**Riconoscimento di volti**

Simeoni Ildebrando

Scanu Mario

Marescalco Christian   
Progetto ESM - gruppo 7

(a.a. 2020-2021)

**1 - Introduzione**

La tecnica del riconoscimento facciale è oggi sempre più utilizzata non solo per scopi di sicurezza, ma anche in molti altri ambiti della vita quotidiana; proprio tale ricchezza di utilizzi, unita allo stato della tecnologia, ne ha permesso una diffusione capillare, tanto che uno studio pubblicato nel giugno 2019 stima che entro il 2024 il mercato globale del riconoscimento facciale genererebbe 7 miliardi di dollari di entrate.

E’ una tecnica biometrica atta ad identificare in modo univoco una persona, ed è attualmente tra le applicazioni di machine learning più diffuse in assoluto.

Di seguito sono riportati alcuni dei casi di utilizzo di tale teconlogia:

* Sblocco dello smartphone; è necessario che l’algoritmo di face recognition abbia passato standard e test di accuratezza (e quindi sicurezza) molto elevati.

Grazie a questa tecnologia, il volto sta assumendo il ruolo di una nuova impronta digitale per riconoscere le persone in modo totalmente univoco.

* Riconoscimento di persone scomparse; alcune autorità e agenzie di investigazione private utilizzano il face recognition per individuare persone scomparse in una folla, in macchina o nei luoghi pubblici a partire da video o immagini di sorveglianza.
* Tecnologie antifurto; soprattutto per i grandi retailer, i furti nei negozi fisici sono un fenomeno che costa miliardi di dollari all’anno. Grazie ai video delle telecamere di sorveglianza è possibile analizzare e riconoscere potenziali criminali a partire da opportuni database.
* Identificazione di persone sui social media; ad esempio per la gestione della reputazione di persone famose divenuta di interesse in particolar modo con l’esplosione di forme di contenuto virali.
* Ricerca per immagini via browser.

Difficoltà relative al problema preso in esame sono:

* Il cambiamento delle condizioni di luce, che causa una sfida significativa per il riconoscimento automatico dei volti e può avere un impatto significativo sui suoi risultati.
* Posa del volto o diverso punto di vista della fotocamera, che possono invariabilmente causare cambiamenti nell'aspetto del viso e generare variazioni intra-classe facendo cadere drasticamente i tassi di riconoscimento automatico dei volti.
* Occlusione del volto, causata da oggetti in scenari reali, ad esempio per immagini estratte da videocamere di sorveglianza.
* Differenti espressioni umane, da tenere in considerazione poiché possono portare a una notevole variazione dei parametri biometrici di un volto in esame.
* Bassa risoluzione, particolarmente sentito nei casi di immagini estratte da camere non di qualità o in condizioni ambientali non ottimali.

Figura 4



Figura 1

**2 - Approccio implementato**

L’approccio sviluppato, rappresentato graficamente in Figura 1, può essere descritto come segue:

1. Download dei dati

Sono state scaricate le immagini appartenenti al dataset precedentemente allineate e ridimensionate .

1. Creazione dei file CSV

Dopo aver estratto il contenuto del file CelebA.txt, è stato necessario tradurlo in un formato csv al fine di individuare successivamente le 1000 identità più presenti nel dataset.

Per fare ciò è stato utilizzato un dizionario python che ha per chiave l’identità(label) e per valore il numero di volte che tale etichetta si ripete nel dataset.

Tale dizionario è stato poi ordinato in modo da individuare le 1000 più presenti che sono state poi inserite in un omonimo vettore.

Essendo l’obiettivo di questa fase la creazione di 3 file csv (training,validation,test),le immagini di ogni identità sono state opportunamente suddivise in: immagini di training(70%), validation(10%) e test(20%).

1. Preparazione dei dati

La rete opera su immagini in formato channel-last (i.e. 224,224,3) che devono essere opportunamente preprocessate prima di poter essere utilizzate per il training.

E’ stato necessario utilizzare la funzione ImageDataGenerator di Keras per specificare il tipo di data augmentation da effettuare, mentre la funzione flow\_from\_dataframe per applicare il generatore alle immagini prelevate dal relativo csv.

1. Architettura

E’ stata utilizzata come modello la rete ResNet50 (Figura 2) con un opportuno Fine-Tuning al fine di inizializzare i pesi della rete con quelli pre addestrati sul dataset ImageNet.



Figura 2

L’ottimizzatore scelto è stato Adam, metodo stochastic gradient descent basato su una stima adattiva dei momenti del primo e secondo ordine.

Il learning rate è stato impostato a un valore di 0.0003, in accordo all’articolo “Calibrating the Adaptive Learning Rate to Improve Convergence of ADAM” [3].

La loss function scelta è stata la categorical\_crossentropy che accetta in ingresso etichette in formato one-hot.

1. Addestramento

Per addestrare la rete è stato utilizzato il metodo fit di Keras, poiché è stato deprecato da Keras stesso l’utilizzo del metodo fit\_generator per addestrare reti fornendo in ingresso un generatore piuttosto che una directory di immagini.

Sono stati scelti i parametri di batch size e numero di epoche in seguito a sperimentazioni discusse nel paragrafo ‘4-Risultati sperimentali’.

Sono stati inoltre introdotti inoltre i parametri di workers e use\_multiprocessing al fine di sfruttare la multiprogrammazione mediante multi threading al fine di velocizzare le operazioni di training.

1. Valutazione delle prestazioni

Le prestazioni della rete sono state valutate mediante le metriche fornite dal modulo sklearn.metrics, in particolare sono state evidenziate la confusion\_matrix sia nella sua versione numerica che in una sua rappresentazione grafica e la metrica di accuracy\_score che evidenzia un’accuracy di test prossima a quella di validation precedentemente rilevata.

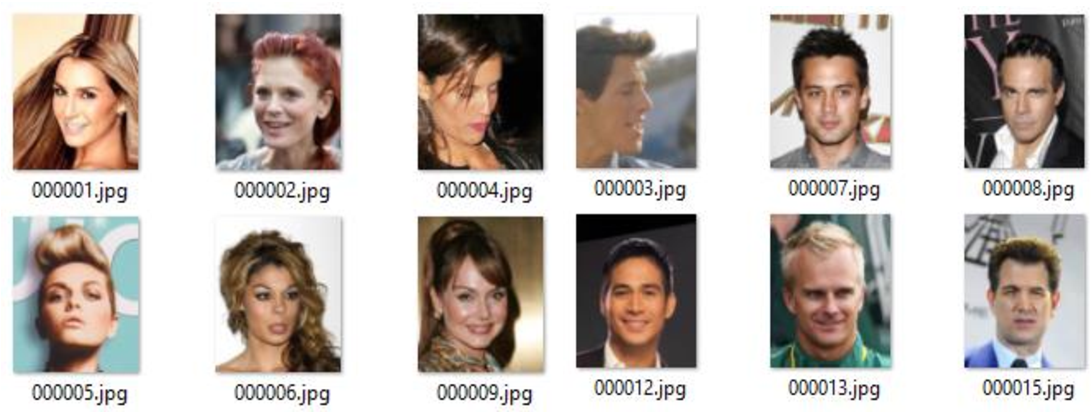


Figura 3

**3 - Dataset**

CelebFaces Attributes Dataset (CelebA) è un dataset di attributi facciali su larga scala che comprende 202.599 immagini relative a 10.177 celebrità, ciascuna delle quali è identificata da un’annotazione univoca.

Esempi del dataset sono riportati in Figura 3.

Ogni volto è descritto mediante 40 etichette binarie che rappresentano la presenza o l’assenza di altrettanti attributi facciali (e.g. borse sotto gli occhi, capelli scuri, naso grande).

Le immagini in questo set di dati coprono grandi variazioni di posa e confusione di sfondo.

In particolare nel progetto sono state utilizzate le immagini precedentemente allineate e ridimensionate in dimensione 218 x 178, ottenute dal lavoro “Deep Learning Face Attributes in the Wild” [1].

**4 - Risultati sperimentali**

L’approccio sviluppato si fonda sulla rete convoluzionale ResNet50 (Figura 2).

Tale rete è stata scelta in seguito a una serie di tentativi effettuati con altre reti, tra cui: VGG19, ResNet50v2 e ResNet152v2.

In particolare in seguito a una serie di esperimenti che prevedevano l’utlizzo delle reti VGG è stato chiaro che pur variando i valori di learning rate a partire da valori ‘ottimali’ per l’ottimizzatore Adam utilizzato [3], i risultati di accuracy erano di circa 2 ordini di grandezza inferiori al previsto.

Dopo aver variato in maniera incrementale anche i valori di batch size (da 32 a 64 fino a 128) senza ottenere effettivi miglioramenti nelle prestazioni di training e validation, si è passati ad effettuare un esperimento valutando non più le 1.000 identità più presenti nel dataset, ma solo le 2 più presenti. In questo modo, dopo aver ottenuto risultati accettabili di accuracy dopo sole 10 epoche (intorno al 40% per il training e 30% per la validazione) si è ritenuto necessario cambiare rete convoluzionale per affrontare un problema di image recognition con un elevato numero di classi e con un insufficiente numero di immagini per classe che permettesse di addestrare un elevato numero di parametri allenanti di una rete come VGG19.

Facendo riferimento all’articolo “Deep Residual Learning for Image Recognition” [2] è stata scelta la famiglia di reti ResNet; in particolare per una maggior semplicità implementativa ed efficienza è stata scelta la versione ResNet50, la quale dopo 15 epoche, utilizzando un valore di 0.003 per il learning rate dell’ottimizzatore Adam, e come loss function la cross-entropy loss, mostrava risultati di accuracy in test e validazione pari alle versioni più avanzate e con un maggior numero di parametri addestrabili della famiglia ResNet (e.g. ResNet50v2, ResNet152V2 Figura 4).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Rete** | **Parametri totali** | **Grandezza** | **Livelli** | **Training Accuracy** | **Validation**  **Accuracy** |
| ResNet50 | 25,636,712 | 98 MB | 50 | 0.9571 | 0.5024 |
| ResNet152V2 | 60,419,944 | 232 MB | 152 | 0.9489 | 0.4929 |
| ResNet50V2 | 25,613,800 | 98 MB | 50 | 0.8661 | 0.4311 |
| VGG19 | 143,667,240 | 549 MB | 26 | 0.001 | 0.001 |

Figura 4

E’ stata poi testato il medesimo modello aumentando il numero di epoche, tuttavia si è notato uno stallo nella diminuzione del valore di loss function nel training e una diminuzione dell’accuracy di validazione causata da un’eccessivo overfitting della rete sui dati di training.

Nonostante la data augmentation effettuata sul dataset ristretto alle 1.000 identità più presenti, le immagini risultanti non erano in numero sufficiente ad evitare un overfitting della rete e quindi il conseguente assestamento del valore di accuracy in validation intorno al 50%; si è ritenuto quindi soddisfacente arrestarsi alla quindicesima epoca.

**Riferimenti bibliografici**

[1] Z. Liu, P. Luo, X. Wang, and X. Tang, “Deep Learning Face Attributes in the Wild” International Conference on Computer Vision (ICCV) 2015.

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, “Deep Residual Learning for Image Recognition” .

[3] Qianqian Tong, Guannan Liang, Jinbo Bi “Calibrating the Adaptive Learning Rate to Improve Convergence of ADAM”