

SKIN DETECTION

ALESSIO MASI

February 2022

1 Introduzione

La skin detection ha un range molto ampio di utilizzi, che vanno dal riconoscimento di alcune parti del corpo come ad esempio i volti oppure all'applicazione su contesti che riguardano la classificazione di alcuni contenuti potenzialmente offensivi come le immagini di nudo.

I principali limiti di utilizzo di questa tecnica sono che essa può essere affetta da cambiamenti di luminosità e sfondi o oggetti che hanno la stessa tonalità della pelle.

Per classificare la pelle nelle immagini ci sono vari approcci e in questo caso verrà utilizzato Multi-Layer Perceptron (MLP) e questo problema verrà modellato nello spazio di colori HLS. I risultati sono ottenuti sulla classificazione di 750 frames provenienti da 12 video reperibili online.

2 MLP

MLP è una rete neurale che connette più layer. Due sono fissi e sono sempre presenti ovvero quello di input e di output, poi ci sono altri layer chiamati hidden layers che possono essere di un numero arbitrario così come può essere arbitrario il numero di neuroni presenti in ogni hidden layer. I vari layer con i loro neuroni (nodi) vengono mappati in un grafo orientato, ciò significa che si può andare in una sola direzione (dall'input verso l'output). Ogni nodo dei layer nascosti ha una funzione di attivazione non lineare. MLP è una tecnica di apprendimento supervisionata in quanto utilizza la backpropagation per allenare la rete. La MLP è una tecnica largamente utilizzata nel riconoscimento di immagini e in generale in tutti quei problemi che non sono linearmente separabili.

Nell'utilizzo di MLP particolare attenzione va posta al numero di hidden layer utilizzati e anche al numero di neuroni presente in ogni layer. Un numero di layer e neuroni troppo elevato potrebbe causare un over-training che non porterebbe poi ad un grande incremento a livello di precisione ottenuta ma di contro farebbe aumentare molto il tempo di esecuzione. D'altro canto invece un numero troppo piccolo di neuroni in rapporto alle dimensioni dei dati in ingresso porterebbe a delle capacità ridotte e quindi un peggioramento dei risultati ottenuti.

2.1 Risultati attesi

Essendo questo esperimento con un dataset uguale già stato condotto da R.Khan[1] abbiamo una idea sui risultati che dovremmo ottenere. In quel caso avevano ottenuto un accuracy score pari a 0.816. Nel nostro caso useremo un dataset ridotto e molto vario quindi è attendibile un risultato simile ma probabilmente meno preciso. Come ulteriore risultato, visivo verranno prodotte due immagini

corrispondenti a due frame distinti a cui sarà applicata una maschera che sarà trasparente nelle zone in cui è stata rilevata pelle mentre sarà nera nelle altre zone.

3 Descrizione esperimento

Per svolgere questo esperimento sono stati usati 750 frames provenienti da 12 video presenti su un sito online[2]. I video contengono persone di etnia diversa e varie condizioni di luce e con una grande varietà di scene e ambienti. La valutazione dei pixel è stata effettuata utilizzando i frame ricavati dai video a colori ed etichettati con bianco o nero a seconda che fosse presente o meno della pelle. Per effettuare l'allenamento sono stati utilizzati altri video presenti all'interno dello stesso dataset ma che corrispondono ad una maschera creata manualmente in cui le zone in cui è presente pelle sono bianche le altre sono nere.

Come risultato è stata prodotta una matrice di confusione e due immagini a cui è stata applicata una maschera sulla base della predizione effettuata dopo la fase di apprendimento.

3.1 Matrice di confusione

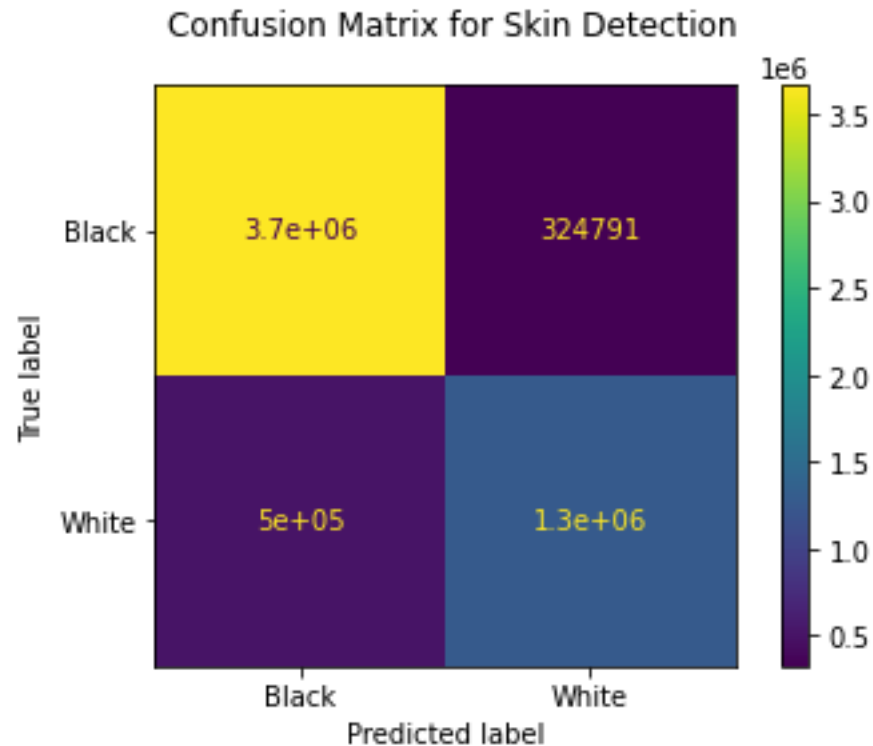


Figura 1: Matrice di confusione che mostra come ci siano più errori nella valutazione dei pixel che contengono effettivamente pelle

Come si può notare dal grafico ottenuto è più frequente etichettare erroneamente un pixel come non contenente pelle quando in realtà è il contrario. Questo è anche facilmente riscontrabile nelle due foto prodotte presenti di seguito.

3.2 Foto Prodotte



4 Conclusioni

Per concludere nonostante un accuracy score prodotto pari a 0.85 che è piuttosto elevato, quando arriviamo all'applicazione effettiva possiamo notare come i risultati non siano molto soddisfacenti. Questo può essere figlio di una scelta non ottimale del dataset di ingresso in quanto molto probabilmente sono stati usati video contenenti situazioni troppo diverse tra loro senza dare più opzioni per le stesse situazioni. Quindi a discapito di un tempo di esecuzione maggiore sarebbe meglio fornire un dataset più bilanciato e che fornisca più esempi della stessa situazione per avere risultati più precisi.

5 References

- [1] SKIN DETECTION: A RANDOM FOREST APPROACH. Rehanullah Khan, Allan Hanbury, Julian Stöttinger
- [2] A 3D-Polar Coordinate Colour Representation Well Adapted to Image Analysis; Allan Hanbury
- [3] Proceedings of the 5th International Symposium on Visual Computing (ISVC), Julian Stöttinger, Allan Hanbury, Christian Liensberger, Rehanulla Khan , Las Vegas, NV, Nov 30 - Dec 2, 2009.