Università degli Studi di Milano
Facoltà di Scienze e Tecnologie
Corso di Laurea in Informatica (L-31)

ANALISI QUANTITATIVA E PERCETTIVA DI VIDEO CREATI CON GENERATORI AI

Relatore: Prof. Raffaella Lanzarotti

Correlatore: Prof. Andrea Gaggioli

Tesi di:
Federico COSCIA
Matricola: 977772

Anno Accademico 2023-2024

a Celestina

Indice

Introduzione

Capitolo 1

Generazione dei video fake

1.1 Funzionamento

1.2 Valutazione delle soluzioni disponibili

Per la generazione dei video fake, sono stati valutati tre applicativi diversi, forniti come Software-as-a-Service (SaaS):

- DupDub.com
- Synthesia.io
- HeyGen.com

I criteri che sono stati valutati sono: la naturalezza dei movimenti generati, l'estensione di questi ultimi, la qualità del lip-sync[1], la qualità e la naturalezza della voce parlata generata, e il grado di realismo generale dato dai video generati. Vediamo per ordine i punti di forza e di debolezza identificati di ognuno, e come si è pervenuti alla scelta finale.

1.2.1 DupDub

DupDub si classifica come un prodotto "Talking-Photo". A partire da una fotografia di un persona, genera il movimento dei muscoli facciali e delle labbra per simulare il parlato. DupDub trova i suoi punti di forza nell'essere molto semplice, ma è stato valutato come troppo semplice per gli scopi di questa ricerca. La più grande limitazione è data dalla limitatezza dei movimenti, limitandosi appunto a generare solo movimenti dei muscoli facciali, e a malapena movimenti della testa, rendendo il

[1] sincronizzazione tra il movimento delle labbra di un soggetto e il suono delle parole pronunciate.

risultato finale poco convincente e innaturale. Inoltre, tutti gli avatar forniti dalla piattaforma per la generazione dei video sono chiaramente soggetti non reali, bensì generati a loro volta tramite IA.

1.2.2 Synthesia.io

Rispetto al precedente, Synthesia.io si mostra molto più capace. Fornisce avatar in mezzo busto, ed è in grado di generare movimenti del viso, della testa, e anche del corpo, producendo risultati decisamente più naturali di DupDub. Gli avatar forniti sembrano essere stati generati a partire da persone reali, ed il servizio offre la possibilità di generare avatar personali. I punti di debolezza individuati sono stati: la qualità del lip-sync e la qualità delle voci generate. In particolare, risultava frequente il disallineamento tra il movimento delle labbra dell'avatar e il suono della voce generato. La voce inoltre è stata valutata come poco espressiva e poco naturale. Vedremo in realtà come questi sono spesso i punti più deboli di questa tecnologia.

Nonostante questo, tale servizio sembrava un buon candidato per la ricerca, ma è stato scartato in base al piano offerto, in quanto offriva un servizio ad abbonamento basato su minuti.

1.2.3 HeyGen

Sin dal primo sguardo, HeyGen.com si è dimostrato essere al di sopra di tutti gli altri, offrendo anche la possibilità di generare video su sfondi reali, angolazioni diverse dello stesso avatar, e implementando movimenti del corpo avanzati come il movimento delle braccia e il gesticolamento delle mani. Gli avatar sono costruiti a partire da un video di riferimento del soggetto, il che li conferisce la possibilità di apprendere ed emulare i movimenti della persona inquadrata, producendo un risultato più naturale e realistico. La voce è stata identificata come un punto debole, ma non perché di scarsa qualità. Le voci generate hanno un timbro molto pulito, secco e "radiofonico", che però può risultare innaturale se utilizzate in un video a sfondo reale, dove il suono della voce potrebbe non essere conforme all'acustica della stanza rappresentata. Questo problema però non si presenta negli avatar tradizionali, i quali sono privi di sfondo. Inoltre, la piattaforma si è dimostrata essere in constante evoluzione e sviluppo, arricchendo il suo catalogo di funzionalità, avatar e di voci durante il periodo di valutazione.

Per queste ragioni, tra le opzioni valutate, HeyGen è stato valutato come il migliore, in termini di qualità e naturalezza dei risultati prodotti, ed è stato quindi scelto come soluzione per la nostra ricerca. Un altro fattore che sicuramente ha giocato a

suo favore è stato anche il piano offerto, il quale ci ha permesso di generare infiniti video durante il periodo di abbonamento, posto che questi fossero sotto i cinque minuti di durata.

1.3 Video real

1.3.1 Scelta dei video real

1.3.2 Pre-Processing

1.4 Generazione dei video fake

Capitolo 2

Setting di acquisizione

2.1 Modalità di acquisizione

2.2 Estrazione delle feature

2.2.1 Video

2.2.2 Dati fisiologici

2.2.3 Eye-tracking

Capitolo 3

Protocollo di acquisizione

3.1 Stesura del protocollo

3.2 Sviluppo dell'interfaccia

3.3 Salvataggio dei dati raccolti

Capitolo 4

Analisi dei dati acquisiti

Conclusioni

Bibliografia

Ringraziamenti