



UNIVERSITÀ DEGLI STUDI DI SALERNO

Dipartimento di Informatica

Corso di Laurea Triennale in Informatica

TESI DI LAUREA

# Un Nuovo Approccio per la Generazione Automatica di Mesh e Texture da una Singola Immagine

RELATORE

Prof. Fabio Palomba

Dott. Viviana Pentangelo

Dott. Dario Di Dario

Università degli Studi di Salerno

CANDIDATO

**Francesco Matteis**

Matricola: 0512105632

Anno Accademico 2023-2024

*Questa tesi è stata realizzata nel*

sesa<sup>lab</sup>  
SOFTWARE ENGINEERING  
SALERNO

*A mia Mamma Pina.*

## **Abstract**

Nel contesto del metaverso, l'immersività gioca un ruolo fondamentale nel facilitare la comunicazione tra individui. Tuttavia, uno dei principali ostacoli all'immersività del metaverso risiede nella complessità della creazione degli avatar, che richiede skill tecniche avanzate. Per affrontare questa sfida, è emersa la necessità di sviluppare avatar per il metaverso che possano essere creati autonomamente da tutti, eliminando così la dipendenza dalle competenze tecniche e rendendo il processo di creazione più accessibile e intuitivo. Al fine di raggiungere questo obiettivo, è stato creato un approccio innovativo per generare le mesh e le texture degli avatar a partire da una singola immagine di riferimento, che utilizza algoritmi avanzati per ottenere i risultati. In particolare, il processo inizia con la generazione parametrica delle mesh della testa utilizzando un algoritmo genetico. Successivamente, attraverso l'impiego di una rete neurale convoluzionale, vengono create le texture da applicare alle mesh. Lo studio condotto ha dimostrato la fattibilità di generare le teste degli avatar a partire da una singola immagine, nonostante le sfide legate all'utilizzo dell'algoritmo genetico e alla limitazione dei dataset per le texture. L'intera implementazione è disponibile sulla sua repository GitHub <sup>1</sup>

---

<sup>1</sup><https://github.com/francM13/Generate-meshes-and-textures-from-a-single-image>

---

# Indice

---

<b>Elenco delle Figure</b>	<b>iii</b>
<b>Elenco delle Tabelle</b>	<b>iv</b>
<b>1 Introduzione</b>	<b>1</b>
1.1 Contesto . . . . .	1
1.2 Motivazione e Obiettivi . . . . .	2
1.2.1 Limiti della Comunicazione a Distanza . . . . .	2
1.2.2 Il Potenziale del Metaverso per Superare i Limiti della Comunicazione a Distanza . . . . .	2
1.2.3 Obiettivi dello Studio . . . . .	3
1.3 Obiettivi e Risultati . . . . .	3
1.3.1 Soluzioni Progettuali . . . . .	4
1.3.2 Risultati e Valutazione . . . . .	4
<b>2 Background</b>	<b>6</b>
2.1 Metaverso . . . . .	6
2.1.1 Realtà Virtuale . . . . .	7
2.1.2 Realtà Aumentata . . . . .	7
2.2 Grafica 3D . . . . .	8
2.2.1 Mesh . . . . .	8

2.2.2	Texture . . . . .	8
<b>3</b>	<b>Strumenti Correlati</b>	<b>10</b>
3.1	Metaversi . . . . .	10
3.1.1	Avatar Meta . . . . .	11
3.1.2	Avatar Apple . . . . .	11
3.2	Strumenti Professionali . . . . .	12
<b>4</b>	<b>Sviluppo</b>	<b>14</b>
4.1	Definizione Della Metodologia di Sviluppo . . . . .	14
4.1.1	Ricerca e Selezione degli Articoli . . . . .	14
4.1.2	Valutazione e Combinazione degli Approcci . . . . .	16
4.2	Implementazione . . . . .	19
4.2.1	Scelte Implementative . . . . .	19
4.2.2	Mesh . . . . .	21
4.2.3	Texture . . . . .	23
4.2.4	Inferenza . . . . .	25
<b>5</b>	<b>Sperimentazione dell'approccio</b>	<b>27</b>
5.1	Metodo di Sperimentazione . . . . .	27
5.2	Valutazione Mesh . . . . .	28
5.3	Valutazione Texture . . . . .	28
5.4	Valutazione Avatar Completo . . . . .	30
<b>6</b>	<b>Conclusioni</b>	<b>31</b>
6.1	Criticità Ricontrate . . . . .	31
6.2	Sviluppi Futuri . . . . .	32
6.3	Usi alternativi . . . . .	32
	<b>Bibliografia</b>	<b>34</b>

---

## Elenco delle figure

---

2.1	Utilizzo del Metaverso nell'ambito lavorativo [1] . . . . .	7
2.2	Visore per la realtà virtuale [2] . . . . .	8
2.3	Esempio di una mesh tridimensionale [3] . . . . .	9
2.4	Visualizzazione dell'applicazione UV Mapping[4] . . . . .	9
3.1	Avatar Sviluppati da Meta[5] . . . . .	11
3.2	Videochiamata FaceTime con Vision Pro[6] . . . . .	12
3.3	Editing di personaggi 3d con l'utilizzo di Metahuman[7] . . . . .	13
4.1	Diagramma modulo generazione mesh . . . . .	20
4.2	Diagramma modulo generazione texture . . . . .	20
4.3	Esempio funzionamento di FLAME[8] . . . . .	21
4.4	Preview landmark Mediapipe . . . . .	22
4.5	Mesh generata utilizzando il GA . . . . .	23
4.6	Elemento nel dataset utilizzato come target[9] . . . . .	24
4.7	Architettura Autoencoder [9][10] . . . . .	25
5.1	Visualizzazione incoerenze mesh . . . . .	28
5.2	Heatmap degli errori su tutto il Test-set . . . . .	29
5.3	Risultato generazione texture . . . . .	29
5.4	Risultato Completo . . . . .	30

---

## **Elenco delle tabelle**

---



# CAPITOLO 1

---

## Introduzione

---

### 1.1 Contesto

La comunicazione a distanza è un elemento fondamentale nella società moderna, influenzando in maniera significativa diversi aspetti della vita quotidiana. La comunicazione a distanza ha rivoluzionato il concetto di distanza geografica, consentendo alle persone di tutto il mondo di connettersi istantaneamente. Grazie a strumenti come videochiamate e messaggistica istantanea come Google Meet<sup>1</sup> e Zoom<sup>2</sup>, le persone possono comunicare in tempo reale, facilitando la collaborazione internazionale in settori come la ricerca scientifica e l'istruzione. Questa interconnessione globale ha aperto nuove opportunità di lavoro, consentendo alle persone di svolgere mansioni da remoto e di lavorare in modo flessibile da qualsiasi luogo con accesso a internet. L'avvento della comunicazione a distanza ha dato vita a nuove forme di socializzazione online attraverso piattaforme come i social media e le app di messaggistica. Queste piattaforme hanno radicalmente trasformato il modo in cui le persone interagiscono con amici e familiari, offrendo la possibilità di rimanere in contatto anche con coloro che si trovano lontano geograficamente. La socializzazione online

---

<sup>1</sup><https://meet.google.com>

<sup>2</sup><https://zoom.us>

ha creato nuove dinamiche relazionali, influenzando la percezione di vicinanza e connessione tra individui separati da distanze fisiche.

## **1.2 Motivazione e Obiettivi**

Nonostante i numerosi vantaggi della comunicazione a distanza, soprattutto tramite strumenti come le videochiamate, esistono alcuni limiti importanti da considerare.

### **1.2.1 Limiti della Comunicazione a Distanza**

- **Mancanza di Comunicazione Non Verbale:** La comunicazione non verbale, che include segnali come il linguaggio del corpo ed espressioni facciali, è essenziale per una comunicazione completa. Le videochiamate, pur consentendo di vedere il volto dell'interlocutore, non trasmettono appieno questi segnali non verbali, potenzialmente causando incomprensioni e ambiguità.
- **Difficoltà di Concentrazione:** Mantenere la concentrazione e prestare attenzione durante le videochiamate può essere difficile a causa di fattori di distrazione presenti nell'ambiente circostante. Questo può influenzare negativamente la produttività e la qualità della comunicazione.
- **Isolamento Sociale:** L'eccessivo affidamento sulla comunicazione mediante videochiamate può comportare diversi effetti negativi sull'interazione umana e sul benessere individuale, come la mancanza di un'interazione fisica e di esperienze condivise nel mondo può contribuire a isolare le persone e a creare barriere emotive.

### **1.2.2 Il Potenziale del Metaverso per Superare i Limiti della Comunicazione a Distanza**

Un metaverso, uno spazio virtuale 3D condiviso e persistente, ha il potenziale di mitigare i limiti della comunicazione a distanza e delle videochiamate offrendo:

- Creazione di avatar e ambienti virtuali realistici per migliorare la sensazione di presenza e ridurre l'isolamento sociale.
- Miglioramento della comunicazione non verbale attraverso espressioni facciali dettagliate e il linguaggio del corpo degli avatar.
- Facilitare l'interazione e la collaborazione tra gli utenti, consentendo di condividere oggetti virtuali, lavorare su progetti insieme e partecipare ad attività comuni.
- Migliora la produttività e il senso di comunità.

Nonostante i progressi negli avatar virtuali, alcune sfide persistono come la mancanza di realismo e fedeltà nella rappresentazione della persona e delle espressioni facciali.

### **1.2.3 Obiettivi dello Studio**

Lo studio mira a creare un nuovo approccio per la generazione di volti realistici per avatar virtuali, utilizzando modelli di Intelligenza Artificiale. L'obiettivo è rendere la creazione degli avatar accessibile a tutti, ricreando fedelmente la persona che li utilizza senza richiedere skill tecniche avanzate. Questo permetterà di superare i limiti attuali dei metaversi e di altri strumenti di comunicazione. Attraverso l'utilizzo di tecnologie all'avanguardia, l'approccio sviluppato mira a migliorare la comunicazione virtuale e a supportare la rapida espansione dei sistemi di metaverso nel presente e nel futuro.

## **1.3 Obiettivi e Risultati**

Ci si è posto l'obiettivo di generare rapidamente volti e teste realistici da usare come avatar virtuali, eliminando la necessità di competenze tecniche avanzate. Questo permetterà di inserirli facilmente in mondi virtuali, indipendentemente dalla piattaforma utilizzata. Tuttavia, durante il processo di sviluppo, sono emersi alcuni limiti e sfide da affrontare per sviluppare un approccio completamente autonomo per la generazione di volti realistici come:

1. **Compatibilità con Ambienti Diversi:** Uno dei principali ostacoli è rappresentato dalla necessità di garantire la compatibilità dei risultati generati con ambienti virtuali diversi, al fine di non avere limiti implementativi per l'approccio sviluppato, e a non limitare possibili sviluppi futuri.
2. **Complessità e Skill Richieste:** La generazione di volti realistici richiede competenze avanzate di modellazione 3D, il che comporta una sfida in termini di complessità e abilità necessarie per ottenere risultati convincenti.
3. **Generazione di Volti in Posizione Neutra:** Un'altra sfida consiste nel generare volti in una posizione neutra, in modo da poter essere facilmente riggabili così da poter seguire le espressioni dell'utente che devono rappresentare.

### **1.3.1 Soluzioni Progettuali**

Per superare tali limiti, sono state adottate diverse scelte progettuali mirate a garantire l'efficacia e la praticità dello strumento sviluppato:

- **Utilizzo di Tecnologie e Standard del Settore:** È stata privilegiata l'adozione di tecnologie e standard consolidati nel settore per la creazione dei volti degli avatar, al fine di garantire una qualità ottimale e una maggiore facilità d'uso.
- **Selezione di Strumenti Ausiliari:** Sono stati scelti accuratamente strumenti ausiliari il cui livello di qualità è stato verificato sulla base della letteratura scientifica, assicurando un supporto affidabile ed efficace durante il processo di generazione dei volti.

### **1.3.2 Risultati e Valutazione**

Dai risultati ottenuti attraverso le varie esecuzioni e test dell'approccio, è emerso che il metodo implementato è in grado di funzionare in modo efficace, sebbene la qualità complessiva dei volti generati non sia eccezionale. È importante sottolineare che allo stato attuale l'approccio è stato concepito principalmente come un prototipo, volto a testare la fattibilità in termini di sviluppo, piuttosto che come un prodotto finito destinato all'uso massivo. In conclusione, nonostante le sfide incontrate durante

il processo di sviluppo, l'approccio per la generazione di avatar realistici in ambienti virtuali si è dimostrato un valido esperimento che apre interessanti prospettive per il futuro sviluppo di metodi simili, con un potenziale impatto significativo nel mondo della realtà virtuale e della modellazione 3D.

## CAPITOLO 2

---

### Background

---

#### 2.1 Metaverso

Il termine "metaverso" è stato coniato da Neal Stephenson nel suo romanzo di fantascienza del 1992 "Snow Crash". Nel libro, il metaverso è un mondo virtuale condiviso in 3D dove le persone possono interagire tra loro e con i computer come avatar. Anche se il metaverso non esiste ancora in questa forma completa, l'idea di base è quella di creare uno spazio virtuale persistente e interoperabile in cui le persone possano svolgere le stesse attività che fanno nel mondo reale, come socializzare, lavorare, giocare e fare acquisti come mostrato nella Figura 2.1. Per accedere al metaverso, saranno necessari dispositivi hardware e software speciali, come visori VR o AR. Tra le sue caratteristiche chiave ci sono l'immersività, il metaverso dovrebbe essere un'esperienza immersiva che faccia sentire gli utenti come se fossero davvero all'interno del mondo virtuale. Questo sarà realizzato utilizzando una combinazione di tecnologie come la realtà virtuale (VR), la realtà aumentata (AR) e il motion capture. Il metaverso dovrebbe essere un mondo persistente che continua a esistere anche quando gli utenti non sono connessi e uno spazio sociale in cui gli utenti possano interagire tra loro.



**Figura 2.1:** Utilizzo del Metaverso nell'ambito lavorativo [1]

### 2.1.1 Realtà Virtuale

La realtà virtuale (VR), è una tecnologia immersiva che proietta l'utente in un qualsiasi luogo 3D mediante l'utilizzo di appositi device che attraverso la combinazione di hardware e software sono in grado di creare un ambiente simulato tridimensionale all'interno del quale è possibile muoversi e interagire come se ci si trovasse nel mondo reale. Per fruire di questa esperienza immersiva, è necessario utilizzare appositi visori VR come si vede nella Figura 2.2, grazie a schermi e sensori, catturano i movimenti della testa e del corpo dell'utente, trasmettendoli al software che, a sua volta, aggiorna l'ambiente virtuale in tempo reale. Oltre ai visori, la realtà virtuale può essere fruita anche attraverso altri dispositivi, come guanti, scarpe e tapis roulant digitali, che permettono un'interazione ancora più realistica con l'ambiente virtuale.

### 2.1.2 Realtà Aumentata

La realtà aumentata (AR) è una tecnologia che integra elementi digitali nel mondo reale, sovrapponendoli alla vista dell'utente attraverso dispositivi come smartphone, tablet, visori AR o occhiali intelligenti. A differenza della realtà virtuale (VR), che crea un ambiente completamente virtuale, la AR arricchisce la realtà percepita con informazioni digitali contestuali, rendendo l'esperienza più interattiva e coinvolgente.



**Figura 2.2:** Visore per la realtà virtuale [2]

## 2.2 Grafica 3D

La grafica 3D, o computer grafica, è un campo dell'informatica che si occupa della creazione di immagini e animazioni tridimensionali. A differenza della grafica 2D, che rappresenta gli oggetti su un piano bidimensionale, la grafica 3D li modella in uno spazio virtuale tridimensionale, permettendo di visualizzare da qualsiasi angolazione e con un livello di realismo molto elevato.

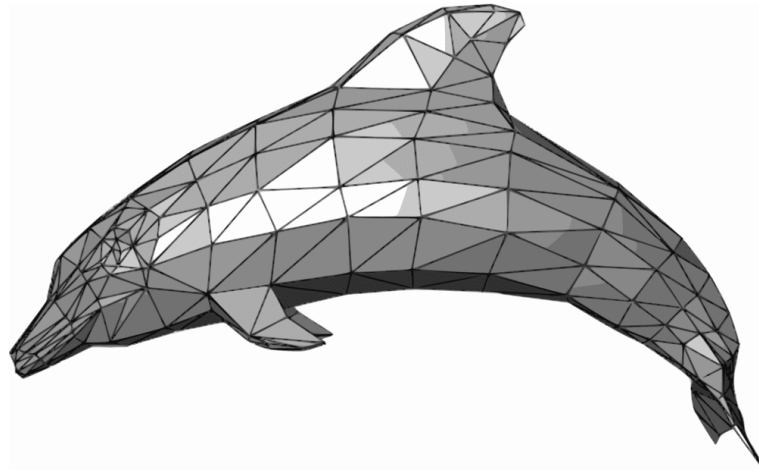
### 2.2.1 Mesh

Le mesh, o maglie poligonali, come si può vedere nella Figura 2.3, sono la rappresentazione digitale più comune degli oggetti 3D nella grafica computerizzata. Sono costituite da un insieme di punti, chiamati vertici, collegati tra loro da segmenti, chiamati spigoli, che definiscono la forma dell'oggetto. Le facce, invece, sono delle chiusure formate da tre o più spigoli che racchiudono i vertici e danno vita alla superficie dell'oggetto.

### 2.2.2 Texture

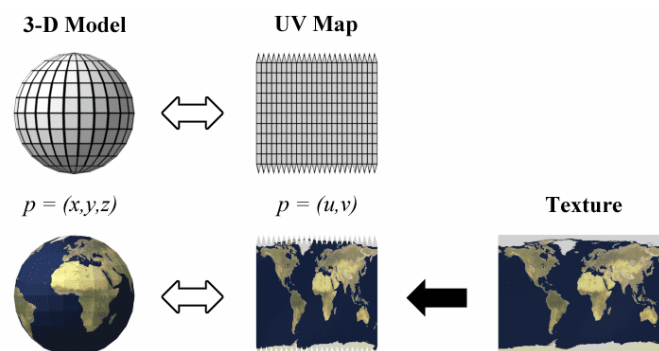
Le texture sono un elemento fondamentale della computer grafica, utilizzate per rivestire la superficie di un oggetto e conferirgli un aspetto più realistico e dettagliato.





**Figura 2.3:** Esempio di una mesh tridimensionale [3]

Come si può vedere dalla Figura 2.4, una texture è un'immagine che viene applicata sulla superficie di un oggetto 3D, per simulare la sua struttura e il suo aspetto, in modo da renderlo più realistico e simile all'oggetto reale. Le texture possono essere applicate sulla superficie degli oggetti in diversi modi, tra cui la mappatura UV, che consiste nel proiettare l'immagine della texture sulla superficie dell'oggetto, come se fosse un tessuto.



**Figura 2.4:** Visualizzazione dell'applicazione UV Mapping[4]

## CAPITOLO 3

---

### Strumenti Correlati

---

#### 3.1 Metaversi

Ci sono diverse piattaforme di metaversi attive al momento, ciascuna con funzionalità e livelli di immersività diversi. Alcuni esempi includono The Sandbox<sup>1</sup>, Decentraland<sup>2</sup>, Meta Horizon Worlds<sup>3</sup>. Attualmente, le aziende leader nello sviluppo dei metaversi sono Meta<sup>4</sup> e Apple<sup>5</sup>. Entrambi i giganti della tecnologia stanno investendo risorse considerevoli nello sviluppo di piattaforme e tecnologie innovative per offrire esperienze coinvolgenti agli utenti. Meta sta sviluppando la sua fascia di visori VR, i Meta Quest. L'obiettivo è di fornire una grafica più realistica e un'interazione naturale con il mondo virtuale. L'ambizione di Meta è quella di creare un metaverso aperto e accessibile a tutti, dove le persone possono lavorare, giocare, socializzare e fare acquisti. Apple, d'altra parte, ha sviluppato e rilasciato su un visore AR-VR chiamato Vision Pro, che punta a combinare la realtà virtuale con quella aumentata, offrendo un'esperienza ibrida unica.

---

<sup>1</sup><https://www.sandbox.game/en>

<sup>2</sup><https://decentraland.org>

<sup>3</sup><https://horizon.meta.com>

<sup>4</sup><https://www.meta.com/it>

<sup>5</sup><https://www.apple.com/it>

### 3.1.1 Avatar Meta

Gli avatar sviluppati da Meta come visibile nella Figura 3.1, rappresentano un passo significativo verso la costruzione di un futuro virtuale interconnesso. Questi avatar digitali hanno lo scopo di fungere da rappresentazioni degli utenti all'interno delle varie piattaforme e esperienze offerte da Meta. Meta ha sviluppato il proprio metaverso, Horizon World, come un insieme di mondi completamente immersivi che non si sovrappongono alla realtà fisica. Questo ambiente virtuale offre agli utenti la possibilità di interagire, creare e connettersi in modi innovativi. Gli avatar creati da Meta sono concepiti per essere inseriti all'interno di questi contesti virtuali. Presentano un design caricaturale, piuttosto che realistico, il che potrebbe rappresentare un ostacolo alla comunicazione tra le persone. Tuttavia, questa scelta di design si configura come un compromesso necessario data la limitata potenza computazionale attuale dei visori per la realtà virtuale.



**Figura 3.1:** Avatar Sviluppati da Meta[5]

### 3.1.2 Avatar Apple

Apple, nota per la sua costante innovazione nel campo della tecnologia, si è recentemente concentrata sulla realtà aumentata. Questa tecnologia permette di mostrare elementi virtuali nel mondo reale in modo che sembrino effettivamente presenti. Per comunicare in modo più coinvolgente, Apple ha adottato un approccio unico. La società ha scelto di ricreare i volti delle persone che indossano il loro visore per tracciare le espressioni facciali degli individui. Come si vede dalla Figura 3.2 il

Vision Pro utilizza algoritmi di machine learning che sono in grado di ricostruire un volto simulato dell'individuo basato su uno scan preliminare. Una volta generati, i volti simulati vengono visualizzati all'interno di finestre mostrate ad altri utenti. Anche se questo approccio potrebbe non garantire un'immersione totale in un mondo completamente virtuale, rappresenta un compromesso interessante per il sistema visionOS di Apple, più orientato verso la Realtà Aumentata.

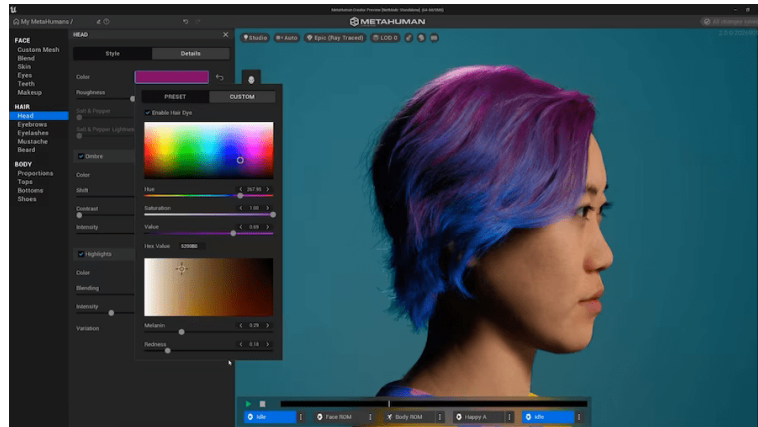


**Figura 3.2:** Videochiamata FaceTime con Vision Pro[6]

## 3.2 Strumenti Professionali

Nel campo della creazione di mesh e texture per volti umani e non, diversi strumenti si sono affermati per la loro capacità di generare modelli fotorealistici di alta qualità. Tuttavia, la maggior parte di questi strumenti sono progettati principalmente per lo sviluppo di videogiochi o effetti visivi (VFX), come Metahuman e Character Creator che come mostra la Figura 3.3 sono due software progettati per la creazione di personaggi digitali fotorealistici di altissima qualità. Questi strumenti creano personaggi che sono completamente riggati e pronti all'uso. Inoltre, consentono una facile personalizzazione e modifica dei personaggi creati[7][11]. Per creare modelli 3D di alta qualità e risoluzione utilizzando questi strumenti, sono necessarie competenze specifiche. Queste competenze includono la conoscenza della modellazione 3D, del rigging e dell'animazione. Queste abilità sono fondamentali per ottenere risultati ottimali nella creazione di personaggi digitali complessi e realistici. Nonostante le

loro capacità avanzate nella creazione di personaggi digitali per VFX e videogiochi, i software come Metahuman e Character Creator presentano alcune limitazioni. Questi strumenti sono progettati per scopi specifici e sono quindi molto pesanti da far girare. Inoltre, non sono pratici per la generazione di mesh e texture in larga scala e non sono adatti per un utilizzo autonomo, come potrebbe essere necessario in un metaverso o in contesti simili.



**Figura 3.3:** Editing di personaggi 3d con l'utilizzo di Metahuman[7]

## 4.1 Definizione Della Metodologia di Sviluppo

### 4.1.1 Ricerca e Selezione degli Articoli

Durante la fase di sviluppo, sono stati esaminati diversi metodi per la generazione sia delle mesh che delle texture, consultando la letteratura scientifica disponibile. Di seguito, vengono presentati alcuni dei metodi che sono emersi come i più promettenti.

#### **Mesh**

Nella ricerca della mesh del volto, sono state identificate varie metodologie presenti nella letteratura scientifica. Il lavoro di Feng et al.[12] propone un'innovativa strategia per la generazione delle mesh dei volti umani. Questo approccio si basa su una rappresentazione 2D del volto nota come "UV position map", la quale registra le coordinate 3D della pointcloud del volto. Il metodo risulta particolarmente rapido nell'esecuzione delle operazioni e risulta essere caratterizzato da un basso impatto sulle risorse computazionali. Tuttavia, nonostante i suoi punti di forza, il metodo presenta delle limitazioni significative tra cui la generazione limitata solo alla zona del volto che quindi non copre l'intera testa, questa limitazione impedisce la genera-

zione di avatar che richiedono una visualizzazione a 360 gradi. Lo studio di Soltani et al.[13] si concentra sulla creazione di mesh generiche, quindi non specializzate per volti o teste umane utilizzando una serie di immagini di uno stesso soggetto da punti di vista diversi. Il metodo utilizza immagini multi vista per generare depthmap e silhouette del soggetto. Sfruttando una rete neurale artificiale, trasforma le informazioni delle immagini in mesh tridimensionali. Le principali limitazioni di questo metodo sono la dipendenza dalla disponibilità di immagini multi vista dei volti dei soggetti, e nell'ottenere un dataset sufficientemente grande e diversificato per garantire la qualità e l'accuratezza della generazione delle mesh. Il lavoro di Kamyab e Azimifar[14] propone un metodo per la ricostruzione della forma tridimensionale utilizzando landmark bidimensionali estratti da un'immagine singola. Il metodo si basa sull'estrazione di landmark bidimensionali da un'immagine tramite l'uso di modelli pre-addestrati, successivamente questi landmark vengono impiegati come punti di riferimento per la ricostruzione della mesh tridimensionale del volto. Il lavoro di Sanyal et al.[15] propone un metodo per la ricostruzione di una mesh 3D dell'intera testa di un individuo. Questo approccio si basa sull'utilizzo di un insieme di molteplici immagini generiche della persona da punti di vista diversi. L'approccio proposto sfrutta FLAME[8], uno strumento progettato per generare mesh parametriche di teste. Attraverso l'impiego di una rete neurale, il sistema è in grado di stimare i parametri di FLAME. Ciò consente di ottenere una ricostruzione della mesh che risulta indipendente da fattori come illuminazione, posa della testa ed espressione facciale. Nonostante i suoi vantaggi, la principale limitazione di questa tecnica risiede nella scarsità di immagini da posizioni diverse di un singolo individuo necessarie per il corretto addestramento della rete neurale. Questa carenza può influenzare l'accuratezza e l'efficacia del processo di ricostruzione della mesh 3D della testa.

## Texture

Per la generazione della texture, sono state valutate 2 diverse metodologie che risultavano essere le più promettenti in letteratura scientifica. Il metodo proposto da Henderson et al.[16] si concentra sulla colorazione delle mesh 3D assegnando un colore distinto a ciascuna faccia. Quest'approccio si basa sulla fotogrammetria, che

coinvolge l'allineamento della mesh con le immagini di riferimento sottostanti. Le mesh vengono posizionate in modo da corrispondere alle immagini 2D sottostanti, successivamente si confrontano i colori presenti nelle immagini del soggetto con quelli della mesh sovrastante per assegnare i colori alle singole facce. Le principali limitazioni dell'approccio per il caso in questione sono dovute alla varietà dei colori presenti in un volto umano che risulta molto più ampia rispetto ad altri oggetti, rendendo limitante l'utilizzo di un solo colore per ogni faccia della mesh, specialmente per mesh non particolarmente definite. Il lavoro di Bai et al.[9] presenta un innovativo metodo per la generazione di texture UV di alta qualità a partire da una singola immagine di un volto. Questo approccio rivoluzionario sfrutta tecniche di editing delle immagini facciali per creare volti normalizzati da diverse prospettive, partendo da un'unica immagine di riferimento. Il processo inizia con l'utilizzo un metodo di editing delle immagini facciali per generare volti normalizzati da diverse angolazioni, tutti derivati da un'unica immagine iniziale successivamente, viene applicata una procedura di estrazione, correzione e completamento della texture UV per garantire coerenza e alta qualità nelle texture generate. Le texture UV così ottenute vengono fuse insieme attraverso un processo di fusione per creare una texture completa pronta per il rendering realistico. Uno dei principali limiti di questo metodo è rappresentato dalla complessità della pipeline necessaria per generare le texture UV.

#### **4.1.2 Valutazione e Combinazione degli Approcci**

Dagli approcci individuati nei precedenti studi, dopo aver individuato punti di forza e debolezza si è proceduto alla creazione di pipeline per la generazione di mesh e texture adatte al caso specifico. Partendo dall'obiettivo di ottenere un'immagine generica di una persona, sono stati selezionati i seguenti approcci principali:

##### **Mesh**

1. Nel tentativo di emulare la modellazione 3D di un volto effettuata manualmente, si è optato per un approccio iniziale che coinvolge la creazione di una rappresentazione 2D del modello 3D del volto simile al lavoro di Feng et al.[12]. Questa rappresentazione 2D viene generata a partire da due depthmap della



faccia di una persona, rispettivamente dalla posizione frontale e laterale. Per generare le due immagini frontali e laterali, è stato scelto un metodo che mira a normalizzare le espressioni facciali, l'illuminazione, nonché a rimuovere elementi di occlusione come occhiali o eventuali indumenti che potrebbero interferire con la visualizzazione del volto come visto nel lavoro di Bai et al.[9]. Inoltre, il processo prevede la segmentazione della faccia e la generazione delle posizioni chiave del volto. Al fine di garantire la corretta corrispondenza e proporzionalità della mesh generata con il volto reale, viene eseguita una comparazione tra la mesh generale dalla prospettiva frontale e laterale e le due depthmap acquisite. Questa fase di confronto è fondamentale per assicurare l'accuratezza della ricostruzione tridimensionale del volto. È importante sottolineare che l'approccio descritto presenta una limitazione significativa, la generazione si limita alla sezione del volto di una persona. Questo potrebbe rappresentare un ostacolo nell'ambito degli ambienti immersivi, dove potrebbe essere necessaria una ricostruzione più dettagliata e completa del volto e della testa per un'esperienza ottimale.

2. Per il secondo approccio, è stato scelto di generare le mesh parametricamente utilizzando FLAME[8]. Questo metodo coinvolge diverse fasi per creare una mesh dettagliata di un volto a partire da un'immagine generica. Inizialmente, si genera una depthmap dell'immagine del volto. La depthmap fornisce informazioni sulla distanza tra il soggetto e la fotocamera. Utilizzando un algoritmo di ottimizzazione, si ottimizzano i parametri FLAME[8] della mesh. Questi parametri definiscono le caratteristiche della mesh in base alla struttura del volto. La mesh generata viene allineata con la depthmap per garantire una corretta corrispondenza. Successivamente, si confrontano i risultati ottenuti per ottimizzare ulteriormente la mesh. Tuttavia, un limite significativo di questo approccio è rappresentato dal singolo punto di vista dell'immagine di riferimento, questo può influire sulla precisione della mesh generata, specialmente nelle aree non sovrapposte con la depthmap che si trova nel lato opposto della mesh.
3. Per il terzo approccio, è stato scelto di generare le mesh parametricamente

utilizzando FLAME[8]. Questo processo inizia con un'immagine generica da cui vengono estratti landmark di riferimento del volto utilizzando uno strumento esterno come visto nel lavoro di Kamyab e Azimifar[14]. Questi landmark sono fondamentali per estrarre le caratteristiche del volto che vogliamo ritrovare nella mesh e ottimizzare i parametri di FLAME tramite un algoritmo di ottimizzazione. Si parte da un'immagine di base per iniziare il processo di generazione delle mesh parametriche. Utilizzando uno strumento esterno, vengono estratti i landmark di riferimento dal volto nell'immagine selezionata. I landmark ottenuti vengono utilizzati per estrarre le caratteristiche necessarie per la creazione delle mesh parametriche. Attraverso un algoritmo di ottimizzazione, i parametri di FLAME[8] vengono regolati per generare le mesh in modo ottimale.

## **Texture**

1. Come primo approccio per generare le texture, è stato scelto un approccio basato sulla fotogrammetria simile al lavoro proposto da Henderson et al.[16]. Inizialmente, la mesh generata viene allineata con un'immagine di riferimento. Successivamente, vengono assegnati colori alle singole facce, confrontando la coerenza del rendering della mesh con l'immagine di riferimento. Per far combaciare la colorazione della mesh con l'immagine di riferimento, si utilizza un algoritmo di ottimizzazione. Tuttavia, questo metodo presenta dei limiti. I colori di un volto umano sono estremamente complessi e variegati, il che rende difficile ottenere una corrispondenza accurata assegnando un solo colore per ogni faccia della mesh.
2. Come secondo metodo è stato adottato un approccio basato sull'utilizzo di reti neurali. Questo metodo prevede l'impiego di una rete neurale artificiale che riceve in input un'immagine generica di un volto e produce la texture corrispondente da applicare successivamente. Nel processo di generazione delle texture UV, è stato scelto di utilizzare un dataset costruito a partire dal lavoro proposto da Bai et al.[9]. Questo dataset fornisce le informazioni necessarie alla

rete neurale artificiale per apprendere e generare in modo accurato le texture da applicare sulle mesh.

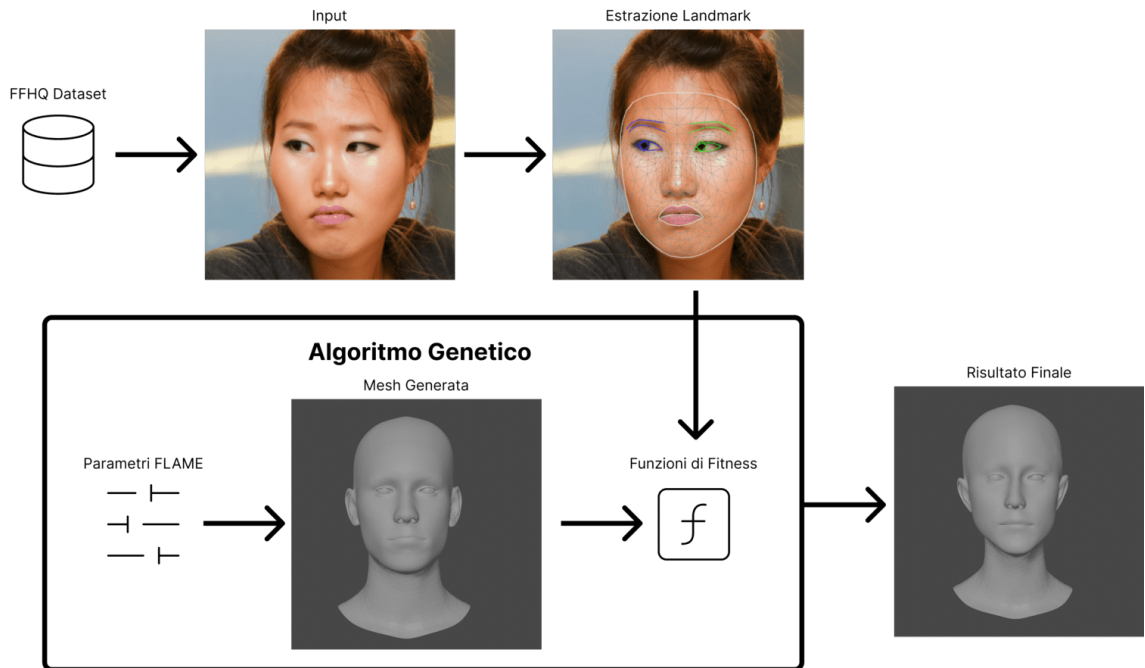
### **Scelta Approcci**

Nel processo di sviluppo di un progetto di modellazione 3D, è essenziale fare una selezione accurata degli approcci per texture e mesh, garantendo la compatibilità tra i due elementi risultanti da tali approcci. In questo contesto, sono stati adottati metodi specifici per la generazione delle mesh e delle texture al fine di ottenere risultati ottimali. Dopo un'attenta valutazione, sono stati scelti i seguenti approcci come metodi finali per la creazione delle mesh e delle texture. Per la generazione delle mesh è stato selezionato il terzo metodo valutato, Figura 4.1. Si è optato per la generazione parametrica delle mesh utilizzando FLAME[8]. Successivamente, è stato impiegato un algoritmo di ottimizzazione per ottimizzare il modello basandosi sui landmark di riferimento. Questa scelta è stata motivata dalla necessità di ottenere precisione nella sezione del volto sia visibile che non. Per la generazione delle texture è stato selezionato il secondo metodo valutato, Figura 4.2. Si è deciso di implementare una rete neurale in grado di elaborare un'immagine generica e restituire la texture UV corrispondente. Questa decisione è stata presa considerando la minore velocità di generazione rispetto ai metodi basati su fotogrammetria, facilitata anche dalla compatibilità delle mesh generate da FLAME[8] con le texture UV. La scelta degli approcci per texture e mesh ha permesso di raggiungere un equilibrio tra precisione, velocità di generazione e compatibilità dei risultati ottenuti.

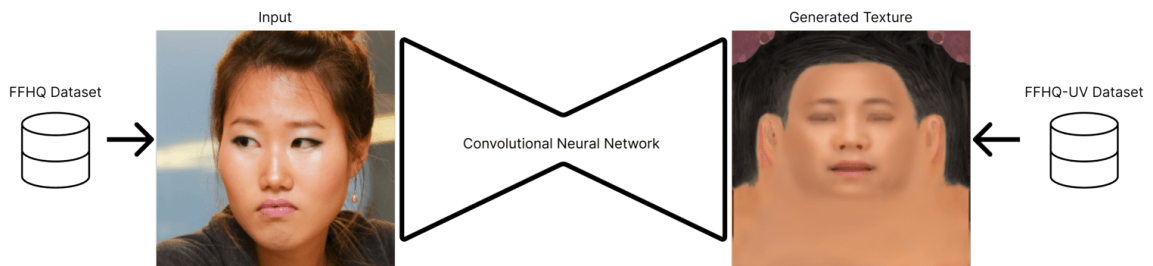
## **4.2 Implementazione**

### **4.2.1 Scelte Implementative**

Per poter ottenere dei volti utilizzabili e compatibili con diversi motori di gioco e metaversi, è fondamentale partire dagli standard del settore per creare Mesh 3D e Texture UV facilmente adattabili. In questo contesto, l'esportazione delle mesh in formato \*.obj e delle texture in formato \*.png rappresenta una scelta comune e funzionale. Per rendere semplice e accessibile la creazione degli avatar a tutti, l'approccio



**Figura 4.1:** Diagramma modulo generazione mesh



**Figura 4.2:** Diagramma modulo generazione texture

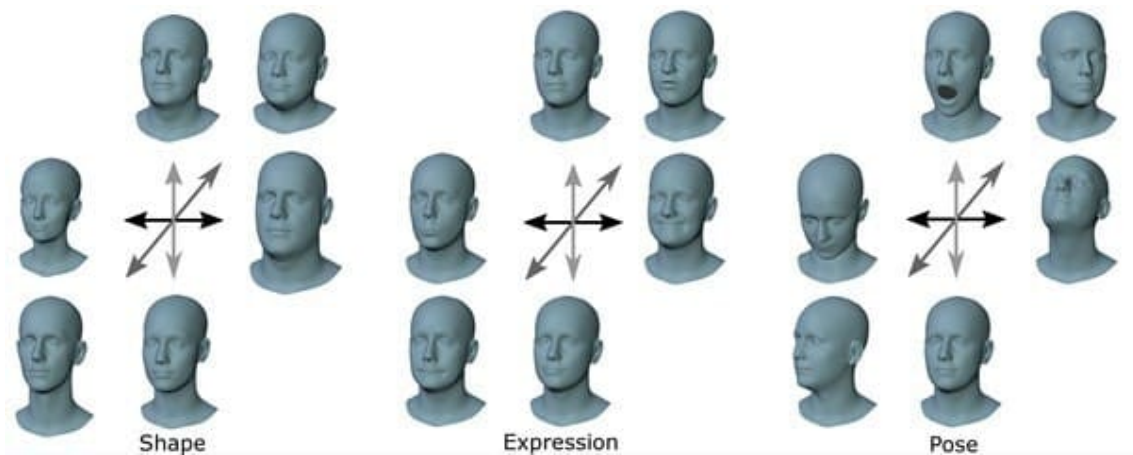
basato sull'utilizzo di immagini rappresenta una soluzione efficace. L'input per la creazione degli avatar consiste in un'immagine della persona da rappresentare, con una posa, un'espressione e un'illuminazione generica. Un obiettivo chiave è rendere gli avatar riggabili, ovvero capaci di emulare le espressioni delle persone che rappresentano. Per raggiungere questo scopo, è essenziale che gli avatar generati si trovino in una posizione neutra. Questo permette una maggiore flessibilità nell'aggiunta di animazioni e espressioni facciali in seguito, contribuendo a creare esperienze più immersive e coinvolgenti nei metaversi.

## 4.2.2 Mesh

### FLAME

Per la generazione delle mesh del volto, si sono valutati diversi metodi e alla fine si è optato per l'utilizzo di FLAME[8] per la sua velocità e precisione.

Cos'è FLAME? FLAME è un framework sviluppato dal lavoro di Li et al.[8] che consente di generare parametricamente le mesh della testa a partire da un insieme di oltre 150 parametri di forma ed espressione come si può vedere dalla Figura 4.3. Questo framework offre la possibilità di creare una vasta gamma di varianti realistiche del volto umano. La decisione di utilizzare FLAME[8] è stata presa dopo aver osservato i risultati eccellenti ottenuti dal lavoro di Sanyal et al.[15] che hanno utilizzato FLAME per realizzare diverse tecniche di generazione di volti con ottimi risultati. Poiché FLAME[8] non è progettato come una libreria e quindi richiede dipendenze con versioni specifiche, si è deciso di containerizzarlo. Questa scelta è stata fatta per evitare conflitti con le librerie necessarie al progetto. Il container accetta i parametri per FLAME tramite una richiesta web e restituisce la mesh del volto generata, offrendo così un'interfaccia efficiente e indipendente dalle specifiche versioni delle librerie.

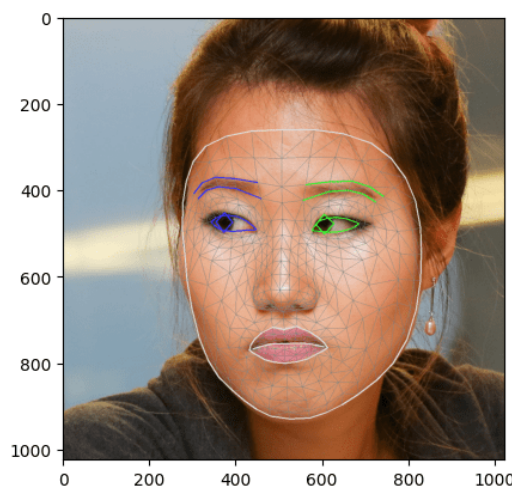


**Figura 4.3:** Esempio funzionamento di FLAME[8]

### Landmark

Nell'ambito della creazione rapida di mesh 3D, è emersa la possibilità di utilizzare i landmark estratti per la ricostruzione delle mesh. Questo approccio si basa sul

lavoro di Kamyab e Azimifar[14]. In particolare, si è scelto di adottare i landmark Mediapipe[17], Mediapipe è una libreria sviluppata da Lugaresi et al.[17] con Google, per l'estrazione dei landmark da un'immagine singola. Mediapipe è stata scelta per la sua capacità di estrarre un elevato numero di riferimenti, più di 400, da un'immagine e dalla qualità dei landmark generati che risulta essere particolarmente precisa e affidabile[17] come si vede nella Figura 4.4. I landmark estratti da Mediapipe vengono convertiti in una pointcloud all'interno di uno spazio tridimensionale. Questo passaggio consente di rappresentare in modo tridimensionale i riferimenti estratti dall'immagine bidimensionale. FLAME[8] genera una mesh 3D che include 68 landmark del viso, seguendo lo standard dlib[18]. Questi landmark costituiscono un sottoinsieme di quelli estratti da Mediapipe, quindi mediante l'utilizzo di open3d[19] abbiamo eseguito una sequenza di operazioni, tra cui ridimensionamento, rotazione e traslazione, che hanno reso possibile allineare la pointcloud dei landmark estratti con le mesh generate da FLAME[8]. L'allineamento della pointcloud con le mesh FLAME consente di combinare le informazioni ricavate dai landmark con la struttura tridimensionale del modello FLAME, producendo risultati più accurati e dettagliati.

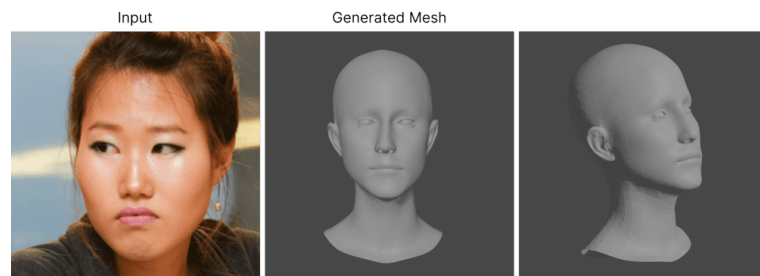


**Figura 4.4:** Preview landmark Mediapipe

### Algoritmo Genetico

Attraverso l'utilizzo di un algoritmo genetico, sono stati ottimizzati i parametri di FLAME per creare una mesh simile ai landmark di partenza. L'implementazione

dell'algoritmo genetico è stata realizzata utilizzando Pygad[20]. È stata creata una funzione di fitness per il GA che, dati la mesh, i riferimenti della mesh e i landmark Mediapipe[17], restituisce un valore di prossimità media tra i riferimenti della mesh e i landmark corrispondenti. E per i landmark rimanenti, viene restituita la distanza media tra la posizione dei punti nella pointcloud dei landmark e le facce della mesh più vicine. Attraverso una ricerca grid search, sono stati selezionati i migliori iperparametri per il GA. Sono state condotte varie analisi quantitative e qualitative per giungere alla configurazione finale del GA. Quest'ultimo viene iterato per 100 epoche su una popolazione di 16 individui per epoca, dei quali solo i migliori 4 vengono selezionati per il crossover. Dalla grid search e dall'analisi qualitativa è emerso che i migliori risultati si ottengono quando i parametri dati in input a FLAME[8] sono limitati tra -2 e +2, quindi il risultato del GA è stato costretto in questo range di valori. Come si può vedere dalla Figura 4.5 attraverso il GA, è stata generata una mesh della testa in posizione neutrale con espressione neutra. Come step finale, è stato eseguito il processo di UV unwrapping, che consiste nel proiettare la superficie della mesh in uno spazio 2D che risulta essere fondamentale per l'applicazione della texture sulla mesh appena generata.



**Figura 4.5:** Mesh generata utilizzando il GA

### 4.2.3 Texture

#### Dataset

Nella creazione del dataset per l'addestramento di reti neurali, è stato fondamentale scegliere un dataset di partenza di alta qualità per ottenere risultati accurati e significativi. In questo contesto, è stato selezionato il Flickr-Faces-HQ dataset[10] come punto di partenza. Questo dataset è composto da 70000 volti in varie posizioni

e condizioni di illuminazione, offrendo una vasta gamma di dati per l'analisi e la generazione di texture. Per la generazione delle texture, ci si è basato sul lavoro di Bai et al.[9], che rappresenta lo stato dell'arte attuale in questo campo. Questo dataset contiene più di 50000 texture[9], che come si vede nella Figura 4.6 possono essere utilizzate come target per i processi di generazione e analisi.



**Figura 4.6:** Elemento nel dataset utilizzato come target[9]

### Modello Deep Learning

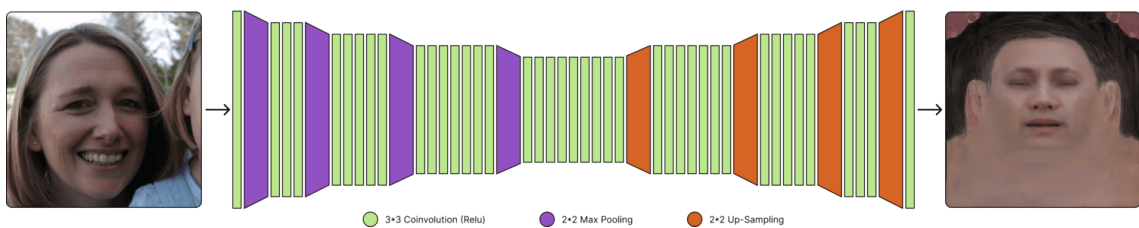
Per la generazione delle texture, è stato scelto un modello di Deep Learning per la sua velocità in fase di inferenza, che prende in input un'immagine di una persona e ne genera la texture. Questo approccio è stato adottato considerando le risorse computazionali limitate disponibili per questo prototipo. Per adattarsi alle limitazioni delle risorse, è stata creata una versione ridotta del dataset iniziale contenente 25000 elementi. Le immagini e le texture sono state ridotte di dimensioni da 1024x1024 pixel a 256x256[9][10]. Il dataset è stato suddiviso in

- 70% per il train-set,
- 15% per il validation-set,
- 15% per il test-set.

Il modello di Deep Learning è stato implementato utilizzando Tensorflow[21] e Keras. È stato progettato per prendere in input un'immagine di una persona con posizione,



espressione e luci generiche e generare la texture corrispondente. L'architettura scelta per il modello di Deep Learning è l'auto-encoder. Questa scelta è stata motivata dalle eccellenti capacità dell'architettura nell'estrazione di feature da immagini e nella generazione di immagini realistiche. Come si può vedere nella Figura 4.7, la rete neurale è composta da 50 livelli distinti, di cui 48 nascosti e alternati con livelli che eseguono il max pooling e l'upsampling per un totale di oltre 12 Milioni di parametri. I livelli convoluzionali utilizzano filtri convoluzionali di dimensione 3x3 con funzione di attivazione relu. Il max pooling e l'upsampling sono eseguiti con filtri 2x2. Durante la fase di training, è stato utilizzato un batch size di 32 elementi per ottenere una stima più precisa del gradiente e favorire una convergenza stabile del modello. Il modello è stato addestrato per 50 epoche sull'intero train-set e successivamente validato sull'intero validation-set. Successivamente, è stata verificata l'effettiva capacità di generalizzazione del modello testandolo sul test-set per valutare le performance in un contesto più realistico, utilizzando per valutare il modello come la funzione Binary cross entropy e come ulteriore metrica il MAE.



**Figura 4.7:** Architettura Autoencoder [9][10]

#### 4.2.4 Inferenza

Nella fase di inferenza, viene selezionata un'immagine di un individuo che funge da input per le pipeline che genereranno texture e mesh. L'immagine scelta viene utilizzata come input per la pipeline che, tramite l'uso di un algoritmo genetico (GA), crea la mesh della testa dell'individuo rappresentato nell'immagine. L'immagine viene anche fornita alla Rete Neurale dedicata alla generazione di texture. Questa rete restituisce una texture iniziale di dimensioni 256x256. Successivamente, viene eseguito un upscaling delle texture generate per portarle a una risoluzione di 1024x1024. Questo processo avviene grazie a un modello pre-addestrato di upscaling

delle immagini creato dal lavoro di Lim et al.[22], il quale migliora la risoluzione delle texture finali. Infine, la texture ottenuta viene applicata alla mesh creata, consentendo di visualizzare il risultato finale con una maggiore definizione e qualità visiva.

---

### Sperimentazione dell'approccio

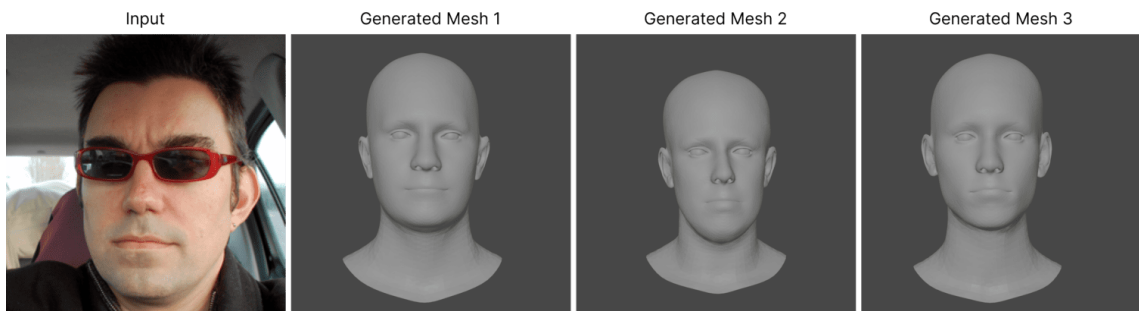
---

#### 5.1 Metodo di Sperimentazione

In un'analisi volta a valutare i risultati ottenuti dalle mesh texturizzate generate partendo da un'immagine di un volto con posa, espressioni e luci generiche, è emerso che non esistono benchmark specifici per questo tipo di valutazione. Pertanto, è stato adottato un approccio che prevede la valutazione delle mesh e delle texture sia singolarmente che combinati per ottenere una visione completa dei risultati. Le mesh generate sono state esaminate attentamente per valutare la qualità e la fedeltà rispetto all'immagine di partenza. Le texture sono state analizzate per verificare la corrispondenza con le caratteristiche del volto e la resa complessiva. Sono stati considerati parametri come la definizione dei dettagli, la coerenza dei colori e la naturalezza delle superfici. Dopo aver valutato singolarmente mesh e texture, è stato eseguito un processo di combinazione per integrare le due componenti. L'obiettivo è stato verificare come l'interazione tra mesh e texture influenzasse l'aspetto complessivo e la somiglianza con l'immagine di riferimento.

## 5.2 Valutazione Mesh

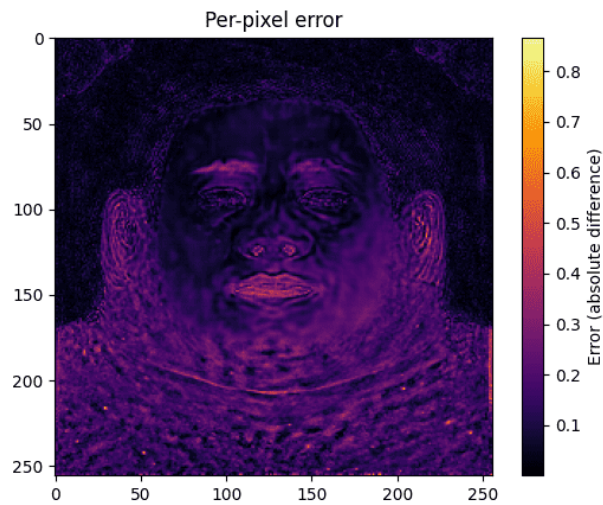
Molti benchmark per la validazione di modelli che elaborano immagini per generare mesh del volto o dell'intera testa considerano anche le espressioni facciali. Tuttavia, poiché il metodo proposto genera mesh in espressione neutra, mancano benchmark in grado di valutarle in modo analitico. Per valutare la qualità delle mesh generate, è possibile inizialmente verificare la consistenza generando molteplici mesh da un'unica immagine. Tutte le mesh generate hanno utilizzato gli stessi settaggi dell'algoritmo genetico, compreso il numero di epoche. Come si può vedere dalla Figura 5.1, le mesh generate mostrano variazioni significative tra loro, soprattutto per quanto riguarda la forma della testa e del collo. Questa discrepanza può essere attribuita al fatto che i landmark Mediapipe[17], da cui sono stati ottenuti i riferimenti tridimensionali, sono concentrati solo nella sezione del volto. Le discrepanze nella zona del volto sono dovute all'algoritmo genetico che, ad ogni epoca, effettua scelte casuali che influenzano sia la coerenza dei risultati ottenuti che la loro qualità finale. Questa criticità impatta sulla consistenza e sulla precisione delle mesh generate.



**Figura 5.1:** Visualizzazione incoerenze mesh

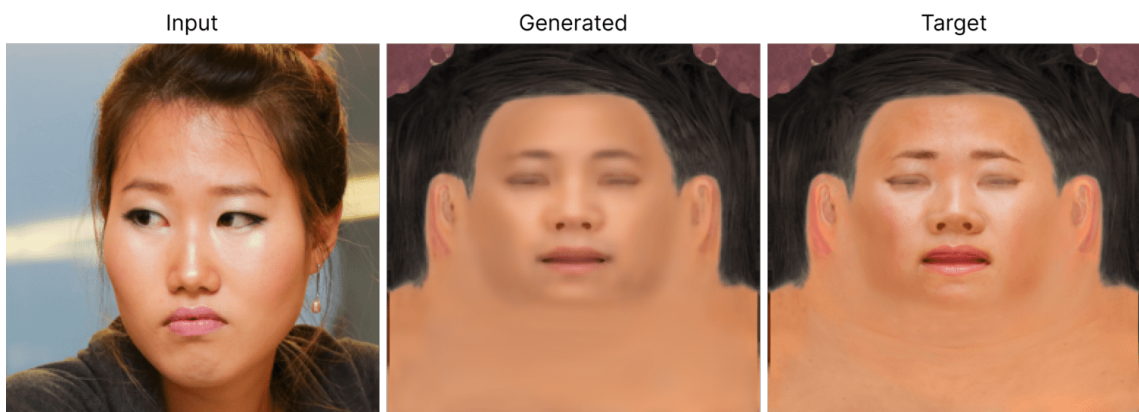
## 5.3 Valutazione Texture

Attualmente non esistono benchmark definiti per valutare le texture in modo oggettivo, il che rende difficile valutarne la qualità in maniera accurata. Tuttavia, è possibile confrontare le texture generate per il test-set con i target di riferimento per ottenere un'analisi più approfondita. Dall'analisi delle texture generate (Figura 5.3) e dalla heatmap pixel per pixel degli errori (Figura 5.2), emerge che i risultati



**Figura 5.2:** Heatmap degli errori su tutto il Test-set

presentano alcune criticità, soprattutto nella sezione del volto. Come si evince dalla Heatmap degli errori Figura 5.2 in queste zone, mancano dettagli cruciali come rughe o barba, oppure dalla Figura 5.3 si può notare che alcuni dettagli della colorazione del volto come le sfumature della pelle o delle labbra risultano non presenti o solo accennati, evidenziando una mancanza di definizione. Questo difetto è attribuibile al fatto che la rete neurale utilizzata potrebbe non essere stata addestrata per un numero sufficiente di epoche o potrebbe non essere abbastanza profonda per affrontare efficacemente il compito assegnato, generando texture approssimative.



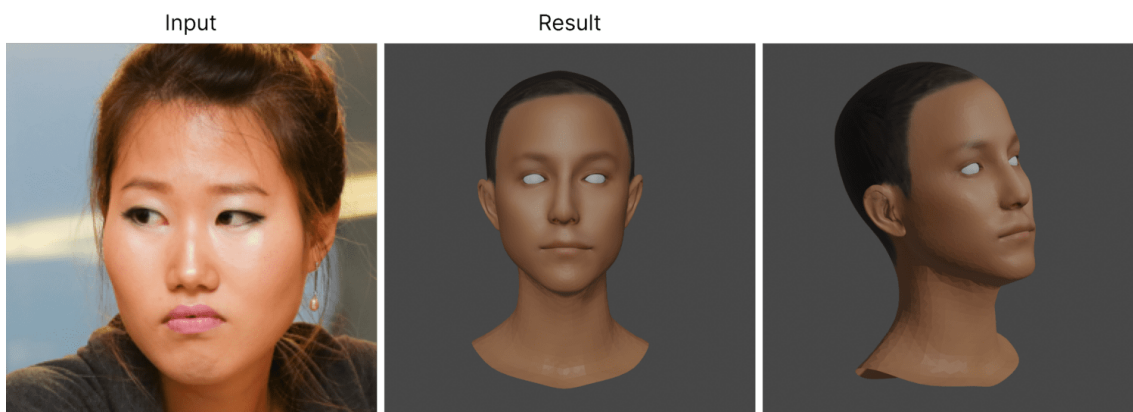
**Figura 5.3:** Risultato generazione texture

Diversamente, la sezione relativa ai capelli e alle aree circostanti delle texture risulta essere particolarmente precisa. Questo risultato positivo è dovuto al dataset

utilizzato che si basa sul lavoro di Bai et al.[9], il quale contribuisce a definire con precisione le zone dei capelli e dei contorni, rendendo le texture uniformi e coerenti in tutti i target analizzati[9]. Nonostante le limitazioni riscontrate nella sezione del volto, è interessante notare che il modello di deep learning riesce comunque a generare texture di qualità anche quando le posizioni dei soggetti nelle immagini di input sono molto diverse tra di loro. Questa capacità di approssimazione suggerisce una certa flessibilità e adattabilità del modello neurale nell'affrontare variazioni significative nei dati di input.

## 5.4 Valutazione Avatar Completo

Successivamente all'applicazione delle texture sulle mesh, si può notare che il risultato è uniforme e coerente su tutte le parti e con l'immagine di partenza. Un aspetto degno di nota che si può notare nella Figura 5.4 è l'allineamento preciso dei vari elementi delle texture con le corrispondenti parti delle mesh sottostanti come gli occhi, la bocca, il naso e altri dettagli che risultano essere perfettamente allineati tra la texture e la struttura delle mesh, contribuendo a creare un effetto visivo armonioso e realistico.



**Figura 5.4:** Risultato Completo

---

### Conclusioni

---

Questa tesi aveva l'obiettivo di sperimentare e trovare un metodo rapido per creare mesh e texture realistiche da poter usare come avatar virtuali partendo da una singola immagine. In base ai risultati ottenuti, è possibile concludere che l'obiettivo prefissato è stato raggiunto con successo. La capacità di generare mesh e texture realistiche a partire da un'unica immagine rappresenta un importante passo avanti nel campo della creazione di avatar virtuali.

#### 6.1 Criticità Riscontrate

L'implementazione di algoritmi genetici per la generazione dei parametri FLAME[8] presenta delle criticità che richiedono attenzione. Tra queste criticità, si evidenzia l'inconsistenza delle generazioni delle mesh che risulta essere una delle principali criticità riscontrate nella generazione delle mesh. Questo problema richiede una verifica manuale approfondita per garantire la coerenza e la qualità delle mesh generate, il che rende il sistema sviluppato non particolarmente adatto a sistemi completamente autonomi come nel caso di un metaverso. L'origine di questa inconsistenza è legata all'implementazione degli algoritmi genetici stessi, che vista la loro natura casuale possono generare risultati vari e non sempre uniformi. Un'altra criticità riscontrata

riguarda la generazione della sezione della testa e del collo delle texture, che mostra un'uniformità e ripetitività tra le varie texture generate. Questo problema è da attribuirsi al dataset che si basa sul lavoro di Bai et al.[9] che durante il processo di generazione delle texture approssima questa sezione della testa.

## 6.2 Sviluppi Futuri

Il progetto potrebbe vedere diversi sviluppi futuri che potrebbero migliorare l'efficienza e le prestazioni. Un possibile sviluppo futuro potrebbe riguardare l'adozione di un modello di Machine Learning per la generazione dei parametri della mesh. Questo approccio potrebbe portare a una maggiore velocità e prevedibilità rispetto agli algoritmi genetici attualmente utilizzati, inoltre nella fase di inferenza, un modello di Machine Learning potrebbe essere in grado di generare i parametri in modo più rapido, consentendo una migliore ottimizzazione del processo. Un altro possibile miglioramento potrebbe riguardare l'adozione di metodi alternativi di estrazione delle feature del volto dalle immagini, diversi dall'utilizzo dei landmark che potrebbe includere anche la forma della testa e del collo, ad esempio tramite l'utilizzo di una depth-map. Questo approccio potrebbe portare a una maggiore precisione nella generazione delle mesh e a una migliore rappresentazione dei tratti distintivi del volto. Infine, un ulteriore sviluppo potenziale potrebbe riguardare il miglioramento della generazione delle texture. Questo potrebbe avvenire attraverso l'ottimizzazione o la creazione di un'alternativa al lavoro di Bai et al.[9]. Migliorare la generazione delle texture potrebbe portare a risultati visivi più realistici e dettagliati degli avatar generati.

## 6.3 Usi alternativi

L'approccio sviluppato è stato concepito fin dall'inizio per il metaverso, con l'obiettivo di permettere a chiunque di creare avatar senza la necessità di competenze tecniche specifiche. Tuttavia, i risultati ottenuti dimostrano che l'approccio implementato può ottenere ottimi risultati anche al di fuori del solo metaverso, come nella stampa 3D, nei VFX, nello sviluppo di videogiochi e nella computer grafica. Inoltre,



i risultati potrebbero essere utili anche per i professionisti della modellazione 3D, offrendo un'avanzata base di partenza che potrebbe far risparmiare tempo nella fase di modellazione.

---

## Bibliografia

---

- [1] M. Inc, "Introducing 'horizon workrooms': Remote collaboration reimaged," 2021, <https://www.meta.com/it-it/blog/quest/workrooms/> [Accessed: (29/04/2024)]. (Citato alle pagine iii e 7)
- [2] D. Nield, "Choose the right vr and ar gear for you," 2019, <https://www.popsci.com/choose-right-vr-gear-for-you/> [Accessed: (29/04/2024)]. (Citato alle pagine iii e 8)
- [3] Wikipedia, "Mesh poligonale — wikipedia, l'enciclopedia libera," [https://it.wikipedia.org/wiki/Mesh\\_poligonale](https://it.wikipedia.org/wiki/Mesh_poligonale) [Accessed: (29/04/2024)]. (Citato alle pagine iii e 9)
- [4] —, "Mappatura uv — wikipedia, l'enciclopedia libera," [https://it.wikipedia.org/wiki/Mappatura\\_UV](https://it.wikipedia.org/wiki/Mappatura_UV) [Accessed: (29/04/2024)]. (Citato alle pagine iii e 9)
- [5] Meta, "Customize your meta avatar with new body shapes, hair and clothing," 2024, <https://about.fb.com/news/2023/04/meta-avatars-new-body-shapes-hair-clothing/> [Accessed: (29/04/2024)]. (Citato alle pagine iii e 11)
- [6] E. Roth, "Apple's vision pro headset will turn you into a digital avatar when facetimeing," 2023, <https://www.theverge.com/2023/6/5/23750096/>

- apple-vision-pro-headset-persona-facetime [Accessed: (29/04/2024)]. (Citato alle pagine iii e 12)
- [7] EpicGames, "Metahuman," 2024, <https://www.unrealengine.com/en-US/metahuman> [Accessed: (29/04/2024)]. (Citato alle pagine iii, 12 e 13)
- [8] T. Li, T. Bolkart, M. J. Black, H. Li, and J. Romero, "Learning a model of facial shape and expression from 4D scans," *ACM Transactions on Graphics, (Proc. SIGGRAPH Asia)*, vol. 36, no. 6, pp. 194:1–194:17, 2017. [Online]. Available: <https://doi.org/10.1145/3130800.3130813> (Citato alle pagine iii, 15, 17, 18, 19, 21, 22, 23 e 31)
- [9] H. Bai, D. Kang, H. Zhang, J. Pan, and L. Bao, "Ffhq-uv: Normalized facial uv-texture dataset for 3d face reconstruction," 2023. (Citato alle pagine iii, 16, 17, 18, 24, 25, 30 e 32)
- [10] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," 2019. (Citato alle pagine iii, 23, 24 e 25)
- [11] Reallusion, "Character creator," 2024, <https://www.reallusion.com/character-creator/> [Accessed: (29/04/2024)]. (Citato a pagina 12)
- [12] Y. Feng, F. Wu, X. Shao, Y. Wang, and X. Zhou, "Joint 3d face reconstruction and dense alignment with position map regression network," 2018. (Citato alle pagine 14 e 16)
- [13] A. Soltani, H. Haibin, J. Wu, T. Kulkarni, and J. Tenenbaum, "Synthesizing 3d shapes via modeling multi-view depth maps and silhouettes with deep generative networks," 07 2017. (Citato a pagina 15)
- [14] S. Kamyab and Z. Azimifar, "Deep-mds framework for recovering the 3d shape of 2d landmarks from a single image," 2022. (Citato alle pagine 15, 18 e 22)
- [15] S. Sanyal, T. Bolkart, H. Feng, and M. Black, "Learning to regress 3D face shape and expression from an image without 3D supervision," in *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2019, pp. 7763–7772. (Citato alle pagine 15 e 21)

- [16] P. Henderson, V. Tsiminaki, and C. H. Lampert, "Leveraging 2d data to learn textured 3d mesh generation," 2020. (Citato alle pagine 15 e 18)
- [17] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. Yong, J. Lee, W.-T. Chang, W. Hua, M. Georg, and M. Grundmann, "Mediapipe: A framework for perceiving and processing reality," in *Third Workshop on Computer Vision for AR/VR at IEEE Computer Vision and Pattern Recognition (CVPR) 2019*, 2019. [Online]. Available: [https://mixedreality.cs.cornell.edu/s/NewTitle\\_May1\\_MediaPipe\\_CVPR\\_CV4ARVR\\_Workshop\\_2019.pdf](https://mixedreality.cs.cornell.edu/s/NewTitle_May1_MediaPipe_CVPR_CV4ARVR_Workshop_2019.pdf) (Citato alle pagine 22, 23 e 28)
- [18] D. King, "Dlib-ml: A machine learning toolkit," *Journal of Machine Learning Research*, vol. 10, pp. 1755–1758, 07 2009. (Citato a pagina 22)
- [19] Q.-Y. Zhou, J. Park, and V. Koltun, "Open3D: A modern library for 3D data processing," *arXiv:1801.09847*, 2018. (Citato a pagina 22)
- [20] A. F. Gad, "Pygad: An intuitive genetic algorithm python library," 2021. (Citato a pagina 23)
- [21] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, software available from [tensorflow.org](https://www.tensorflow.org/). [Online]. Available: <https://www.tensorflow.org/> (Citato a pagina 24)
- [22] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017. (Citato a pagina 26)

*Questa tesi ha contribuito a piantare un albero in Madagascar tramite il progetto Treedom.*

<https://www.treedom.net/it/user/sesalab/event/sesa-random-forest>