Report

Table of contents

Dataset Split

- Il dataset è stato diviso in due porzioni, una destinata al training(stima del *threshold*) e una per la validazione;
- Entrambe le porzioni contengono lo stesso numero di identità, tuttavia hanno immagini diverse;
  - **TRAIN** → 15 immagini *real*, 10 immagini *fake*
  - **VAL** → 10 immagini *real*, 5 immagini *fake*

Re-Training e Validation

- Tutti i modelli sono stati addestrati e validati sul nuovo dataset. Seguono in questa sezione i risultati ottenuti nel training e nella validation;
- Successivamente, analizzeremo le strategie impiegate per provare ad aumentare l'accuracy;

Model	Threshold (test data)	F1 Score (train)	AUC	Train Accuracy	Val Accuracy	Validation F1 Score
VGG-Face	0.5	0.7654	0.7791	0.7486	0.7318	0.7432
Facenet	0.7	0.8065	0.8392	0.7932	0.7790	0.7881
Facenet512	0.72	0.777	0.8186	0.7832	0.7763	0.7597
OpenFace	0.45	0.6669	0.6262	0.5045	0.5081	0.6666
DeepFace	0.2	0.6667	0.6411	0.5	-	-
DeepID	0.2	0.6667	0.554	0.5	-	-
ArcFace	0.58	0.8406	0.7225	0.83	0.8154	0.82332
SFace	0.54	0.7529	0.7657	0.75	0.7281	0.7187
GhostFaceNet	0.5	0.8113	0.7498	0.8232	0.7891	0.7721

Quorum Size	Accuracy	F1 Score
Top 3	.8045	.7939
Top 5	0.786	0.7945

Weighted Distances

- In questa strategia ho provato ad inserire più funzioni di distanza, ciascuna associata ad un peso;
- La similarity totale era data dalla somma pesata di tutte le distanze;

💡 Questa soluzione ha fallito in quanto non tutte le metriche di distanza variano tra 0 e 1, di conseguenza alcune metriche impattavano in maniera troppo importante la similarity;

Clip

- In questa strategia ho provato a impiegare solo gli embeddings di Clip per il calcolo della similarity

💡 Si ottiene 50% di accuracy dimostrando che clip non riesce a catturare informazioni sull'identità

Clip + FaceRec Weighted Distance

- In questa strategia ho impiegato sia Clip e sia un Modello di face recognition (GhostFaceNet);
- Ho definito un peso f_w associato alla similarity del Modello di face recognition, e il complementare associato a clip, definito come $f_c = 1 - f_w$
- La similarity di una coppia di immagini è data dalla somma pesata delle similarity degli embeddings di Clip e di GhostFaceNet;

Threshold (test data)	Face Weight	Clip Weight	F1 Score (train)	AUC	Train Accuracy	Val Accuracy	Validation F1 Score
0.26	0.6	0.4	0.6689	0.4923	0.505	-	-
0.22	0.75	0.25	0.6665	0.4919	0.505		
0.52	0.8	0.2	0.8079	0.8406	0.805		
0.52	0.9	0.1	0.808	0.8086	0.8177		

💡 Anche in questo caso Clip non riesce a aiutare il modello a distinguere le immagini. Difatti con il tendere di di $f_w \to 1$, l'accuracy migliora tendendo al valore ottimale che si ottiene senza clip;

Threshold Sensitivity

- In questo approccio ho cambiato la pipeline di riconoscimento, inserendo due nuovi parametri:
  - La sensibilità del threshold S
  - Il numero minimo di match richiesti $MinID$

- Data un'identità e un treshold t, definiamo $MaxSim$ con la *similarity* più alta raggiunta confrontando l'immagine con il Reference Set:
  - Se $MaxSim > t * S$, il modello classificherà l'immagine come Real

- Altrimenti, richiediamo che esistano almeno $MinID$ immagini la cui similarity sia maggiore di t per classificare il sample Real. Se questa condizione non si verifica, l'immagine sarà classificata come Fake.

Model	Threshold (test data)	Sensitivity	Min Matches	F1 Score (train)	Train Accuracy	Val Accuracy	Validation F1 Score
Vgg-facenet	0.4	1.15	9	0.7574	0.7222	0.71	0.7421
Facenet	0.66	1.1	7	0.8085	0.805	0.7980	0.7940
Facenet512	0.68	1.05	2	0.7779	0.776	0.7955	0.7878
GhostFace Net	0.45	1.1	3	0.813	0.8218	0.7943	0.7826
ArcFace	0.45	1.3	5	0.8438	0.8273	0.8234	0.8358
Sface	0.5	1.1	7	0.7532	0.7573	0.7390	0.724

Quorum Size	Accuracy	F1 Score
Top 3	0.6972	0.7676
Top 5	0.6618	0.7472

L'approccio migliora minimamente l'accuracy (1%-2%) peggiorando però l'accuracy dell'ensemble.