ottimali in termini di FRR.

Resta comunque spazio per ulteriori miglioramenti: alcuni valori di sco-

re hanno infatti evidenziato margini di ottimizzazione che meritano di essere esplorati. Tra le prospettive future vi sono l'integrazione di strategie di fusione tra più metriche di qualità, l'ampliamento e la diversificazione del dataset e la valutazione in scenari operativi più ampi. Nonostante i suoi limiti, questo lavoro contribuisce alla comprensione delle potenzialità e dei limiti delle tecniche video-based per la rilevazione delle morphing attack, ponendo le basi per futuri sviluppi orientati a soluzioni pratiche e affidabili.

Ringraziamenti

Giungere a questo traguardo non è stato semplice. Sono stati necessari giorni e notti di studio intenso, spesso fino allo sfinimento, accompagnati da momenti di scoraggiamento e di dubbio sulle mie capacità. Tuttavia, oggi desidero ringraziare la me stessa di allora, che non ha mai abbandonato la strada intrapresa. Ricorderò sempre l'impegno profondo per affrontare il mio primo esame, quando la tentazione di arrendermi era forte. Allora non avrei mai immaginato che, tre anni dopo, sarei riuscita a completare il mio percorso accademico. Oggi, guardandomi indietro, provo una profonda gratitudine per la determinazione che mi ha permesso di arrivare fino a questo momento.

Vorrei esprimere la mia sincera gratitudine al mio relatore, Prof. Matteo Ferrara. La scelta di un supervisore era inizialmente motivo di incertezza, ma il suo approccio serio, paziente e accurato mi ha subito fatto capire di aver fatto la scelta giusta. Grazie alla sua guida competente e costante, ho potuto crescere e maturare non solo dal punto di vista accademico, ma anche personale. A lui devo gran parte del percorso che mi ha portato a questo risultato. Un ringraziamento va anche alla mia correlatrice, Prof.ssa Annalisa Franco, per la sua disponibilità, la sua gentilezza e la capacità di mettere a proprio agio ogni studente, qualità che hanno reso il mio cammino più sereno e stimolante.

Un pensiero speciale va alla mia sorella, Elena, che per me è sempre stata una guida e un punto di riferimento. Nei momenti più difficili, la sua presenza mi ha sostenuta e incoraggiata. La considero la mia forza più grande e il mio modello, un esempio che continuerà a guidarmi anche in futuro.

Non possono mancare i miei genitori, ai quali devo tutto ciò che sono. Ho sempre cercato di dimostrare loro la mia crescita e la mia indipendenza, pur sapendo che, per loro, resterò sempre la bambina di un tempo. In particolare, desidero ringraziare mia madre, che non mi ha mai fatto mancare sostegno, vicinanza e incoraggiamento, anche nei momenti più complessi. A lei devo un grazie speciale, con l'impegno a continuare a dare il meglio di me.

Un grazie sincero va anche alle mie compagne e ai miei compagni di università, alcuni dei quali mi hanno accompagnato sin dai tempi delle superiori. La loro presenza ha reso più leggere le giornate di studio e ha arricchito di ricordi

e sorrisi anche i momenti più impegnativi. Condividere questo percorso con loro è stato un vero privilegio.

In conclusione, ciascuna delle persone che ho incontrato lungo questo cammino ha contribuito, con la propria presenza e il proprio sostegno, a formare la persona che sono oggi. Con la volontà di continuare a crescere e scoprire la mia strada, guardo con fiducia alle sfide che mi attendono.

轻舟已过万重山

Elisa Yan
Settembre 2025

Bibliografia

[1] Abrosoft. Fantamorph. `https://www.fantamorph.com/`, 2021. Accessed: 2025-09-10.

[2] Alyaa Qusay Aloraibi. Image morphing techniques: A review. *Technium*, 9, 2023.

[3] Ilias Batskos and Luuk J. Spreeuwers. Improving fully automated landmark-based face morphing. In *Proceedings of the 12th International Workshop on Biometrics and Forensics (IWBF 2024)*, pages 1–6. IEEE, 2024.

[4] Guido Borghi, Nicolò Di Domenico, Annalisa Franco, Matteo Ferrara, and Davide Maltoni. Revelio: A modular and effective framework for reproducible training and evaluation of morphing attack detectors. *IEEE Access*, 11:120419–120437, 2023.

[5] Guido Borghi, Annalisa Franco, Nicolò Di Domenico, Matteo Ferrara, and Davide Maltoni. V-mad: Video-based morphing attack detection in operational scenarios. In *2024 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–10. IEEE, 2024.

[6] Fadi Boutros, Meiling Fang, Marcel Klemt, Biying Fu, and Naser Damer. Cr-fiqa: face image quality assessment by learning sample relative classifiability. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5836–5845, 2023.

[7] Oya Çeliktutan, Sezer Ulukaya, and Bülent Sankur. A comparative study of face landmarking techniques. *EURASIP Journal on Image and Video Processing*, 2013:1–27, 2013.

[8] Naser Damer, Viola Boller, Yaza Wainakh, Fadi Boutros, Philipp Terhörst, Andreas Braun, and Arjan Kuijper. Detecting face morphing attacks by analyzing the directed distances of facial landmarks shifts. In *German Conference on Pattern Recognition*, pages 518–534. Springer, 2018.

[9] Naser Damer, Fadi Boutros, Alexandra Mosegui Saladie, Florian Kirchbuchner, and Arjan Kuijper. Realistic dreams: Cascaded enhancement of gan-generated images with an example in face morphing attacks. In *2019 IEEE 10th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–10. IEEE, 2019.

[10] Naser Damer, Alexandra Mosegui Saladie, Andreas Braun, and Arjan Kuijper. Morgan: Recognition vulnerability and attack detectability of face morphing attacks created by generative adversarial network. In *2018 IEEE 9th international conference on biometrics theory, applications and systems (BTAS)*, pages 1–10. IEEE, 2018.

[11] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4690–4699, 2019.

[12] Department of Internal Affairs (DIA), New Zealand. Passport photo requirements. `https://www.passports.govt.nz/passport-photos/passport-photo-requirements/`. [Online; accessed 27-May-2020].

[13] M. Ferrara, A. Franco, D. Maio, and D. Maltoni. Face image conformance to iso/icao standards in machine readable travel documents. *IEEE Transactions on Information Forensics and Security*, 7(4):1204–1213, August 2012.

[14] Matteo Ferrara and Annalisa Franco. Morph creation and vulnerability of face recognition systems to morphing. In *Handbook of Digital Face Manipulation and Detection: From DeepFakes to Morphing Attacks*, pages 117–137. Springer International Publishing Cham, 2022.

[15] Matteo Ferrara, Annalisa Franco, Dario Maio, and Davide Maltoni. Face image conformance to ISO/ICAO standards in machine readable travel documents. *IEEE Trans. Inf. Forensics Secur.*, 7(4):1204–1213, 2012.

[16] Matteo Ferrara, Annalisa Franco, and Davide Maltoni. The magic passport. In *IEEE International Joint Conference on Biometrics*, pages 1–7, 2014.

[17] Matteo Ferrara, Annalisa Franco, and Davide Maltoni. Face demorphing. *IEEE Transactions on Information Forensics and Security*, 13(4):1008–1017, 2018.

[18] Matteo Ferrara, Annalisa Franco, and Davide Maltoni. Face demorphing in the presence of facial appearance variations. In *2018 26th European Signal Processing Conference (EUSIPCO)*, pages 2365–2369. IEEE, 2018.

[19] Matteo Ferrara, Annalisa Franco, and Davide Maltoni. Decoupling texture blending and shape warping in face morphing. In *2019 international conference of the biometrics special interest group (BIOSIG)*, pages 1–5. IEEE, 2019.

[20] Matteo Ferrara, Annalisa Franco, and Davide Maltoni. Face morphing detection in the presence of printing/scanning and heterogeneous image sources. *IET Biometrics*, 10(3):290–303, 2021.

[21] Matteo Ferrara, Annalisa Franco, Davide Maltoni, and Yunlian Sun. On the impact of alterations on face photo recognition accuracy. In *Image Analysis and Processing–ICIAP 2013: 17th International Conference, Naples, Italy, September 9-13, 2013. Proceedings, Part I 17*, pages 743–751. Springer, 2013.

[22] FRONTEX. Frontex — european union agency. `https://www.frontex.europa.eu/`, July 2014. [Online; accessed August 2025].

[23] RDU Frontex. Best practice operational guidelines for automated border control (abc) systems. *European Agency for the Management of Operational Cooperation, Research and Development Unit,. https://bit.ly/2KYBXhz Accessed*, 9(05):2013, 2012.

[24] GIMP. Gimp animation package. `https://www.gimp.org/news/2009/06/05/gimp-animation-package-260-released/`, 2021. Accessed: 2025-09-10.

[25] HM Government. Passport photo requirements. `https://www.gov.uk/photos-for-passports/photo-requirements`. [Online; accessed 27-May-2020].

[26] Yuge Huang, Yuhan Wang, Ying Tai, Xiaoming Liu, Pengcheng Shen, Shaoxin Li, Jilin Li, and Feiyue Huang. Curricularface: adaptive curriculum learning loss for deep face recognition. In *proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5901–5910, 2020.

[27] TAG MRTD ICAO. Ntwg. biometrics deployment of machine readable travel documents. *Development and Specification of Globally Interoperable Biometric Standards for Machine Assisted Identity Confirmation using Machine Readable Travel Documents*, 2004.

[28] ISO/IEC. Information technology – biometric data interchange formats – part 5: Face image data. Technical Report ISO/IEC 19794-5:2011, International Organization for Standardization, Geneva, Switzerland, 2011.

[29] Minchul Kim, Anil K Jain, and Xiaoming Liu. Adaface: Quality adaptive margin for face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.

[30] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.

[31] John Lewis. Automated lip-sync: Background and techniques. *The Journal of Visualization and Computer Animation*, 2(4):118–122, 1991.

[32] Min Long, Quantao Yao, Le-Bing Zhang, and Fei Peng. Face de-morphing based on diffusion autoencoders. *IEEE Transactions on Information Forensics and Security*, 19:3051–3063, 2024.

[33] Luxand. Facemorpher. `http://www.facemorpher.com/`, 2021. Accessed: 2025-09-10.

[34] Andrey Makrushin, Tom Neubert, and Jana Dittmann. Automatic generation and detection of visually faultless facial morphs. In *International conference on computer vision theory and applications*, volume 7, pages 39–50. SciTePress, 2017.

[35] Qiang Meng, Shichao Zhao, Zhida Huang, and Feng Zhou. Magface: A universal representation for face recognition and quality assessment. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14225–14234, 2021.

[36] Johannes Merkle, Christian Rathgeb, Benjamin Herdeanu, Benjamin Tams, Day-Parn Lou, André Dörsch, Maxim Schauberst, Jonas Dehen, Liming Chen, Xiangnan Yin, Di Huang, Anna Stratmann, Marcel Ginzler, Marcel Grimmer, and Christoph Busch. Ofiq project – short public report v1.0. Technical report, German Federal Office for Information Security (BSI), September 2024.

[37] Ramachandra Raghavendra, KiranB Raja, Sushma Venkatesh, and Christoph Busch. Face morphing versus face averaging: Vulnerability and detection. In *2017 IEEE International Joint Conference on Biometrics (IJCB)*, pages 555–563. IEEE, 2017.

[38] Kiran Raja, Matteo Ferrara, Annalisa Franco, Luuk Spreeuwers, Ilias Batskos, Florens De Wit, Marta Gomez-Barrero, Ulrich Scherhag, Daniel Fischer, Sushma Krupa Venkatesh, et al. Morphing attack detection-database, evaluation platform, and benchmarking. *IEEE transactions on information forensics and security*, 16:4336–4351, 2020.

[39] Kiran Raja, Sushma Venkatesh, RB Christoph Busch, et al. Transferable deep-cnn features for detecting digital and print-scanned morphed face images. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 10–18, 2017.

[40] Raghavendra Ramachandra, Sushma Venkatesh, Kiran Raja, and Christoph Busch. Towards making morphing attack detection robust using hybrid scale-space colour texture features. In *2019 IEEE 5th International Conference on Identity, Security, and Behavior Analysis (ISBA)*, pages 1–8. IEEE, 2019.

[41] Christian Rathgeb, Ruben Tolosana, Ruben Vera-Rodriguez, and Christoph Busch. *Handbook of digital face manipulation and detection: from DeepFakes to morphing attacks*. Springer Nature, 2022.

[42] Ulrich Scherhag, Dhanesh Budhrani, Marta Gomez-Barrero, and Christoph Busch. Detecting morphed face images using facial landmarks. In *International conference on image and signal processing*, pages 444–452. Springer, 2018.

[43] Ulrich Scherhag, Ramachandra Raghavendra, Kiran B Raja, Marta Gomez-Barrero, Christian Rathgeb, and Christoph Busch. On the vulnerability of face recognition systems towards morphed face attacks. In *2017 5th international workshop on biometrics and forensics (IWBF)*, pages 1–6. IEEE, 2017.

[44] Ulrich Scherhag, Christian Rathgeb, Johannes Merkle, Ralph Breithaupt, and Christoph Busch. Face recognition systems under morphing attacks: A survey. *IEEE Access*, 7:23012–23026, 2019.

[45] Ulrich Scherhag, Christian Rathgeb, Johannes Merkle, and Christoph Busch. Deep face representations for differential morphing attack detection. *IEEE transactions on information forensics and security*, 15:3625–3639, 2020.

[46] Torsten Schlett, Christian Rathgeb, Olaf Henniger, Javier Galbally, Julian Fierrez, and Christoph Busch. Face image quality assessment: A literature survey. *ACM Computing Surveys (CSUR)*, 54(10s):1–49, 2022.

[47] Eli Shechtman, Alex Rav-Acha, Michal Irani, and Steve Seitz. Regenerative morphing. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 615–622. IEEE, 2010.

[48] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[49] Jag Mohan Singh and Raghavendra Ramachandra. Fusion of deep features for differential face morphing attack detection at automatic border control gates. In *2022 10th European Workshop on Visual Information Processing (EUVIP)*, pages 1–5. IEEE, 2022.

[50] Philipp Terhorst, Jan Niklas Kolf, Naser Damer, Florian Kirchbuchner, and Arjan Kuijper. Ser-fiq: Unsupervised estimation of face image quality based on stochastic embedding robustness. In *Proceedings of the IEEE/-CVF conference on computer vision and pattern recognition*, pages 5651–5660, 2020.

[51] Luan Tran, Xi Yin, and Xiaoming Liu. Disentangled representation learning gan for pose-invariant face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1415–1424, 2017.

[52] Sushma Venkatesh, Raghavendra Ramachandra, Kiran Raja, and Christoph Busch. Face morphing attack generation and detection: A comprehensive survey. *IEEE transactions on technology and society*, 2(3):128–145, 2021.

[53] Sushma Venkatesh, Raghavendra Ramachandra, Kiran Raja, Luuk Spreeuwers, Raymond Veldhuis, and Christoph Busch. Morphed face detection based on deep color residual noise. In *2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pages 1–6. IEEE, 2019.

[54] Sushma Venkatesh, Raghavendra Ramachandra, Kiran Raja, Luuk Spreeuwers, Raymond Veldhuis, and Christoph Busch. Detecting morphed face attacks using residual noise from deep multi-scale context aggregation network. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 280–289, 2020.

[55] Sushma Venkatesh, Haoyu Zhang, Raghavendra Ramachandra, Kiran Raja, Naser Damer, and Christoph Busch. Can gan generated morphs

threaten face recognition systems equally as landmark based morphs?–vulnerability and detection. *arXiv preprint arXiv:2007.03621*, 2020.

[56] Hao Wang, Yitong Wang, Zheng Zhou, Xing Ji, Dihong Gong, Jingchao Zhou, Zhifeng Li, and Wei Liu. Cosface: Large margin cosine loss for deep face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5265–5274, 2018.