

Rilevamento automatico del trucco facciale

Michele Nardini

0001125986



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Corso di Laurea Magistrale in Ingegneria Informatica
UNIVERSITÀ DI BOLOGNA

Gennaio 2024

Table of Contents

List of Figures	1
List of Tables	2
1: Introduzione	3
1.1 Contesto	4
1.2 Problema Affrontato	4
2: Stato dell'arte	6
2.1 L'Impatto del Trucco sul Riconoscimento Facciale	7
2.2 Algoritmi di Localizzazione dei Landmark e Estrazione delle ROI	8
2.3 Algoritmi di Estrazione delle Feature	9
2.4 Algoritmi di Classificazione	10
2.5 Dataset esistenti per il riconoscimento e classificazione visi senza trucco e con	12
3: Approccio Sviluppato	14
3.1 Rilevamento del viso e localizzazione dei landmark	16
3.2 Estrazione delle feature	18
3.2.1 Feature basate sul colore	18
3.2.2 Feature basate sulla forma	19
3.2.3 Feature basate su texture	22

3.3	Classificazione	23
3.3.1	Support Vector Machine(SVM)	24
3.3.2	Adaboost	24
4:	Valutazione Sperimentale	27
4.1	Dataset	27
4.2	Sperimentazione	28
5:	Conclusioni	34

List of Figures

3.1	Visualizzazione di un viso nei corrispondenti canali dei colorspace RGB(sopra) e HSV(sotto)	15
3.2	Visualizzazione dei landmark di interesse	17
3.3	Visualizzazione di μ , σ e skewness nelle due tipologie di ROI.	19
3.4	Applicazione dei filtri di Gabor sulle due tipologie di visi.	21
3.5	Applicazione di EOH sulle due tipologie di visi.	22
4.1	Esempio di visi truccati (sopra) e non truccati (sotto).	28
4.2	Dalla prima immagine alla quinta la curva ROC del detector SVM e Adaboost per tutte e 5 iterazioni nella cross-validation. L'ultima immagine mostra la curva ROC dei due detector testati su un dataset diverso da quello di training.	31
4.3	Matrice Confusione	32
4.4	Risultati delle previsioni	33

List of Tables

1.1	Esempi di prodotti di trucco che alterano l'aspetto del viso.	5
3.1	Dimensionalità del vettore delle features	23
4.1	Tassi di classificazione (Classification Rate) ottenuti con SVM e AdaBoost su diverse ROI e combinazioni di ROI.	30

1 Introduzione

Negli ultimi decenni, il riconoscimento facciale è diventato una delle principali aree di ricerca nell'ambito della visione artificiale e dell'intelligenza artificiale. Questa tecnologia trova applicazioni in settori come la sicurezza, l'assistenza sanitaria, il marketing e l'intrattenimento. Tuttavia, l'accuratezza e la robustezza dei sistemi di riconoscimento facciale sono spesso influenzate da fattori esterni, come variazioni di illuminazione, pose, espressioni facciali e, in particolare, la presenza di trucco. Il trucco può infatti influire sui sistemi di riconoscimento facciale, sebbene l'entità di questo effetto possa variare in base a diversi fattori, tra cui il tipo e l'applicazione del trucco.

Effetti del trucco sul riconoscimento facciale:

- **Alterazione delle caratteristiche:** Il trucco può cambiare l'aspetto delle caratteristiche facciali. Ad esempio un eyeliner marcato può creare ombre e punti luce che confondono gli algoritmi di riconoscimento facciale, i quali si basano su caratteristiche stabili per l'identificazione.
- **Variabilità:** Il trucco può introdurre variazioni nel tono e nella texture della pelle, complicando ulteriormente il processo di riconoscimento. I sistemi di riconoscimento facciale sono addestrati su una vasta gamma di volti, ma cambiamenti estremi causati dal trucco possono portare a errori di identificazione.
- **Sensibilità degli algoritmi:** Alcuni sistemi di riconoscimento facciale sono più sensibili ai cambiamenti nelle caratteristiche facciali rispetto ad altri. I sistemi avanzati basati su deep learning possono essere più capaci di adattarsi alle variazioni causate dal trucco, mentre i sistemi più semplici possono avere maggiori difficoltà.

1.1 Contesto

Il trucco facciale rappresenta quindi una sfida unica per i sistemi di visione artificiale. Esso può alterare significativamente le caratteristiche visive di una persona, mascherando dettagli importanti o enfatizzando altri, e quindi ridurre l'affidabilità dei modelli di riconoscimento. La tabella 1 da alcuni esempi. Questa problematica è particolarmente rilevante in contesti dove è necessario garantire la coerenza tra le immagini facciali, come nei sistemi di sicurezza o nei dataset utilizzati per l'addestramento di modelli.

Di fronte a queste sfide, emerge la necessità di sviluppare metodi che possano distinguere accuratamente tra visi truccati e non truccati, migliorando così l'affidabilità delle applicazioni basate su immagini facciali. Questo progetto si inserisce in questo contesto, con l'obiettivo di affrontare le difficoltà legate alla classificazione dei visi e di proporre un classificatore robusto ed efficace per il riconoscimento del trucco facciale.

1.2 Problema Affrontato

Il problema affrontato in questo progetto consiste nella classificazione di immagini facciali in due categorie: makeup (trucco presente) e no makeup (trucco assente). La difficoltà principale risiede nel fatto che il trucco facciale può alterare le caratteristiche visive in modo non uniforme. Ad esempio:

- Il trucco degli occhi può coprire le ombre naturali, modificare il contrasto o aggiungere texture artificiali.
- Il rossetto può cambiare il colore naturale delle labbra, introducendo variazioni significative nei canali cromatici.
- Il trucco del viso (fondotinta, cipria) può uniformare la texture della pelle, alterando le informazioni di profondità e colore.

L'obiettivo principale è progettare una pipeline in grado di:

- Estrarre con precisione le ROI (Regioni di Interesse) dal viso, includendo:
 - Occhi: Escludendo sopracciglia e altre aree non pertinenti.
 - Bocca: Limitando la regione al contorno delle labbra.

- Viso: Escludendo capelli e orecchie.
- Analizzare e classificare ciascuna ROI individualmente per comprendere il contributo di ogni parte del viso alla classificazione.
- Combinare le informazioni delle ROI per migliorare le prestazioni del classificatore complessivo.

Sfide

- Dataset Sbilanciati: I dataset disponibili (ad esempio, CelebA) presentano spesso una distribuzione non uniforme tra le classi makeup e no makeup.
- Accuratezza delle ROI estrapolate: La qualità delle ROI dipende dalla precisione dei landmark facciali utilizzati.
- Rumore nei Dati: Presenza di variazioni legate a illuminazione, espressioni facciali e pose non standard.
- Combinazione delle ROI: Determinare come combinare efficacemente le informazioni provenienti da occhi, bocca e viso per ottenere una classificazione accurata.

Regione del viso	Prodotti di trucco correlati
ROI degli occhi	Mascara, ombretto, ciglia finte, matite per sopracciglia, creme, cere, gel e polveri
ROI delle labbra	Rossetto, lucidalabbra, volumizzanti, balsamo
Aspetto generale della pelle	Correttore, fondotinta, cipria, rouge, blush o fard, polveri/creme per contouring, illuminante, bronzer

Table 1.1: Esempi di prodotti di trucco che alterano l'aspetto del viso.

2 Stato dell'arte

L'impatto del trucco sui sistemi di riconoscimento facciale rappresenta una sfida rilevante nel campo della visione artificiale e del machine learning. Negli ultimi anni, l'attenzione della ricerca si è focalizzata su come il trucco possa alterare la capacità dei sistemi di identificare correttamente un individuo, nonché sulle soluzioni tecniche per mitigare queste problematiche.

In questo capitolo analizzeremo i seguenti aspetti principali:

- **L'impatto del trucco sul riconoscimento facciale:** Approfondiremo come il trucco possa alterare tratti del viso e introdurre variazioni che sfidano gli algoritmi di riconoscimento. Verranno discussi gli effetti su texture, geometria e caratteristiche locali del viso.
- **Algoritmi e tecnologie utilizzati nel riconoscimento facciale:** Verranno presentati i principali approcci tecnologici, inclusi i modelli deep learning, tecniche di allineamento dei landmark e metodi ensemble per l'analisi delle regioni facciali specifiche come occhi, bocca e viso completo.
- **Dataset rilevanti per lo studio del trucco e del riconoscimento facciale:** Discuteremo i principali dataset utilizzati per addestrare e valutare i modelli, con un'attenzione particolare ai dataset bilanciati e ricchi di annotazioni relative al trucco.
- **Metodi di valutazione delle prestazioni dei modelli:** Esamineremo le metriche comunemente usate per valutare i classificatori, con un focus sulla gestione degli sbilanciamenti nel dataset e sulla robustezza dei modelli.

Questo capitolo fornirà una visione complessiva delle ricerche e delle tecnologie attualmente esistenti, nonché delle sfide aperte nel riconoscimento facciale in presenza di

trucco. L'obiettivo è contestualizzare il lavoro svolto in questo progetto, evidenziando i punti di forza e i limiti delle soluzioni attuali, oltre a definire le basi teoriche e pratiche su cui si fonda il nostro approccio.

2.1 *L'Impatto del Trucco sul Riconoscimento Facciale*

Il trucco è una delle variabili più significative che possono alterare l'aspetto del viso e, di conseguenza, influire sulle prestazioni dei sistemi di riconoscimento facciale. Sebbene i sistemi avanzati di riconoscimento facciale siano progettati per essere robusti di fronte a variazioni nelle condizioni di illuminazione, angolazione e espressioni facciali, il trucco rappresenta una sfida unica a causa della sua capacità di alterare profondamente le caratteristiche visive del viso.

- **Regione degli occhi:** Elementi come ombretti, eyeliner e ciglia finte possono mascherare o modificare la forma naturale degli occhi, che rappresentano una delle caratteristiche più distintive del viso.
- **Regione delle labbra:** Prodotti come rossetto e gloss introducono variazioni significative nei contrasti cromatici e nella texture delle labbra, rendendo più complessa la segmentazione di quest'area.
- **Aspetto globale della pelle:** Fondotinta e correttori uniformano la tonalità della pelle, mascherando imperfezioni e modificando la texture naturale.

Queste alterazioni hanno quindi un impatto che varia a seconda della complessità e tipologia di algoritmi di riconoscimento facciale adottati. I metodi tradizionali, come LBP (Local Binary Patterns) e HOG (Histogram of Oriented Gradients), si basano su caratteristiche visive statiche e non adattative. Di conseguenza, questi metodi soffrono una significativa perdita di accuratezza in presenza di trucco marcato. Al contrario, gli algoritmi di deep learning, come FaceNet o ArcFace, sono più robusti grazie alla loro capacità di apprendere rappresentazioni astratte e invarianti delle caratteristiche facciali. Tuttavia, nonostante i progressi tecnologici, le performance di questi sistemi possono ancora essere compromesse in presenza di modifiche estreme causate dal trucco. Un aspetto interessante riguarda l'uso deliberato del trucco per eludere i sistemi di riconoscimento facciale. Tecniche come l'applicazione strategica di eyeliner e ombretti, o il contouring pesante, vengono talvolta utilizzate per oscurare o distorcere landmark chiave come gli angoli degli occhi o la forma del viso. Questi approcci mettono in evidenza i limiti di

molti algoritmi, anche quelli basati su reti neurali avanzate, nel distinguere un volto reale da uno modificato intenzionalmente. Tuttavia, è importante sottolineare che, nonostante le difficoltà introdotte dal trucco, i sistemi di riconoscimento facciale più avanzati riescono comunque a gestire variazioni moderate, grazie all'utilizzo di dataset diversificati e alla progettazione di architetture più robuste.

2.2 Algoritmi di Localizzazione dei Landmark e Estrazione delle ROI

La localizzazione dei landmark e l'estrazione delle regioni di interesse (ROI) sono due fasi fondamentali nei sistemi di elaborazione dei volti. Questi processi rappresentano il punto di partenza per molte applicazioni, come il riconoscimento facciale, l'analisi delle emozioni e lo studio dell'impatto estetico del trucco. In questo capitolo esploreremo i principali algoritmi utilizzati per individuare i landmark e definire accuratamente le ROI.

Localizzazione dei landmark

I landmark facciali sono punti chiave che identificano le caratteristiche geometriche del volto, come occhi, naso, bocca e il contorno del viso. Tra gli approcci più comuni si trovano:

- **Modelli basati su cascata di Haar:** I primi approcci alla localizzazione dei landmark si basavano su tecniche come il classico algoritmo Haar-Cascade. Questo metodo utilizza una cascata di classificatori addestrati su feature Haar per individuare regioni del viso. Questo metodo soffre in condizioni di illuminazione variabile o in presenza di volti non frontali.
- **CNN:** Con l'avvento del deep learning, i metodi basati su CNN hanno rivoluzionato la localizzazione dei landmark. Architetture come Face Alignment Networks (FAN) e Deep Alignment Networks (DAN) utilizzano reti neurali profonde per mappare con precisione i punti chiave del volto. Questi approcci riescono a gestire pose estreme, illuminazioni variabili e occlusioni, rendendoli ideali per contesti reali.
- **MediaPipe Face Mesh:** MediaPipe Face Mesh è una soluzione moderna che utilizza pipeline basate su modelli neurali per individuare fino a 468 landmark 3D del viso. Questo approccio è particolarmente utile per applicazioni avanzate come realtà aumentata, poiché consente di ottenere una mappatura completa e dettagliata del volto in tempo reale.

Estrazione delle regioni di interesse

Una volta localizzati i landmark, l'estrazione delle ROI consente di isolare specifiche aree del viso, come occhi, bocca e contorno facciale. Questo processo è cruciale per focalizzarsi su dettagli rilevanti, riducendo il rumore introdotto da parti non significative del viso.

- **Definizione delle ROI tramite bounding box:** La tecnica più comune per estrarre le ROI consiste nell'utilizzare bounding box definite intorno ai landmark chiave. Ad esempio, gli occhi possono essere racchiusi in box centrati sui punti chiave corrispondenti agli angoli degli occhi, mentre la bocca può essere definita utilizzando i landmark delle labbra. Questo approccio è rapido e flessibile, ma richiede un'accurata definizione dei margini per evitare di includere dettagli non rilevanti.
- **Tecniche basate su deep learning:** Gli algoritmi moderni possono apprendere direttamente come segmentare le ROI utilizzando reti neurali convoluzionali. Questo approccio elimina la necessità di definire manualmente i margini e consente al modello di adattarsi a diverse variazioni del viso, come pose o illuminazioni non standard.

2.3 Algoritmi di Estrazione delle Feature

L'estrazione delle feature è uno dei passaggi fondamentali nei sistemi di riconoscimento facciale, poiché consente di convertire dati grezzi, come le immagini, in rappresentazioni numeriche utili per i classificatori. La scelta delle feature influenza direttamente le prestazioni del sistema, soprattutto in contesti complessi come il riconoscimento di volti truccati, dove il trucco introduce variabilità non naturale.

I metodi di estrazione delle feature possono essere suddivisi in due categorie principali: approcci tradizionali basati su texture, colore e forma, e tecniche moderne che sfruttano reti neurali profonde. Entrambe le categorie hanno i loro vantaggi, ma la loro efficacia dipende fortemente dal problema specifico e dal tipo di dati utilizzati.

Feature Basate su Texture

Le texture del viso rappresentano un aspetto cruciale nella discriminazione tra volti truccati e non truccati, poiché il trucco modifica frequentemente la struttura visiva delle regioni facciali. Tra i metodi più utilizzati, troviamo i filtri di Gabor, i Local Binary Patterns (LBP) e gli Histogram of Oriented Gradients (HOG).

I filtri di Gabor sono progettati per rilevare bordi e texture a diverse scale e orientamenti, rendendoli particolarmente adatti per analizzare dettagli locali come ciglia, ombretti e rughe. Gli LBP, invece, analizzano la texture confrontando ciascun pixel con i suoi vicini e trasformando i risultati in modelli binari. Questo metodo si è dimostrato robusto rispetto a variazioni di illuminazione, motivo per cui è spesso impiegato in applicazioni biometriche. Gli HOG rappresentano invece le immagini in termini di istogrammi di gradienti orientati, fornendo una descrizione accurata della struttura e dei contorni facciali. Questi metodi, benché tradizionali, offrono una base solida per il riconoscimento di volti con variazioni estetiche moderate.

Feature Basate su Colore

Il colore è un altro aspetto fondamentale, soprattutto per il riconoscimento di volti truccati, poiché il trucco introduce cambiamenti significativi nella distribuzione cromatica del viso. Lo spazio di colore HSV, che separa tonalità, saturazione e valore, è particolarmente utile per catturare variazioni introdotte da prodotti come rossetti e ombretti. Ad esempio, analizzando gli istogrammi di tonalità all'interno delle ROI (regioni di interesse), è possibile individuare variazioni di colore che distinguono volti con trucco da quelli senza. Inoltre, la combinazione di spazi di colore diversi (RGB e HSV) consente di rappresentare sia la componente cromatica che luminosa, migliorando l'efficacia complessiva delle feature.

Feature Basate su Forma

Le caratteristiche geometriche del viso giocano un ruolo complementare nell'identificazione di volti truccati. Tecniche come gli Edge Orientation Histograms (EOH) analizzano l'orientamento dei bordi per descrivere la forma di regioni specifiche, come labbra e sopracciglia. Questi metodi sono particolarmente efficaci per rilevare modifiche introdotte da tecniche di contouring o dall'uso di prodotti che alterano visivamente i tratti del viso. Inoltre, i landmark facciali possono essere utilizzati per calcolare distanze e proporzioni tra le diverse parti del volto, catturando cambiamenti strutturali introdotti dal trucco.

Ad esempio, un volto truccato con contouring potrebbe presentare una distribuzione dei bordi e delle proporzioni diversa rispetto a un volto naturale, influenzando l'identificazione della mascella e degli zigomi.

2.4 Algoritmi di Classificazione

La fase di classificazione rappresenta il cuore del sistema di riconoscimento facciale, poiché utilizza le feature estratte per distinguere tra le diverse classi di interesse, come

volti con trucco e volti senza trucco. Gli algoritmi di classificazione possono essere suddivisi in due categorie principali: approcci tradizionali e metodi basati su deep learning.

Classificatori Tradizionali

I metodi tradizionali sono ampiamente utilizzati per problemi di classificazione binaria e sono noti per la loro semplicità e robustezza in scenari con un numero moderato di feature.

Uno dei classificatori più comuni è il **Support Vector Machine (SVM)**, che cerca di massimizzare il margine tra le classi utilizzando un iperpiano di separazione. Gli SVM sono particolarmente efficaci in problemi di classificazione binaria ad alta dimensionalità, come il riconoscimento facciale. La variante kernel SVM, con kernel come Radial Basis Function (RBF), consente di gestire problemi non linearmente separabili, offrendo flessibilità nella rappresentazione dei dati. Tuttavia, gli SVM richiedono un'attenta selezione degli iperparametri, come il parametro C e il tipo di kernel, per ottenere prestazioni ottimali.

Un altro approccio popolare è **Random Forest**, un metodo ensemble che combina più alberi decisionali per migliorare l'accuratezza e ridurre l'overfitting. Ogni albero viene addestrato su un sottoinsieme dei dati, e le predizioni vengono aggregate tramite voto maggioritario. Questo metodo è particolarmente utile per dataset sbilanciati o con feature eterogenee.

Infine, **AdaBoost (Adaptive Boosting)** è un algoritmo che costruisce un forte classificatore combinando più classificatori deboli, come alberi decisionali poco profondi. AdaBoost assegna pesi maggiori ai campioni classificati erroneamente, consentendo al sistema di concentrarsi su esempi difficili. Questo approccio è spesso utilizzato in combinazione con feature tradizionali, come LBP o Gabor.

Classificatori Basati su Deep Learning Con l'avvento delle reti neurali profonde, i classificatori basati su deep learning sono diventati lo stato dell'arte in molti problemi di classificazione, incluso il riconoscimento facciale. I modelli come le **Convolutional Neural Networks (CNN)** sono progettati per apprendere automaticamente feature rilevanti dai dati grezzi, eliminando la necessità di progettare manualmente le feature.

Modelli pre-addestrati come **VGGFace** e **ResNet** possono essere utilizzati per estrarre rappresentazioni ad alta discriminatività del viso. Questi modelli, addestrati su dataset di grandi dimensioni, come CelebA o LFW, possono essere riutilizzati per compiti specifici tramite tecniche di transfer learning. Inoltre, le CNN possono essere combinate con metodi di pooling e normalizzazione per migliorare la robustezza alle variazioni estetiche, come il trucco pesante.

Un'altra tecnica avanzata è **Deep Metric Learning**, utilizzata da modelli come FaceNet e ArcFace. Questi modelli apprendono uno spazio embedding in cui le immagini della stessa identità sono vicine, mentre quelle di identità diverse sono lontane. Sebbene non siano classificatori nel senso tradizionale, questi metodi possono essere combinati con SVM o reti Fully Connected per risolvere problemi di classificazione specifici.

2.5 Dataset esistenti per il riconoscimento e classificazione visi senza trucco e con

I dataset sono fondamentali per l'addestramento, la validazione e il test dei sistemi di riconoscimento facciale. La qualità e la diversità dei dati influiscono direttamente sulle prestazioni dei modelli. In questo capitolo, vengono presentati i principali dataset utilizzati per analizzare l'impatto del trucco sui sistemi di riconoscimento facciale.

CelebA (CelebFaces Attributes Dataset)

CelebA è uno dei dataset più utilizzati nella ricerca sul riconoscimento facciale. Contiene oltre 200.000 immagini di celebrità, ciascuna annotata con 40 attributi binari, tra cui "Heavy Makeup" e "Wearing Lipstick". Questo dataset è particolarmente utile per studiare l'impatto del trucco, grazie alla varietà di volti, pose ed espressioni. Tuttavia, presenta uno sbilanciamento tra classi, che può influire negativamente sull'addestramento di modelli sensibili alla distribuzione dei dati.

Makeup in the Wild (MIW)

Il dataset MIW è progettato specificamente per l'analisi del trucco facciale. Include immagini catturate in ambienti naturali, con annotazioni dettagliate sul tipo di trucco applicato (occhi, labbra, viso). La varietà di condizioni di illuminazione e angolazione rende questo dataset adatto per sviluppare modelli robusti.

Chicago Face Database (CFD)

Il CFD è un dataset ampiamente utilizzato per studi psicologici e biometrici. Contiene immagini standardizzate di volti maschili e femminili, con e senza trucco. La standardizzazione delle condizioni di acquisizione riduce le variabili non controllate, consentendo un'analisi più precisa degli effetti del trucco.

Nonostante la disponibilità di dataset ricchi di annotazioni, esistono ancora sfide significative:

- **Sbilanciamento delle Classi:** I dataset tendono ad avere più immagini di volti non truccati rispetto a quelli truccati, influenzando negativamente i modelli di classificazione.

- **Diversità Limitata:** Molti dataset non rappresentano adeguatamente diverse etnie, generi e fasce d'età.
- **Annotazioni Incomplete:** Sebbene dataset come CelebA forniscano annotazioni dettagliate, non tutti includono informazioni specifiche sul tipo di trucco applicato.

3 Approccio Sviluppato

L'approccio sviluppato in questo progetto si concentra sulla progettazione e implementazione di un sistema automatizzato per la rilevazione del trucco facciale, con un focus particolare sull'impatto di tale trucco su sistemi di riconoscimento facciale. La pipeline proposta integra tecniche avanzate di elaborazione delle immagini, localizzazione dei landmark e classificazione. L'obiettivo principale è quello di creare un sistema accurato, robusto e adattabile a diversi scenari, sfruttando informazioni provenienti da diverse regioni facciali (occhi, bocca e viso) e combinando tali informazioni per migliorare le prestazioni del classificatore.

Colorspace: RGB vs HSV

Un aspetto fondamentale nella progettazione del sistema è la scelta del color space per la rappresentazione delle immagini. In questo progetto, è stato deciso di utilizzare lo spazio dei colori HSV (Hue, Saturation, Value) anziché RGB (Red, Green, Blue). Di seguito, analizziamo le differenze sostanziali tra i due spazi dei colori e i motivi di questa scelta.

Il colorspace RGB rappresenta ogni pixel come una combinazione di tre componenti: rosso (R), verde (G) e blu (B). Questo modello è ampiamente utilizzato in display elettronici e fotocamere. Tuttavia, presenta alcune limitazioni:

- Dipendenza dalla luminosità: Le componenti RGB sono fortemente influenzate dalla luminosità dell'immagine, rendendolo poco robusto a variazioni di illuminazione.
- Mancanza di separazione tra colore e intensità: Le informazioni cromatiche (colore) e di intensità (luminosità) sono combinate, rendendo difficile analizzare il colore indipendentemente dalla luminosità.

Il colorspace HSV separa il colore in tre componenti:

- Hue (H): Tonalità del colore, rappresenta la lunghezza d'onda dominante.
- Saturation (S): Intensità cromatica, misura la purezza del colore.
- Value (V): Luminosità o brillantezza del colore.

Come si può vedere dalla figura 3.1 le informazioni riguardanti il trucco possono essere estratte al meglio nel canale S del colorspace HSV.

Questo spazio è particolarmente utile per analizzare immagini con variazioni di illuminazione, poiché separa la componente di colore (H) dalla componente di intensità (V). Matematicamente, i valori HSV sono derivati dalle componenti RGB attraverso trasformazioni non lineari:

$$H = 60 \times \begin{cases} 0 + \frac{G-B}{\max-\min}, & \text{se } \max = R \\ 2 + \frac{B-R}{\max-\min}, & \text{se } \max = G \\ 4 + \frac{R-G}{\max-\min}, & \text{se } \max = B \end{cases}$$

$$S = \frac{\max(R, G, B) - \min(R, G, B)}{\max(R, G, B)}$$

$$V = \max(R, G, B)$$

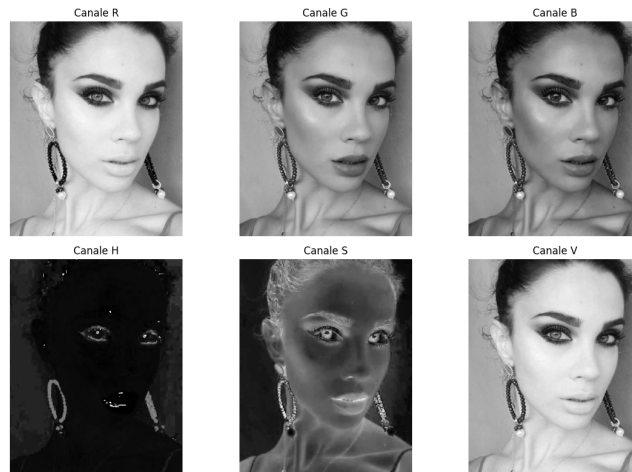


Figure 3.1: Visualizzazione di un viso nei corrispondenti canali dei colorspace RGB(sopra) e HSV(sotto)

La pipeline proposta include una serie di fasi fondamentali. Si parte dalla localizzazione dei landmark facciali per individuare i punti chiave del volto. Questi punti sono poi utilizzati per eseguire una normalizzazione geometrica delle immagini tramite trasformazioni affini, al fine di ridurre le variazioni dovute a pose e dimensioni. Successivamente, vengono isolate le regioni di interesse, ovvero gli occhi, la bocca e l'intero viso, utilizzando i landmark localizzati. Queste regioni vengono poi preprocessate e ridimensionate per garantire uniformità nei successivi passaggi.

Una volta estratte le ROI, si procede con l'estrazione delle feature. Sono stati considerati metodi per catturare informazioni relative a colore, texture e forma, che sono poi utilizzate per l'addestramento e la validazione di classificatori dedicati. La pipeline è stata progettata in modo flessibile, consentendo sia l'analisi individuale delle ROI sia la combinazione delle informazioni provenienti da più regioni attraverso tecniche di ensemble learning.

In questo capitolo, analizzeremo nel dettaglio ogni passaggio della pipeline, partendo dalla localizzazione dei landmark, fino ad arrivare all'implementazione della classificazione finale. Ogni componente sarà descritta nei suoi aspetti teorici e implementativi, evidenziando il contributo al raggiungimento dell'obiettivo generale.

3.1 Rilevamento del viso e localizzazione dei landmark

Il rilevamento dei volti è stata implementata utilizzando il modulo MediaPipe Face Mesh. Questo strumento non solo identifica la presenza di un volto all'interno dell'immagine, ma fornisce anche una mappa dettagliata di punti di riferimento che descrivono la geometria del volto stesso. MediaPipe Face Mesh è particolarmente adatto per applicazioni come questa grazie alla sua capacità di rilevare fino a 468 punti di riferimento in modo accurato.

L'algoritmo è stato configurato per operare in modalità statica, ottimizzando così il processo per immagini individuali invece che per flussi video. I volti vengono rilevati anche in condizioni di illuminazione variabili e con diverse pose, grazie alla robustezza del modello pre-addestrato.

Il rilevamento del volto in MediaPipe si basa su un modello convoluzionale profondo (CNN), che genera una mappa di calore per individuare la posizione del volto nell'immagine. L'immagine $I(x, y, c)$, con $c \in \{R, G, B\}$, viene normalizzata a una dimensione $W \times H$. L'immagine viene poi processata attraverso una CNN che genererà una mappa di calore che evidenzierà le regioni con alta probabilità di contenere un volto. Viene applicata una funzione di soglia τ :

$$M(x,y) = \begin{cases} 1, & \text{se } P(x,y) \geq \tau \\ 0, & \text{altrimenti} \end{cases}$$

Una volta rilevato il volto, MediaPipe utilizza una rete neurale per identificare i landmark specifici. Le coordinate del bounding box del volto rilevato vengono scalate in un intervallo normalizzato $[0, 1]$. Una rete di regressione prevede la posizione dei landmark $L = \{(x_i, y_i)\}_{i=1}^{468}$ minimizzando una funzione di perdita L_{MSE} :

$$L_{MSE} = \frac{1}{N} \sum_{i=1}^N |L'_i - L_i|^2$$

dove L'_i sono le posizioni predette e L_i quelle reali. Nella figura 3.2 si possono vedere i landmark identificati per le roi di interesse.



Figure 3.2: Visualizzazione dei landmark di interesse

Dopo il rilevamento dei volti e la localizzazione dei landmark, normalizziamo geometricamente le immagini dei volti utilizzando una trasformazione affine per rimuovere le variazioni dovute alla scala e posa. Tutte le immagini dei volti normalizzate vengono ritagliate e ridimensionate ad una dimensione di 150×130 pixel, così anche: ROI occhio sinistro: 52×52 ; ROI occhio destro: 52×52 ; ROI della bocca: 56×62 .

3.2 Estrazione delle feature

Le caratteristiche proposte per il rilevamento del trucco si basano su descrittori di forma, texture e colore. La scelta delle feature è stata guidata da alcune osservazioni chiave. Visivamente, il trucco ha un impatto dominante sugli attributi cromatici del volto. Per questo motivo, vengono impiegate feature basate sul colore, che consentono di catturare le variazioni cromatiche introdotte dal trucco. Inoltre, il trucco può modificare la forma locale e le informazioni di texture del volto. Per rappresentare queste variazioni, vengono utilizzati filtri di Gabor, che sono in grado di estrarre informazioni legate alla forma e alla texture attraverso diverse scale spaziali e orientamenti. Un'altra osservazione rilevante riguarda la capacità del trucco di alterare piccoli dettagli del viso, come i micro-patterns. Per rappresentare tali modifiche, viene applicato l'operatore Local Binary Pattern (LBP), particolarmente efficace nell'analisi delle texture locali.

Nelle sezioni successive, verranno illustrate nel dettaglio le tecniche di estrazione delle feature.

3.2.1 Feature basate sul colore

Le informazioni di colore sono particolarmente rilevanti per il riconoscimento del trucco, poiché prodotti cosmetici come fondotinta, rossetti e ombretti alterano significativamente la distribuzione cromatica del viso.

Per rappresentare il colore, è stato utilizzato lo spazio HSV (Hue, Saturation, Value), che separa la tonalità, la saturazione e la luminosità. Per estrarre le feature colore ho suddiviso ogni ROI in blocchi 5×5 e, per ciascun blocco, sono calcolate le seguenti statistiche:

- Media(μ): Rappresenta il valore medio dei pixel nel blocco.
- Deviazione standard(σ): Misura la dispersione dei valori rispetto alla media.
- Asimmetria(skewness): Quantifica la simmetria della distribuzione dei pixel.

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i, \quad \sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}, \quad \text{skewness} = \frac{1}{N} \sum_{i=1}^N \frac{(x_i - \mu)^3}{\sigma^3 + \varepsilon}$$

dove x_i è l'intensità di un pixel e N è il numero totale di pixel.

Per estrarre queste informazioni dall'intero viso invece l'immagine è stata suddivisa in blocchi 3×3

I grafici in figura 3.3 evidenziano differenze significative tra occhi truccati e non truccati nel canale di saturazione. La media è più alta negli occhi truccati, il che indica una maggiore intensità cromatica causata dall'uso di cosmetici che intensificano i colori. La varianza è più elevata nei truccati, segnalando una maggiore diversità cromatica dovuta all'applicazione di ombretti e eyeliner. L'asimmetria (skewness) è più accentuata negli occhi truccati, suggerendo una concentrazione dei valori di saturazione verso toni più intensi, mentre negli occhi non truccati i valori risultano più uniformi. Questi risultati dimostrano come i descrittori basati sul colore possano distinguere in modo efficace le caratteristiche cromatiche introdotte dal trucco.

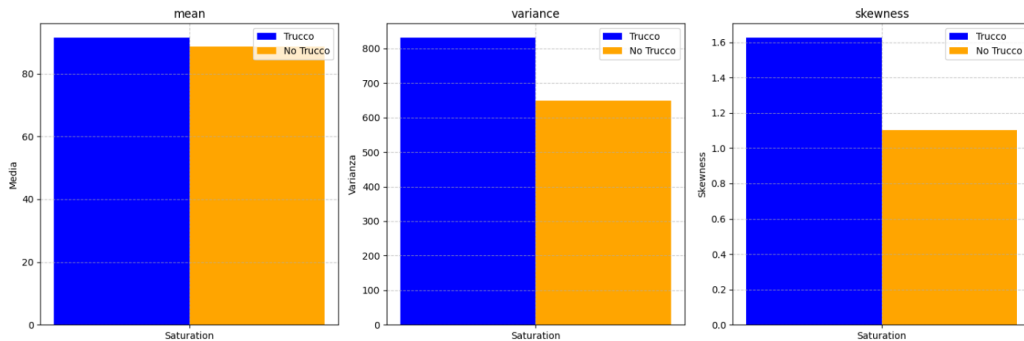


Figure 3.3: Visualizzazione di μ , σ e skewness nelle due tipologie di ROI.

3.2.2 Feature basate sulla forma

L'approccio prevede l'utilizzo di descrittori di forma, tra cui Gabor Filters, GIST e EOH (Edge Orientation Histogram). Questi metodi sono fondamentali per catturare le caratteristiche strutturali delle regioni del viso, permettendo una rappresentazione robusta di elementi come contorni, bordi e orientamenti che possono essere influenzati dal trucco. Di seguito verrà descritto il funzionamento di ciascun algoritmo e la loro utilità per il problema affrontato.

1. Filtri di Gabor

I filtri di Gabor sono utilizzati per analizzare i contorni e le frequenze spaziali in immagini. Sono particolarmente utili per identificare pattern locali, come quelli generati dall'ombretto o dal trucco sugli occhi.

$$g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos\left(2\pi\frac{x'}{\lambda} + \psi\right)$$

dove:

- $x' = x \cos \theta + y \sin \theta$
- $y' = -x \sin \theta + y \cos \theta$
- λ è la lunghezza d'onda,
- θ è l'orientamento,
- ψ è la fase,
- σ è la deviazione standard,
- γ è il rapporto di aspetto.

I filtri vengono applicati in diverse direzioni e scale per estrarre caratteristiche come media, varianza e asimmetria dalle immagini filtrate. Questo consente di catturare informazioni multi-scala e multi-orientazione, fondamentali per descrivere i dettagli delle forme locali del viso. Ho utilizzato una dimensione di 64×64 per ogni Gabor Kernel, e per ogni immagine risultante dall'applicazione dei filtri ho calcolato i valori $\mu, \rho, \text{skewness}$. Nella figura 3.4 possiamo notare l'applicazione di questi filtri su diverse direzioni e scale sulle due tipologie di volti e si può notare come nel viso truccato, le regioni intorno agli occhi e alla bocca appaiono più marcate, con contrasti più evidenti dovuti alla presenza di trucco come eyeliner o rossetto. Al contrario, nei visi non truccati, queste stesse aree risultano meno accentuate, con contorni più sfumati e uniformi. Le differenze sono particolarmente visibili nelle zone chiave del viso, mentre altre aree, come la fronte e le guance, mostrano minime variazioni.

2. Edge Orientation Histogram (EOH)

Il descrittore EOH si basa sull'analisi degli angoli di orientamento dei bordi. Questo approccio è efficace per rilevare contorni e spigoli creati da trucchi come eyeliner o matita per sopracciglia. L'algoritmo calcola i gradienti G_x e G_y usando filtri di Sobel, e successivamente converte i gradienti in ampiezza e orientazione:

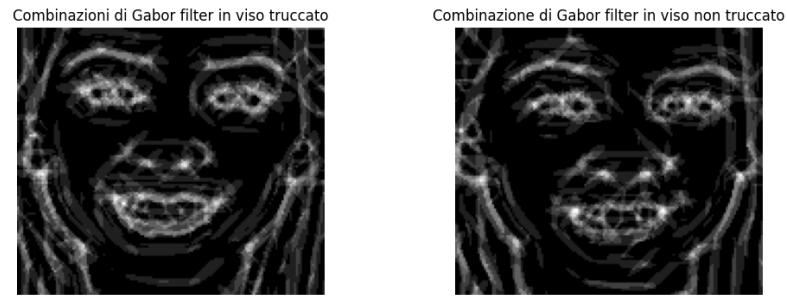


Figure 3.4: Applicazione dei filtri di Gabor sulle due tipologie di visi.

$$\text{magnitude} = \sqrt{G_x^2 + G_y^2}, \quad \text{angle} = \arctan\left(\frac{G_y}{G_x}\right)$$

Successivamente, le orientazioni vengono discretizzate in un istogramma normalizzato. Ogni bin dell'istogramma rappresenta la densità di bordi in una specifica direzione, offrendo un quadro dettagliato dei pattern geometrici nel viso.

Come si può notare nella figura 3.5 ci sono differenze significative tra l'applicazione degli Edge Orientation Histogram (EOH) in visi truccati e non truccati. Nei visi truccati, i contorni risultano molto più marcati, specialmente nelle regioni degli occhi e delle labbra, dove il trucco accentua le linee e le forme. Questo si traduce in immagini dei contorni più nitide e ricche di dettagli. Inoltre, gli istogrammi mostrano picchi più pronunciati, indicando una maggiore presenza di bordi con orientazioni specifiche, che riflettono l'intensità e la precisione introdotte dal trucco. Nei visi non truccati, invece, i contorni appaiono più morbidi e meno definiti. Le linee delle ROI sono meno evidenti e gli istogrammi risultano più uniformi, senza picchi dominanti, riflettendo una minore complessità nei dettagli strutturali. In sintesi, l'uso di EOH evidenzia come il trucco modifichi le caratteristiche visive del viso, enfatizzando i dettagli e introducendo variazioni significative nei bordi, che possono essere sfruttate per la classificazione.

3. GIST

Gist è un descrittore utilizzato per catturare le informazioni globali di una scena o immagine, sintetizzandone la struttura principale in modo compatto. Questo metodo analizza l'immagine suddividendola in regioni spaziali e applicando filtri Gabor con diverse frequenze e orientamenti. L'obiettivo è rappresentare l'immagine attraverso una firma numerica che descriva caratteristiche come l'orientamento, la frequenza spaziale e la distribuzione delle texture.

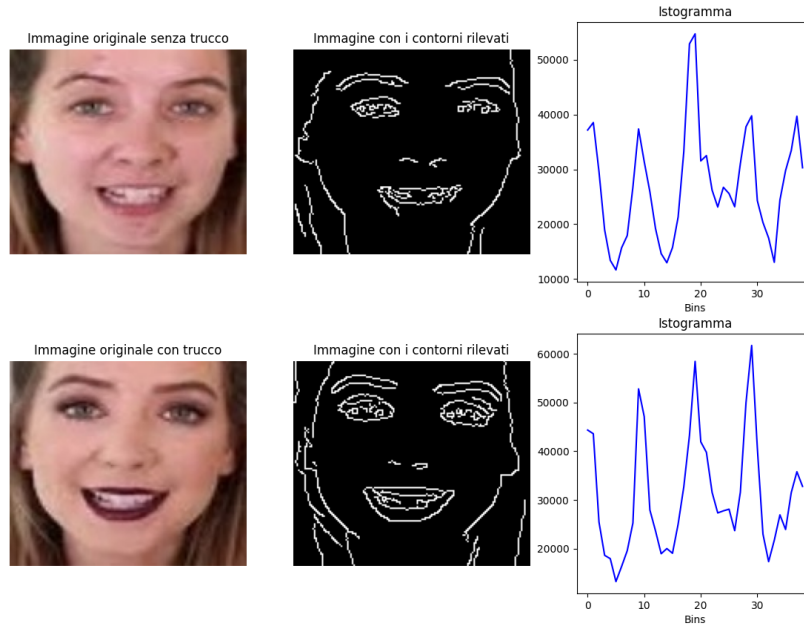


Figure 3.5: Applicazione di EOH sulle due tipologie di visi.

Nel contesto del nostro problema, Gist è particolarmente utile perché riesce a catturare le variazioni globali nella struttura del viso dovute al trucco. Con il suo approccio globale, Gist fornisce una rappresentazione che riflette queste variazioni, contribuendo così a migliorare l'accuratezza del classificatore.

3.2.3 Feature basate su texture

Le Local Binary Patterns (LBP) rappresentano uno dei metodi più utilizzati per l'analisi delle texture nelle immagini grazie alla loro semplicità ed efficacia. Questo metodo si basa su un confronto locale dei valori di intensità di pixel per estrarre informazioni sui pattern di texture. Nel nostro contesto, l'uso di LBP risulta particolarmente utile per analizzare le regioni del viso truccate e non truccate, poiché il trucco può modificare in modo significativo le microstrutture superficiali della pelle, generando pattern caratteristici che l'algoritmo è in grado di catturare.

Il funzionamento di LBP si basa su un confronto tra il valore di intensità di un pixel centrale I_c e i valori dei pixel circostanti in una finestra locale di dimensione prefissata. Per ogni pixel circostante I_p , il confronto produce un valore binario:

$$LBP_p = \begin{cases} 1, & \text{se } I_p \geq I_c, \\ 0, & \text{altrimenti} \end{cases}$$

I valori binari risultanti vengono poi combinati in un numero decimale tramite una somma pesata, che definisce il valore LBP associato al pixel centrale:

$$LBP = \sum_{p=0}^{P-1} s(I_p - I_c) \times 2^p,$$

dove P è il numero di pixel circostanti e $s(x)$ è la funzione di segno, definita come:

$$s(x) = \begin{cases} 1, & \text{se } x \geq 0, \\ 0, & \text{altrimenti} \end{cases}$$

Questo processo viene ripetuto per ogni pixel dell'immagine, generando una mappa LBP che rappresenta i pattern locali di texture.

Nel contesto della rilevazione del trucco, LBP è utile perché il trucco tende a modificare la texture della pelle introducendo variazioni nei pattern locali. Ad esempio, il trucco per gli occhi o il rossetto possono generare microstrutture regolari che vengono catturate dalla rappresentazione LBP. Inoltre, LBP è resistente alle variazioni di illuminazione, un aspetto cruciale quando si analizzano immagini raccolte in condizioni di luce diverse.

L'insieme totale delle feature estratte dalle ROI e dal viso è definito nella tabella 3.1.

Attributi	Feature	Dim-Viso	Dim-ROI
Colore	Moments	81	225×3
Forma	Gabor	120	-
	GIST	512	-
	EOH	37	-
Texture	LBP	59	-
Totale		809	675

Table 3.1: Dimensionalità del vettore delle features

3.3 Classificazione

Per il compito di classificazione delle immagini in base alla presenza o assenza di trucco, abbiamo utilizzato approcci basati su Support Vector Machines (SVM), AdaBoost e tec-

niche di ensemble learning. Inoltre, è stato implementato un processo di selezione delle feature tramite Random Forest per migliorare le prestazioni del classificatore, riducendo il rumore e mantenendo solo le informazioni rilevanti. Di seguito, descriviamo nel dettaglio ciascun approccio.

3.3.1 Support Vector Machine(SVM)

L'algoritmo SVM è particolarmente efficace per problemi di classificazione binaria. L'obiettivo principale di un SVM è trovare un iperpiano $h(x) = w^T x + b = 0$ che separi i dati di due classi nel modo più ampio possibile. La distanza tra l'iperpiano e i punti più vicini di ciascuna classe, chiamata margine, viene massimizzata.

Il problema di ottimizzazione da risolvere può essere formulato come:

$$\min_{w,b} \frac{1}{2} ||w||^2$$

soggetto a:

$$y_i(w^T x_i + b) \geq 1, \forall i.$$

Dove:

- x_i rappresenta i dati di input.
- $y_i \in \{-1, 1\}$,
- w è il vettore dei pesi,
- b è il termine di bias.

Siccome i dati non sono linearmente separabili, si è utilizzato una funzione di kernel $K(x_i, x_j)$ RBF, che consente di modellare decisioni non lineari. I valori ottimali per C e i parametri del kernel sono ottenuti da una grid search dello spazio dei parametri in base al set di addestramento, e dopo una serie di prove sono stati definiti i seguenti parametri: $C = 8$, kernel = 'rbf', $\gamma = 0.001$.

3.3.2 Adaboost

AdaBoost è stato scelto come secondo classificatore. Questo algoritmo si basa sull'idea di combinare deboli classificatori, come gli alberi decisionali con profondità ridotta, per

creare un classificatore forte. Ad ogni iterazione, AdaBoost assegna pesi maggiori ai campioni mal classificati, consentendo al modello di concentrarsi su di essi nelle iterazioni successive. La funzione di classificazione risultante è data da:

$$H(x) = \sum_{t=1}^T \alpha_t h_t(x)$$

dove:

- $h_t(x)$ è il classificatore debole alla t -esima iterazione.
- α_t è il peso assegnato a $h_t(x)$ in base alla sua accuratezza.
- T è il numero di classificatori deboli adoperati

Per adaboost sono stati usati classificatori deboli di tipo Albero Decisionale con una max depth pari a 1. Per il numero di classificatori deboli mi sono basato sul numero di feature selezionate d_{eff} e il numero di campioni N :

$$T = \min(10 \times d_{eff}, \frac{N}{\log(N)}) \approx 1000$$

Nel nostro approccio, la combinazione delle informazioni provenienti da più ROI, come occhi + bocca o occhi + viso, è stata implementata utilizzando un meccanismo di votazione soft. In questo metodo, i classificatori allenati su ciascuna ROI non forniscono semplicemente una predizione binaria, ma restituiscono una probabilità associata a ciascuna classe. Questo approccio consente di integrare in modo più raffinato le informazioni provenienti dalle diverse ROI, sfruttando il grado di confidenza di ciascun classificatore.

Nella votazione soft, l'output finale del classificatore ensemble viene calcolato aggregando le probabilità predette da ciascun classificatore ROI. Supponiamo di avere M ROI e c classi. La probabilità assegnata alla classe y è data da:

$$P(y = c|x) = \sum_{m=1}^M w_m P_m(y = c|x)$$

dove:

- $P_m(y = c|x)$ è la probabilità predetta dal classificatore allenato sulla ROI m per la classe c .

- w_m è il peso assegnato alla ROI m , che riflette la sua rilevanza per il problema.
- M è il numero totale di ROI

La classe finale assegnata al campione è quella con la probabilità più alta.

4 Valutazione Sperimentale

In questa sezione vengono presentate le valutazioni sperimentali condotte per verificare l'efficacia dell'approccio proposto. Gli esperimenti sono stati progettati per analizzare le prestazioni del sistema di classificazione nel distinguere immagini di visi truccati da quelli non truccati, utilizzando le tecniche descritte nei capitoli precedenti. Il processo comprende l'uso di un dataset eterogeneo, la definizione di un protocollo di training e testing rigoroso, e l'adozione di metriche di valutazione standard per confrontare le prestazioni dei modelli sviluppati.

4.1 Dataset

Il dataset utilizzato negli esperimenti è composto da immagini di 698 soggetti truccati e 628 soggetti non truccati, per un totale di 1326 immagini. Come si può vedere nella figura 4.1 le immagini includono volti catturati con differenti pose e condizioni di illuminazione, il che rende il dataset variegato e rappresentativo di scenari realistici. Nonostante queste variazioni, tutte le immagini hanno in comune una caratteristica fondamentale: il viso è chiaramente visibile e messo in risalto, consentendo una facile identificazione delle regioni di interesse (ROI) come occhi, bocca e l'intero viso.

Ogni immagine del dataset è stata etichettata manualmente per distinguere tra soggetti truccati e non truccati, garantendo l'accuratezza delle informazioni di classificazione. La diversità delle pose e delle illuminazioni introduce una sfida significativa per il modello, in quanto richiede un'alta capacità di generalizzazione per distinguere le caratteristiche intrinseche del trucco indipendentemente dai fattori esterni.

Per la fase di preprocessing, le immagini sono state normalizzate e ridimensionate a una risoluzione uniforme, preservando i dettagli rilevanti per le ROI. Questo passaggio è stato fondamentale per garantire un'analisi consistente delle regioni chiave durante l'estrazione delle feature.

L'utilizzo di questo dataset è cruciale per valutare le capacità del modello di gestire variabilità realistiche, come quelle introdotte dall'illuminazione o dalle pose non frontali. Inoltre, il bilanciamento relativo tra classi (698 truccati e 628 non truccati) riduce il rischio di bias nei risultati, permettendo di misurare l'accuratezza e la robustezza del modello in modo affidabile.



Figure 4.1: Esempio di visi truccati (sopra) e non truccati (sotto).

4.2 Sperimentazione

Per valutare l'efficacia del sistema di classificazione per è stato definito un rigoroso protocollo di training e testing. Questo processo garantisce che i risultati siano affidabili e che il modello sia in grado di generalizzare su dati non visti.

Il dataset, composto da 1325 immagini, è stato suddiviso in modo stratificato per preservare la proporzione tra classi (truccati e non truccati):

- 80% del dataset è stato utilizzato per il training.
- 20% del dataset è stato utilizzato per il testing.

Per addestrare i classificatori, ho utilizzato tecniche di validazione incrociata (cross-validation) con 5 fold. In ogni iterazione, il dataset di training è stato suddiviso in 5 parti:

- 4 fold per l'apprendimento.
- 1 fold per la validazione

Le immagini usate nella cross-validation non sono presenti nel testing set. Per il problema di classificazione binaria affrontato ho adottato come metriche per la valutazione delle prestazioni del modello il **Classification Rate**: $\frac{TP+TN}{TP+TN+FP+FN}$ e **AUC(Area Under the Curve)**. L'AUC rappresenta quindi la probabilità che il modello assegni un punteggio più alto a un campione positivo rispetto a uno negativo. Un valore di AUC pari a:

- 1.0 indica un modello perfetto.
- 0.5 equivale a un modello casuale, incapace di distinguere tra le classi.
- ≤ 0.5 significa che il modello è peggiore di un'assegnazione casuale (in questo caso, potrebbe esserci un'inversione nella classificazione).

I risultati riportano i tassi di classificazione (Classification Rate, CR) ottenuti utilizzando diversi modelli di classificazione (SVM e AdaBoost) applicati a varie regioni di interesse (ROI) del viso. Questi risultati offrono una chiara indicazione delle prestazioni dei modelli in base alle ROI analizzate e alle loro combinazioni.

- Per il modello SVM, il tasso di classificazione è più alto quando vengono combinate più ROI, raggiungendo il valore massimo di 0.95 con la combinazione di eyes+mouth+face. Ciò dimostra l'efficacia dell'integrazione delle informazioni provenienti da diverse regioni del viso.
- Anche il modello AdaBoost segue una tendenza simile, con il tasso di classificazione più alto pari a 0.92, sempre per la combinazione eyes+mouth+face.
- Con entrambi i modelli, le singole ROI, come mouth o left-eye, ottengono tassi di classificazione inferiori rispetto alle combinazioni multiple, indicando che informazioni locali limitate non sono sufficienti per ottenere una classificazione ottimale.

ROI	SVM(%)	Adaboost(%)
Occhio sinistro	≈ 85	≈ 84
Occhio destro	≈ 84	≈ 83
Bocca	≈ 78	≈ 81
Viso	≈ 85	≈ 85
Occhi+Bocca	≈ 90	≈ 89
Occhi+Bocca+Viso	≈ 95	≈ 92

Table 4.1: Tassi di classificazione (Classification Rate) ottenuti con SVM e AdaBoost su diverse ROI e combinazioni di ROI.

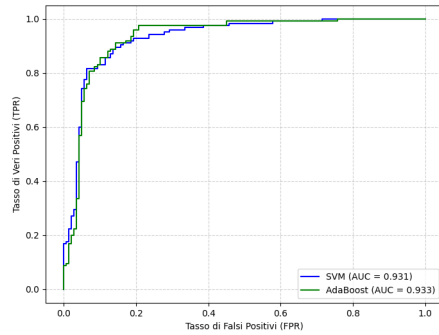
- In particolare, la ROI mouth presenta valori di CR più bassi (0.78 per SVM e 0.81 per AdaBoost), probabilmente perché le caratteristiche estratte da questa regione non sono sufficientemente distintive da sole.

Come si nota in figura prime cinque immagini rappresentano le curve AUC ottenute da una procedura di cross-validation con cinque iterazioni, in cui per ciascuna iterazione sono stati addestrati e valutati due classificatori, SVM e AdaBoost. Ogni immagine corrisponde a una diversa iterazione della cross-validation. L'ultima immagine mostra la curva AUC ottenuta utilizzando un dataset di test completamente separato, non visto durante l'addestramento.

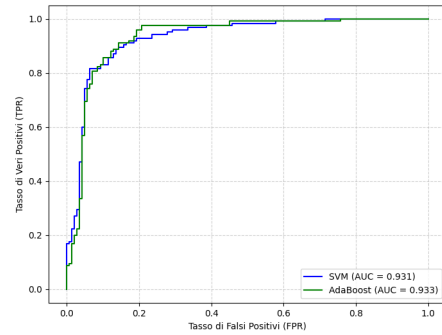
Osservando le curve AUC delle iterazioni di cross-validation, notiamo che entrambi i classificatori mostrano una buona capacità predittiva, con valori di AUC generalmente alti in tutte le iterazioni. La performance di SVM tende ad essere leggermente superiore rispetto a quella di AdaBoost in alcune iterazioni, con una curva più vicina all'angolo superiore sinistro del grafico, che rappresenta un miglior compromesso tra il tasso di veri positivi (TPR) e il tasso di falsi positivi (FPR). Tuttavia, le differenze sono minime, indicando che entrambi i classificatori sono efficaci per questo problema.

Nell'ultima immagine, dove i classificatori sono stati testati su un dataset completamente separato, si osserva un comportamento simile. Entrambi i classificatori mantengono un'alta AUC, confermando la loro capacità di generalizzare a nuovi dati. SVM mantiene un leggero vantaggio rispetto ad AdaBoost, ma il margine è contenuto.

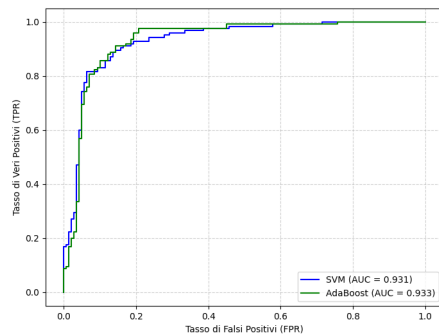
Questi risultati indicano che entrambi i classificatori sono adatti per il problema, con SVM che potrebbe essere una scelta preferibile per via della sua consistenza nelle iterazioni e della capacità di mantenere un alto valore di AUC. AdaBoost, d'altra parte, mostra una buona robustezza e potrebbe essere vantaggioso in scenari con dati più rumorosi o con più variazioni, grazie alla sua natura ensemble.



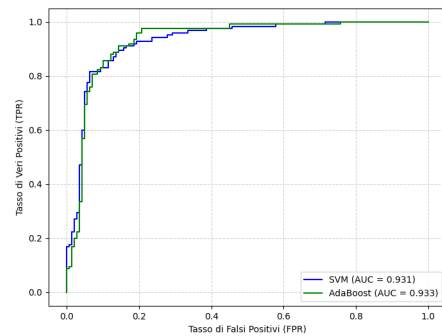
(a) Iterazione 1



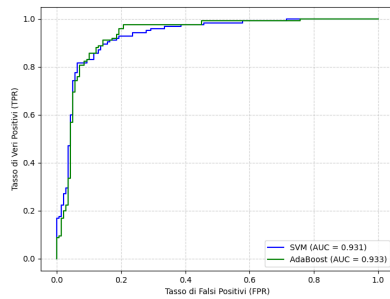
(b) Iterazione 2



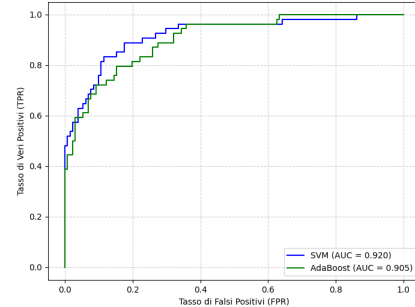
(c) Iterazione 3



(d) Iterazione 4



(e) Iterazione 5



(f) Testing su dataset separato

Figure 4.2: Dalla prima immagine alla quinta la curva ROC del detector SVM e Adaboost per tutte e 5 iterazioni nella cross-validation. L'ultima immagine mostra la curva ROC dei due detector testati su un dataset diverso da quello di training.

La matrice di confusione ottenuta in figura 4.3 durante il testing sul dataset indipendente evidenzia alcuni aspetti interessanti da analizzare. Si osserva che il modello è particolarmente efficace nel riconoscere i volti non truccati, con solo un errore su un totale di 54 campioni. Questo suggerisce che le caratteristiche distintive dei volti non truccati sono

state apprese in modo molto accurato durante l'addestramento. Al contrario, nel caso dei volti truccati, il modello ha commesso 21 errori su 131 campioni, classificandoli erroneamente come non truccati. Questo potrebbe indicare che alcune immagini di volti truccati presentano caratteristiche simili ai volti non truccati, forse a causa di un trucco leggero o della mancanza di pattern distintivi rilevabili.

Un aspetto positivo è che il modello mostra una buona capacità di generalizzazione, data l'alta precisione raggiunta su un dataset completamente distinto da quello utilizzato in fase di addestramento. Tuttavia, gli errori nella classificazione dei volti truccati suggeriscono che il modello potrebbe non essere altrettanto robusto nel catturare la variabilità all'interno di questa classe. Differenze in illuminazione, pose o l'intensità del trucco potrebbero aver contribuito a questi errori. È possibile che l'inclusione di ulteriori dati di addestramento con una maggiore variabilità o l'uso di feature più discriminative possa migliorare le prestazioni.

La performance complessiva del modello appare comunque molto promettente. Nonostante i 21 errori sui volti truccati, il basso tasso di errore nei volti non truccati e l'alto numero di classificazioni corrette indicano che il sistema è in grado di affrontare in modo efficace il compito assegnato. In figura si possono vedere delle previsioni applicate a visi con diverse pose, illuminazioni ecc

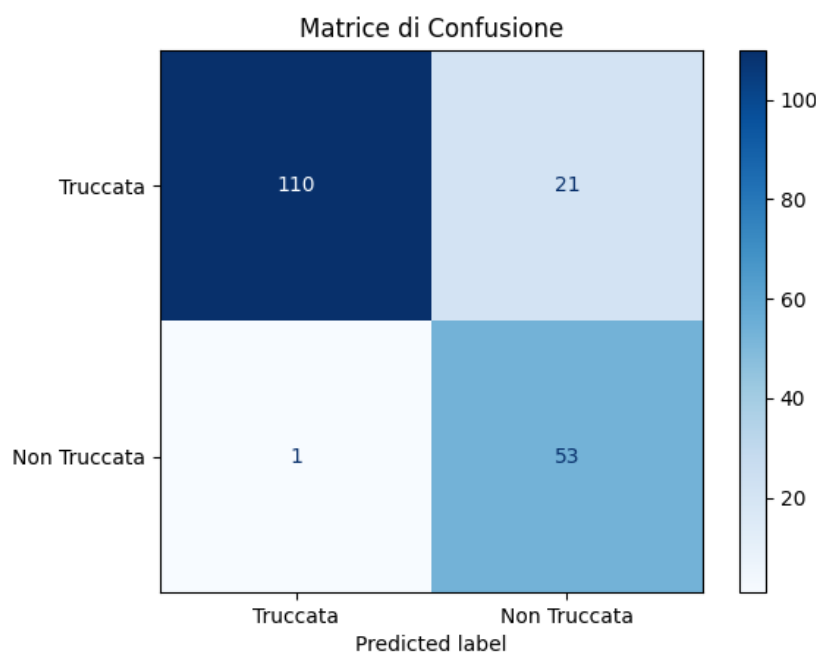


Figure 4.3: Matrice Confusione

In figura 4.4 si possono vedere delle previsioni applicate a visi con diverse pose, illuminazioni ecc

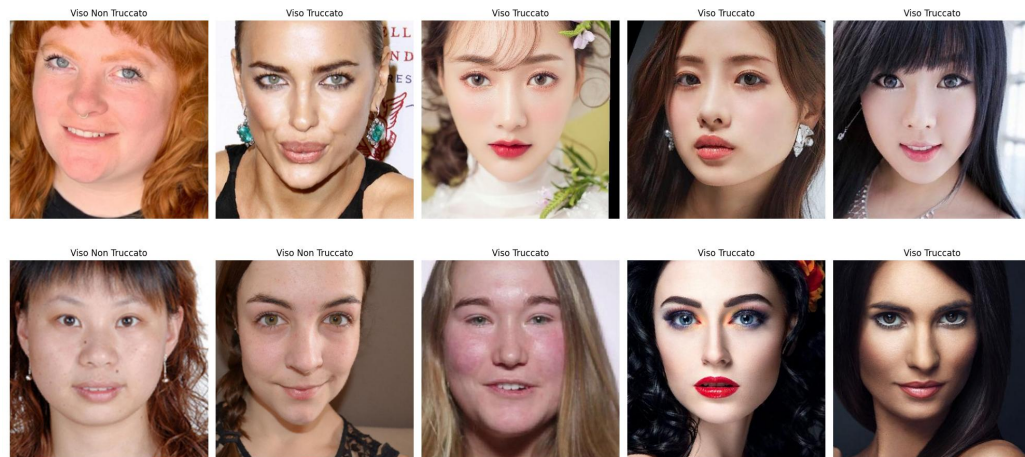


Figure 4.4: Risultati delle previsioni

5 Conclusioni

Questo progetto rappresenta un contributo significativo nell'ambito del riconoscimento facciale e della classificazione automatica, affrontando una sfida complessa e affascinante come la rilevazione del trucco facciale. I risultati raggiunti dimostrano un'attenta pianificazione metodologica, una solida implementazione tecnica e un'analisi approfondita delle prestazioni. La combinazione di tecniche classiche di elaborazione delle immagini con approcci di machine learning, in particolare con algoritmi come SVM, AdaBoost e tecniche di ensemble learning, ha evidenziato la validità dell'approccio e la sua efficacia in scenari controllati.

Uno degli aspetti che merita di essere sottolineato è l'attenzione dedicata alla costruzione delle feature. L'integrazione di descrittori basati su colore, forma e texture ha consentito di rappresentare in maniera esaustiva le differenze visive introdotte dal trucco. Le analisi dei risultati, supportate da visualizzazioni dettagliate come matrici di confusione e curve AUC, hanno fornito una chiara panoramica delle capacità e dei limiti del sistema sviluppato. Tuttavia, è emerso anche come l'accuratezza possa essere influenzata da condizioni esterne, come pose particolari o livelli di trucco meno evidenti, evidenziando la necessità di ulteriori miglioramenti.

Dal punto di vista metodologico, l'uso di tecniche di feature selection basate su modelli e la scelta di classificatori deboli in AdaBoost sono scelte ben motivate e che riflettono una buona comprensione del problema. Tuttavia, un possibile sviluppo futuro potrebbe includere una maggiore esplorazione dei parametri ottimali attraverso analisi più avanzate, andando oltre la semplice selezione manuale o iterativa.

Gli sviluppi futuri nel campo del face recognition, in particolare considerando la possibilità di rilevare e analizzare il trucco facciale, offrono numerose opportunità per migliorare la precisione, la robustezza e l'applicabilità di questi sistemi in vari contesti. Un approccio interessante consiste nell'utilizzare tecniche avanzate di elaborazione delle immagini per ridurre o eliminare virtualmente gli effetti del trucco facciale, ripristinando i tratti naturali del viso. Questo processo, noto come de-makeup processing, potrebbe avere

un impatto significativo in scenari dove l'identità autentica di una persona è essenziale, come nei sistemi di sicurezza o di autenticazione biometrica.