

UNIVERSITÀ DEGLI STUDI DI VERONA

CORSO DI LAUREA IN INFORMATICA

Riconoscimento di bottiglie di vino

Autore

Alessia BODINI

Supervisori

Marco CRISTANI

Pietro LOVATO

28 maggio 2019

1 Introduzione

Il progetto realizzato ([?]) si basa sull'applicazione di un software di riconoscimento di scritte a un dataset di bottiglie di vino, e in particolare sulle loro etichette, con lo scopo di affibiare a ognuna di esse il nome del vino ed eventuali informazioni aggiuntive (quali produttore e luogo di provenienza).

Per l'analisi ci si è basati su un dataset contenente immagini di vini suddivise in tre categorie:

- *gt (ground-truth)*: immagini ufficiali rese disponibili dai produttori sul sito di Vinitaly;



Figura 1: Bottiglia di Amarone prodotta da Corte Giara, caricata sul sito di Vinitaly.

- *raw*: foto degli stessi vini pubblicate sui social network;



Figura 2: Immagine caricata da un utente su Intstagram che presenta una bottiglia identica alla precedente.

- *bottles*: immagini presenti in *raw* appositamente ritagliate per lasciare le bottiglie in primo piano.



Figura 3: Bottiglia di Amarone di Corte Giara ritagliata dalla foto in *raw*.

A partire dal dataset, lo scopo del programma è quello di riconoscere per ogni immagine nel dataset le parole presenti e, in seguito, di associare a ciascuna foto in *raw* e *bottles* il nome della bottiglia raffigurata, a seconda delle corrispondenze trovate con le immagini in *gt*. All'interno di *raw* e *bottles* le immagini sono disposte in modo tale che ognuna di esse sia presente nella sottodirectory con il nome della bottiglia in *gt* a cui fa riferimento, per verificare poi facilmente l'accuratezza dei risultati raggiunti.

2 Fase di riconoscimento

Per riconoscere le etichette inizialmente si è ricorsi a una demo del programma di riconoscimento sviluppato da T. Wang: *End-to-End Text Recognition with Convolutional Neural Networks* ([?]). A causa però degli scarsi risultati raggiunti si è deciso di cambiare strategia.

Si è quindi passati all'attivazione di una chiave per l'utilizzo dei *Cognitive Service* di Microsoft, che per la sezione di *Computer Vision* mettono a disposizione una serie di API per il riconoscimento del testo, sia scritto a mano che stampato. Richiamando all'interno di un programma Python le due API *Recognize Text* ([?]) e *Get Recognize Text Operation Results* ([?]) si è stati in grado di ottenere dei file .json con le parole riconosciute dalle immagini presenti nel dataset. I file risultato vengono posti all'interno delle stesse directory contenenti le immagini di riferimento e contengono due oggetti principali:

- *status*, posto uguale a *Succeeded* se l'operazione ha avuto successo.

"status": "Succeeded"

- *recognitionResult*, che comprende a sua volta *lines* e per ognuna di queste *boundingBox*, *text* e *words*. Quest'ultime sono quelle che a cui si è interessati e su cui verranno effettuati i successivi controlli. Per l'immagine sottostante ad esempio una parte dei risultati prodotti si può vedere nella pagina successiva.



Figura 4: Bottiglia di Amarone prodotta da Corte Giara.

```

"recognitionResult": {
  "lines": [
    {
      "boundingBox": [
        21,
        348,
        114,
        347,
        115,
        360,
        21,
        361
      ],
      "text": "CORTE GIARA",
      "words": [
        {
          "boundingBox": [
            22,
            348,
            66,
            348,
            66,
            361,
            22,
            361
          ],
          "text": "CORTE"
        },
        {
          "boundingBox": [
            72,
            348,
            115,
            348,
            115,
            361,
            72,
            361
          ],
          "text": "GIARA"
        }
      ]
    }
  ],
  ...

```

Il procedimento riguardante la chiamata alle API di riconoscimento e il salvataggio dei risultati conseguiti nei file .json è visibile all'interno di *recognition.py*.

3 Fase di validazione

Una volta terminata la fase di riconoscimento, si è passati a confrontare i risultati raggiunti per le bottiglie nei dataset *bottles* e *raw* con quelli ottenuti a partire dalle immagini in *gt*. Di seguito si elencano i passi eseguiti per ottenere i riscontri sulle immagini in *bottles* (visibili in *validation.py*); similmente vengono raggiunti i risultati per *raw*, semplicemente modificando il nome del dataset da cui importare le immagini.

1. I nomi delle immagini da analizzare vengono importati dai dataset *gt* e *bottles*.
2. L'analisi prosegue esaminando una foto in *bottles* per volta. In particolare, per ognuna di esse:
 - a. vengono estratti i risultati ottenuti nella fase di riconoscimento dal corrispondente file *.json*;
 - b. per trovare la bottiglia di appartenenza, si scorrono tutte le bottiglie presenti in *gt*, leggendo per ognuna di esse le parole ottenute durante il riconoscimento e confrontando quest'ultime con quelle trovate per la bottiglia corrente in *bottles*. In particolare, per ogni parola comune trovata si aumenta il punteggio per quell'immagine *gt*;
 - c. si crea per l'immagine corrente un nuovo file *.json*, contenente, in ordine di somiglianza (cioè a secondo del punteggio ottenuto nella fase precedente), i nomi delle possibili bottiglie di appartenenza;
 - d. si verifica quale nome nel nuovo file *.json* corrisponde esattamente alla directory di appartenenza dell'immagine. In particolare, se il nome della bottiglia corretto è ritrovato in posizione x , nell'array *ranks* si aumenta di uno il numero di immagini riconosciute in posizione x .
3. Al termine dell'analisi, i valori presenti in *ranks* vengono rappresentati in un apposito grafico. Ques'ultimo presenta sull'asse delle ascisse le posizioni in cui è avvenuto il match e su quello delle ordinate il numero di immagini riconosciute in quella posizione durante il precedente controllo.

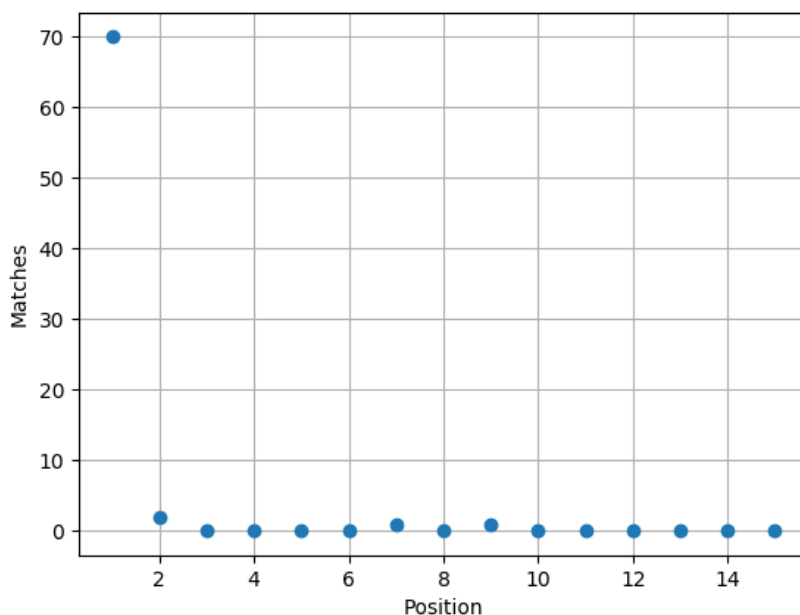


Figura 5: Grafico dei match

4. In seguito, si è proceduto ricavando da *ranks* il grafico relativo alla CMC (*Cumulative Match Curve*).

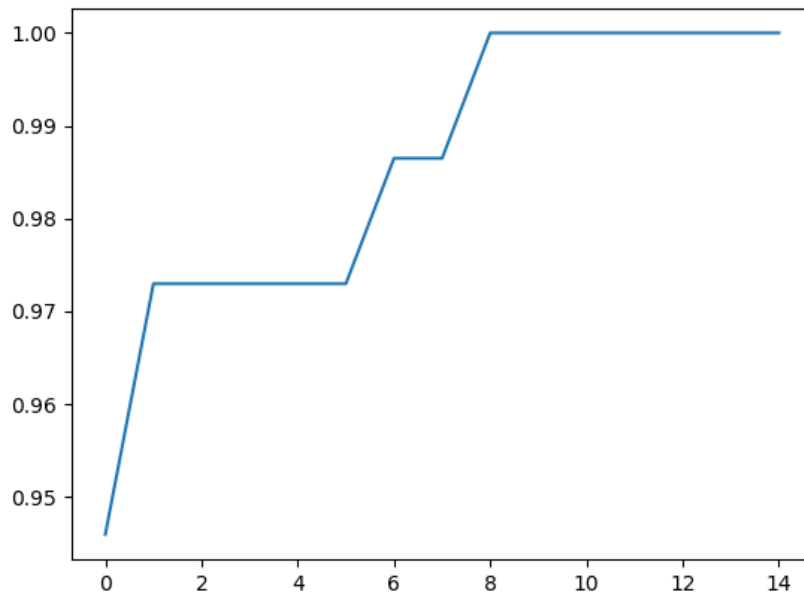


Figura 6: Grafico della CMC

5. Per concludere, viene stampato il valore dell'accuratezza del sistema, ottenuto calcolando l'area sotto la curva della CMC.

I valori di accuratezza raggiunti in base ai due dataset sono i seguenti:

	<i>Bottles</i>	<i>Raw</i>
<i>Accuracy</i>	0.9207	0.9144

Tabella 1: Tabella con valori di accuratezza

4 Conclusioni

I risultati finali sono piuttosto soddisfacenti. Visti livelli di accuratezza raggiunti, mostrando l'immagine che presenta al suo interno la bottiglia da riconoscere la probabilità di ottenere il nome e relative specifiche del vino è pari al 91% (addirittura maggiore se viene messa in primo piano). Al momento, al termine del riconoscimento, solo il nome ufficiale viene stampato per ogni bottiglia, ma future modifiche al programma potrebbero consistere nell'aggiungere una serie di documenti da cui poter estrarre tutte le informazioni riguardanti la bottiglia richiesta. Inoltre, il dataset di immagini *gt* finora utilizzato risulta molto piccolo, e per poter soddisfare tutte le possibili richieste dovrebbe essere ampliato il più possibile.

In conclusione, il progetto è modificabile su più punti e ha ampi margini di miglioramento per quello che riguarda la risposta all'utente.

Riferimenti bibliografici

- [1] Alessia Bodini. *Progetto di Wine Bottles Recognition*.
- [2] Tao Wang, David J. Wu, Adam Coates and Andrew Y. Ng.. *End-to-End Text Recognition with Convolutional Neural Networks*. In *Proceedings of the Twenty-First International Conference on Pattern Recognition (ICPR 2012)*.
- [3] *Computer Vision API - v2.0, Recognize Text*
- [4] *Computer Vision API - v2.0, Get Recognize Text Operation Result*