



Università degli Studi di Verona

DIPARTIMENTO DI INFORMATICA

Corso di Laurea in Informatica

TESI DI LAUREA TRIENNALE

Riconoscimento statistico di immagini raffiguranti bottiglie di vino

Laureanda:

Alessia Bodini

Matricola **VR411353**

Relatore:

Marco Cristani

Correlatore:

Pietro Lovato

Anno Accademico 2018–2019

Indice

1	Introduzione	2
2	Stato dell'arte	4
2.1	Wine-Searcher	5
2.2	Delectable	7
2.3	Vivino	8
3	Tecniche di riconoscimento già sviluppate	10
3.1	Riconoscimento di etichette di bottiglie di vino	10
3.2	Riconoscimento di etichette di farmaci	11
3.3	Classificazione di bottiglie in base alla loro sagoma	13
4	Sistema di riconoscimento scelto	14
4.1	Metodo utilizzato	14
4.2	Applicazione su immagini <i>in the wild</i>	16
4.2.1	Caparzo	16
4.2.2	Gulfì	20
4.2.3	Primitivo	24
4.2.4	Etna Rosso	28
4.3	Risultati raggiunti	29
5	Classificazione dei prodotti	30
6	Validazione del sistema	32
7	Conclusioni	35

Capitolo 1

Introduzione

Il progetto realizzato ([1]) 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 affibbiare 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 1.1: Bottiglia di Amarone prodotta da Corte Giara, caricata sul sito di Vinitaly.

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



Figura 1.2: Immagine caricata da un utente su Instagram che presenta una bottiglia identica quella in fig. 1.1.

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



Figura 1.3: Bottiglia di Amarone di Corte Giara ritagliata da fig. 1.2.

A partire dal dataset, lo scopo del programma è quello di riconoscere all'interno di ogni immagine 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, in modo da poter verificare facilmente l'accuratezza dei risultati raggiunti.

Un tale programma potrebbe essere adattato anche all'ambito mobile, così da consentire a possibili acquirenti di avere informazioni, possibilmente nella loro lingua madre, riguardanti il vino che sono interessati a comprare.

Capitolo 2

Stato dell'arte

Programmi come quello appena descritto sono già stati realizzati da diverse compagnie con lo scopo di diffondere notizie e valutazioni di vini in tutto il mondo. Applicazioni di ricerca di vini come Vivino e Wine-Searcher sono ormai famose per possedere ampi database e una grande community su cui fare affidamento, e negli ultimi anni hanno ampliato le loro tecniche di ricerca abbracciando il metodo del wine label scanning, ovvero del riconoscimento di vini tramite scannerizzazione da smartphone delle loro etichette.

Molte delle app disponibili al giorno d'oggi presentano in realtà delle tecniche di riconoscimento poco sofisticate, che presentano tassi di matching relativamente bassi. In base all'analisi del 2016 ([2]) effettuata da Richard Hemming, esperto scrittore e consulente di vini, vengono esposti di seguito le tre app maggiormente conosciute e con le più elevate valutazioni dal punto di vista del software di riconoscimento utilizzato. Per ognuna è stato eseguito un test del servizio di riconoscimento utilizzando la bottiglia Amarone di Corte Giara presentata al capitolo precedente ed in particolare dell'etichetta ritagliata da fig. 1.2:



Figura 2.1: Etichetta di una bottiglia di Amarone di Corte Giara ritagliata da fig. 1.2.

2.1 Wine-Searcher

Il sito neozelandese wine-searcher.com (disponibile anche come app per iOS e Android) presenta un database e un motore di ricerca che insieme permettono di unire vini e relativi prezzi di vendita in tutto il mondo. Ideato nel 1999, oggi viene utilizzato da milioni di utenti per localizzare, confrontare e comprare diversi tipi di vini. Il database viene costantemente aggiornato e dei processi manuali e automatici permettono il mantenimento delle liste di vini presenti.

Tra i servizi che Wine-Searcher mette a disposizione si trova quello di ricerca tramite foto del vino di interesse. Il riconoscimento si ottiene attraverso la sua etichetta, che deve risultare già centrata e messa a fuoco nell'immagine per ottenere risultati concreti. Nonostante la praticità del sito e l'estesa disponibilità di prezzi e informazioni per ogni vino, il servizio di riconoscimento non si posiziona tra i migliori. Secondo i dati rilevati da Richard Hemming ([2]) nei suoi test (su 5 differenti bottiglie di vino) la velocità risulta piuttosto elevata ma le identificazioni non sono mai al 100% corrette e, in alcuni casi, si sono dimostrate del tutto non accurate.

Sottoponendo al programma l'etichetta di Amarone il programma ritorna la corretta bottiglia di riferimento:



Figura 2.2: Riscontro da Wine-Searcher per fig. 2.1.

Ma se al posto della precedente etichetta viene passata una foto della bottiglia a distanza si ottiene come riscontro un marchio completamente diverso:



Figura 2.3: Riscontro da Wine-Searcher per fig. 1.2.

Riassumendo quindi l'applicazione permette sì una buona identificazione nel momento in cui l'etichetta risulta in primo piano, ma nel caso in cui le vengono sottoposte foto della bottiglia *in the wild*, come può capitare se si vuole pubblicare sui social network il prodotto, Wine-Searcher non riesce nemmeno ad avvicinarsi al risultato aspettato.

2.2 Delectable

Delectable è un applicazione web che nella sua versione mobile per iOS e Android permette agli utenti di scannerizzare etichette di vini tramite smartphone ed avere così accesso a recensioni, foto e valutazioni da parte di altri utenti. Fornisce inoltre, in collaborazione con il servizio Banquet, un tool per consentire ai medesimi utenti di localizzare e comprare il vino desiderato.

Il servizio di riconoscimento messo a disposizione da Delectable risulta nella maggior parte dei casi più lento di quello in uso dal suo collega Wine-Searcher, ma garantisce risultati migliori, come si può vedere di seguito:



Figura 2.4: Riscontro da Delectable per fig. 2.1.

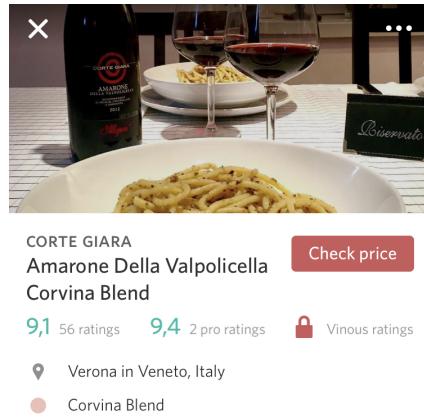


Figura 2.5: Riscontro da Delectable per fig. 1.2.

Le due risposte del sistema dimostrano in entrambi i casi che il riconoscimento è stato più che buono, ad eccezione di alcune sottilizzze. Caratteristica però non presente in questo software rispetto al precedente è quella di non estrarre dall'etichetta l'anno della bottiglia; la ricerca viene infatti lasciata all'utente nel momento in cui decide di scorrere tra i prezzi di vendita dei fornitori consigliati.

2.3 Vivino

Vivino è il mercato di vini online più grande del mondo e l'app di vini più scaricata, basata su una community di milioni di persone. Fondata a Copenhagen nel 2010 da Heini Zachariassen e Theis Søndergaard, oggi Vivino ha un database che conta oltre 10 milioni di vini differenti e 35 milioni di utenti. Come app Vivino supporta la ricerca tramite scannerizzazione dell'etichetta del vino interessato. La tecnica di riconoscimento utilizzata si basa su una combinazione di pattern recognition e OCR (optical character recognition). Il primo decide quale delle foto esistenti rappresenta il miglior match; parallelamente, il secondo esegue un'analisi con lo scopo di migliorare quelli che sono i dettagli meno distinguibili, come spesso risultano l'anno e le informazioni sull'uva di provenienza. L'intero processo impiega meno di un secondo sui server.

In questo caso si è deciso di eseguire un test più approfondito per quanto riguarda l'accuratezza raggiunta dal sistema di riconoscimento. Visto infatti che Vivino sembra essere l'applicazione al momento più popolare in commercio per quanto riguarda la ricerca di vini, gli si è sottoposto l'intero dataset utilizzato per il progetto corrente. In particolare al software di riconoscimento di Vivino sono state passate tutte le foto presenti nella directory *raw*, appositamente ritagliate in modo da mostrare esclusivamente la parte riguardante l'etichetta (come richiesto dall'app). Di seguito vengono riportati un paio di esempi dei riscontri ottenuti:



Figura 2.6: Riscontro da Vivino per fig. 2.1.



Figura 2.7: Riscontro da Vivino per la foto di una bottiglia di Donna Walda.

Come si può notare, nel secondo caso l'applicazione non produce alcun risultato, bensì si limita a presentare due opzioni: quella di ricerca manuale o quella di richiesta di identificazione da parti di esperti. Si deduce dall'insieme dei risultati ottenuti che sì, il software di riconoscimento di Vivino risulta in grado di identificare una grande quantità di vini, ma talvolta gli effetti di luci/ombre e qualità della foto possono impedire una lettura efficace dell'etichetta.

Tramite l'analisi dell'intero dataset (165 immagini in totale) si è ottenuto un livello di accuratezza pari allo 69.1%, con ben 114 vini riconosciuti.

Dall'analisi delle tre principali applicazioni in commercio per la ricerca e valutazione di vini, se ne ricava che i risultati raggiunti attraverso le tecniche di machine learning sono appena soddisfacenti, se si considera che il maggior risultato raggiunto (circa il 70% di riconoscimenti corretti) è quello riportato da Vivino. Sarebbe quindi utile scoprire se l'utilizzo di altre tecniche di riconoscimento permetta di raggiungere dei livelli di accuratezza maggiori.

Capitolo 3

Tecniche di riconoscimento già sviluppate

Analizzando diversi articoli pubblicati online si sono trovate tre principali tecniche di riconoscimento, applicate nell'ambito di bottiglie di vino, contenitori di farmaci e soft drinks, che possono risultare utili per questo progetto.

3.1 Riconoscimento di etichette di bottiglie di vino

Il sistema ([3]) progettato da alcuni studenti della Chonnam National University della Corea del Sud include il ritrovamento e l'estrazione del testo da un'immagine per il riconoscimento di un'etichetta appartenente a una bottiglia di vino. Il procedimento tiene conto sia dei diversi font presenti sia dell'illuminazione (a seconda dei differenti scenari) e di altri effetti causati dalla possibile presenza di rumore. Il programma si suddivide in due parti fondamentali: *text detection* e *text recognition*.

La prima fase consiste in:

- ricerca della regione di testo: in seguito a un (temporaneo) processo di binarizzazione sull'intera immagine, viene applicato il *canny edge detector* con l'obiettivo di identificare i contorni degli oggetti presenti. Tramite l'analisi dell'istogramma dei contorni si estraggono poi le regioni che presentano un certo picco: esse rappresenteranno le zone di testo nell'immagine;

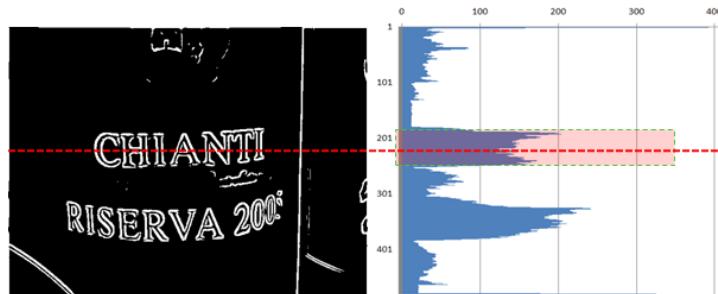


Figura 3.1: Metodo di localizzazione della regione di interesse tramite *edge detection*.

- rilevazione dei singoli caratteri: si prendono in considerazione i *gap* localizzati tra i contorni verticali all'interno delle regioni trovate. L'utilizzo di questa tecnica prima della binarizzazione (definitiva) della regione si è dimostrato dare risultati migliori rispetto al mantenimento del risultato della fase precedente. Segue poi l'applicazione del *FMC (fuzzy c-means) clustering* su ogni carattere trovato per la fase di segmentazione;

- unione dei caratteri trovati nella sequenza corretta.



Figura 3.2: Esempi di segmentazione dei caratteri.

La fase di riconoscimento del testo invece comporta l'utilizzo di un approccio statistico, che consiste nel ricercare le differenze di ogni carattere analizzato con gli altri presenti, facendo uso dell'informazione sulla forma del carattere “a maglia” (*mesh-based shape information*). Questo metodo risulta piuttosto semplice da implementare e risulta la scelta ottimale vista il diverso tipo di illuminazione presente nei vari punti delle immagini e la possibile presenza di font corsivi e/o con decorazioni particolari.

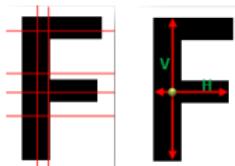


Figura 3.3: Esempio di estrazione delle caratteristiche per la fase di riconoscimento.

L'accuracy di quest'ultima fase ha raggiunto il 97.5%.

3.2 Riconoscimento di etichette di farmaci

Il secondo esempio di sistema di riconoscimento preso in considerazione è stato progettato per contenitori/bottigliette di farmaci. Inizialmente la confezione di una serie di pillole può sembrare avere poco in comune con una bottiglia di vino, eppure dal punto di vista del riconoscimento si possono trovare diversi aspetti che li accomunano; tra questi la presenza di un'etichetta che avvolge il farmaco e la sua forma (spesso) cilindrica che provoca le stesse distorsioni alle parole come per una classica bottiglia.

Il problema di riconoscimento viene in primo luogo trattato con un'operazione di *image processing* che migliora la qualità dell'immagine, a cui seguono localizzazione, segmentazione e riconoscimento dei singoli caratteri pervenuti.

La parte di *image processing* comprende:

- conversione dell'immagine in scala di grigi;
- correzione della distorsione dei caratteri, tramite un processo di amplificazione non lineare sui punti più laterali (e quindi distorti) del farmaco;
- smoothing con un filtro mediana 3x3 per eliminare la possibile presenza di rumore data dalla scena circostante;

- rilevazione dei contorni tramite algoritmo LOG (*Laplacian Of Gaussian*).



Figura 3.4: Effetti delle operazioni di pre-processing.

Una volta migliorata la qualità dell’immagine, si passa alla localizzazione delle regioni di testo. Il processo avviene tramite suddivisione dell’immagine in regioni con conseguente estrazione delle regioni interessate mediante *contour tracking*. Per definire in modo ancor più accurato i simboli trovati nel testo si analizzano poi i pixel neri presenti tra un carattere e l’altro.

Ad esso segue la fase di segmentazione, prima tramite utilizzo della proiezione verticale e in seguito, per migliorare il risultato, attraverso l’applicazione dell’algoritmo *drop-fall*.

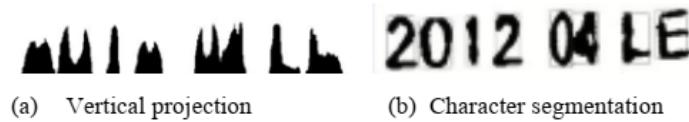


Figura 3.5: Effetti della segmentazione dei caratteri.

Per finire, il riconoscimento del testo si ottiene in tre passaggi:

- applicazione di un’interpolazione bilineare per normalizzare la grandezza dei caratteri;
- estrazione delle caratteristiche strutturali e statistiche più semplici in modo da classificare i caratteri in tre gruppi fondamentali (secondo le caratteristiche dei numeri di Eulero);
- applicazione della trasformata K-L (*Karhunen-Loeve transform*) sulle feature ottimali estratte e introduzione del risultato in una rete neurale con *back propagation* e tasso di apprendimento adattivo.

Dagli esperimenti svolti si sono ottenuti i seguenti tassi di riconoscimento (in base alla primitiva classificazione sui numeri di Eulero):

<i>Numeri di Eulero</i>	-1	0	1
<i>Tasso</i>	97.3%	96.9%	97.1%

Tabella 3.1: Risultati del testing su BP neural network.

3.3 Classificazione di bottiglie in base alla loro sagoma

Un altro metodo ([4]) già utilizzato nel riconoscimento delle bottiglie, progettato da alcuni studenti dell'*Universidad Icesi* della Colombia, si basa sull'analisi della silhouette di alcune soft drinks. Il sistema prende in input una foto in alta risoluzione, scattata in un ambiente con luminosità controllata, e da essa estrae le principali caratteristiche fisiche della bottiglia: il centroide X e Y, l'altezza e larghezza sia della bottiglia in sé che del tappo che la chiude e l'area dell'intera sagoma.

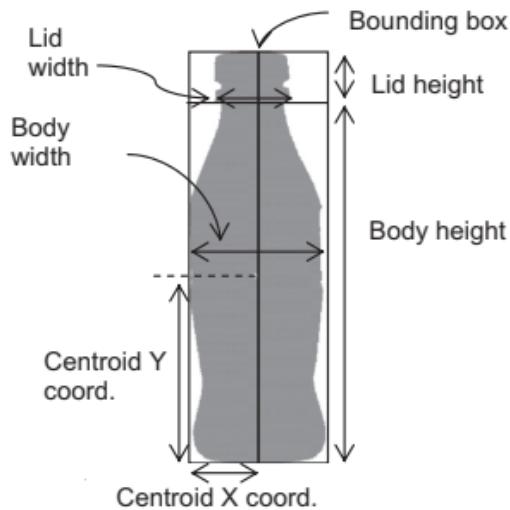


Figura 3.6: Proprietà fisiche misurate.

Le feature appena estratte, insieme ad altre calcolate da quelle principali, vengono poi utilizzati per la classificazione in categoria di prodotto: acqua, succhi di frutta, bevande sportive, bibite gassate o altro.



Figura 3.7: Bottiglie e relativa classificazione.

Dai risultati ottenuti si è potuto constatare che i quattro parametri indicanti altezza e larghezza delle due parti di prodotto sono sufficienti per la classificazione in una delle categorie, con un tasso di successo del 82.7%.

Capitolo 4

Sistema di riconoscimento scelto

Per il compito scelto, ovvero quello classificare bottiglie di vino, si è escluso inizialmente il metodo basato sulla forma del prodotto, sapendo che la silhouette di alcuni vini risulta praticamente identica quando il produttore è lo stesso. Basti pensare al caso di Corte Giara, che produce sia bottiglie di Amarone che di Ripasso, i cui prodotti si distinguono solamente dal nome presente all'interno dell'etichetta. Entrambi gli altri due metodi, visti i tassi di riconoscimento piuttosto elevati (superiori al 97%), risultano adatti al caso preso in considerazione, ma non essendo i relativi codici disponibili si è deciso di ricorrere all'utilizzo di altri sistemi più generici.

4.1 Metodo utilizzato

Si è inizialmente ricorsi a una demo del programma di riconoscimento sviluppato da T. Wang: "End-to-End Text Recognition with Convolutional Neural Networks" ([5]). 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 tre oggetti: *lines*, *text* e *words*, uno all'interno dell'altro. *text* e *words* contengono a loro volta un *boundingBox*, cioè il riquadro che include rispettivamente la riga di testo o la parola corrispondente, espresso tramite otto coordinate *x* e *y* (i quattro punti del riquadro). I risultati in *words* sono quelli a cui si è interessati e su cui verranno effettuati i successivi controlli. Per darne un esempio, si riportano di seguito i risultati (parziali) ottenuti dalla bottiglia di Corte Giara mostrata in precedenza:

```

    "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*.

4.2 Applicazione su immagini *in the wild*

Di seguito verranno elencate una serie di immagini con corrispettivi risultati ottenuti nel riconoscimento tramite le due API Microsoft *Recognize Text* e *Get Recognize Text Operation Results*. Per prima verrà presentata una bottiglia di Caparzo, riconosciuta in tutte le immagini in *raw* e *bottles*; seguiranno poi bottiglie di Gulfi, Primitivo e Etna Rosso che hanno invece riscontrato alcuni problemi nel riconoscimento.

4.2.1 Caparzo

La bottiglia di Caparzo sul sito di Vinitaly è mostrata come segue:



Figura 4.1: Bottiglia di Caparzo

Si distinguono chiaramente nell’immagine le seguenti parole: “CAPARZO”, “BRUNELLO”, “DI”, “MONTALCINO”. Tutte e quattro vengono identificate durante il riconoscimento. La frase centrale in corsivo e l’ultima frase, presente verso la fine dell’etichetta, non vengono però riconosciute: la prima a causa della del font molto particolare utilizzato, e la seconda per la scarsa qualità. Inoltre viene identificata erroneamente la scritta “CAPARZU” sul collo della bottiglia.

Di seguito si elencano le foto di bottiglie di Caparzo con relativi risultati dal dataset *raw*.

- 2013
- CAPARZO
- BRUNELLO
- DI
- MONTALCINO
- DENOMINAZIONE
- DI
- ORIGINE
- CONTROLLATA
- E
- GARANTITA
- *altro dall'ambiente*



Tabella 4.1: La buona qualità dell'immagine ha permesso di riconoscere anche le scritte più piccole.

- 2013
- CAPARZO
- BRUNELLO
- DI
- MONTALCINO
- DENOMINAZIONE
- DI
- ORIGINE
- CONTROLLATA
- E
- CARAMIA



Tabella 4.2: La parola "GARANTITA" non è stata individuata a causa della disposizione della bottiglia, ruotata durante l'acquisizione della foto.

- CAPARZO
- 2011
- CAPARZO
- BRUNELLO
- DI
- MONTALCINO



Tabella 4.3: La frase finale sull'etichetta in questo caso non è stata riconosciuta a causa della bassa qualità della foto, ma in compenso è stata identificata correttamente la parola "CAPARZO" anche sul collo della bottiglia.

- 2011
- CAPARZO
- BRUNELLO
- DI
- MONTALCINO
- DENOMINAZIONE
- DI
- ORIGINE
- CONTROLLATA
- E
- GARANTITA



Tabella 4.4: L'ottima qualità dell'immagine e la posizione in primo piano della bottiglia hanno qui permesso di riconoscere in modo completo l'etichetta.

- LAZO
- CAPARE
- 2014
- CAPARZO
- BRUNELLO
- DI
- MONTALE
- *altro dall'ambiente*



Tabella 4.5: In questo caso la bottiglia di nostro interesse (la seconda da sinistra) è stata presa molto di lato, ottenendo così delle parole spezzate o, come accade sul collo della bottiglia, un mixto di quelle effettivamente presenti.

Si può concludere l'analisi di questo insieme di bottiglie dicendo che tutte le parole principali sono state sempre riconosciute, permettendo di associare correttamente tutte e cinque le immagini alla bottiglia di Caparzo presentata inizialmente. In questo caso il compito è risultato piuttosto facile, sia per la generale buona qualità delle foto riportate che per il formato delle frasi sull'etichetta (quasi completamente in stampatello maiuscolo).

4.2.2 Gulfi

La bottiglia di Gulfi sul sito di Vinitaly è mostrata come segue:



Figura 4.2: Bottiglia di Gulfi

Dalla figura il riconoscitore è stato in grado di estrarre due sole parole: “GULFI” e “Maccarj”. La seconda di queste non risulta in verità corretta poiché, come si vedrà in seguito in immagini di migliore qualità, la parola esatta sarebbe “NeroMaccarj”. La parola su cui contare per l’associazione con altre bottiglie rimane quindi solo “GULFI”.

Di seguito si elencano le foto di bottiglie di Gulfi con relativi risultati dal dataset *raw*.

*solo alcune parole ricavate
dall'ambiente circostante*



Tabella 4.6: La ripresa dall'alto della bottiglia non ha permesso di riconoscere alcuna parola utile da essa; sono stati ricavati altri vocaboli all'interno dell'immagine ma nessuno di questi che possa ricondurre all'origine della bottiglia.

- GULFI
- organic
- wines
- NerMaccarj
- 2009



Tabella 4.7: Nonostante la parziale presenza di un ostacolo (il bicchiere) davanti alla bottiglia, buona parte dell'etichetta è stata correttamente riconosciuta, e in particolare anche la parola di nostro interesse “GULFI”, permettendo l'associazione alla bottiglia di provenienza. Anche in questo caso la parola “NeroMaccarj” non è stata riconosciuta in modo esatto, a causa della rotazione della bottiglia.

- GULFI
- NerMaccarj
- 2011



Tabella 4.8: Il riconoscimento in questo caso è stato perfetto, avendo a disposizione la bottiglia in primo piano e con l'etichetta "dritta" durante l'acquisizione della foto.

- GULFI
- NerMaccarj
- 2010



Tabella 4.9: L'immagine piuttosto scura non ha permesso un riconoscimento perfetto come il precedente, ma comunque buono abbastanza da poter associare alla bottiglia la corretta provenienza.

- G
- ULFI
- organic
- wines
- NextMaccar)
- 2008



Tabella 4.10: In questo caso la percezione della bottiglia nell’immagine da parte del riconoscitore non ha dato come risultato la parola “GULFI” (necessaria per effettuare il match con la bottiglia di appartenenza).

Si è notato però un cambiamento sostanziale utilizzando al posto dell’ultima immagine la sua corrispettiva nel dataset *bottles*, come si può vedere nell’analisi che segue.

- GULFI
- organic
- wines
- NewMaccar)
- 2608



Tabella 4.11: Dai risultati raggiunti si nota che, in questo caso, l’immagine è stata correttamente associata alla bottiglia Gulfi.

L'analisi di questo secondo insieme di bottiglie fa notare come in certi casi l'operazione di ritaglio e la presenza dell'etichetta in primo piano possano portare ad avere un riconoscimento anche solo un minimo migliore ma comunque essenziale per permettere il match con la bottiglia desiderata. Inoltre, la presenza di luce/ombra e la disposizione della bottiglia risultano altri aspetti importanti nella fase di riconoscimento.

4.2.3 Primitivo

La bottiglia di Primitivo sul sito di Vinitaly viene così presentata:



Figura 4.3: Bottiglia di Primitivo

Il programma di riconoscimento, applicato sulla bottiglia di Primitivo, ha riportato come risultato tutte le parole presenti sull'etichetta in modo corretto: "PRIMITIVO", "Puglia", "CANTINE" e "LOSITO". In questo caso quindi l'immagine ufficiale rappresenta un' ottimo punto di partenza per il riconoscimento delle bottiglie nelle foto scattate.

Di seguito si elencano le foto di bottiglie di Primitivo con relativi risultati dal dataset *raw*.

- PRIMITIVO
- Puglia
- CANTINE
- LOSITO
- *altro dall'ambiente*

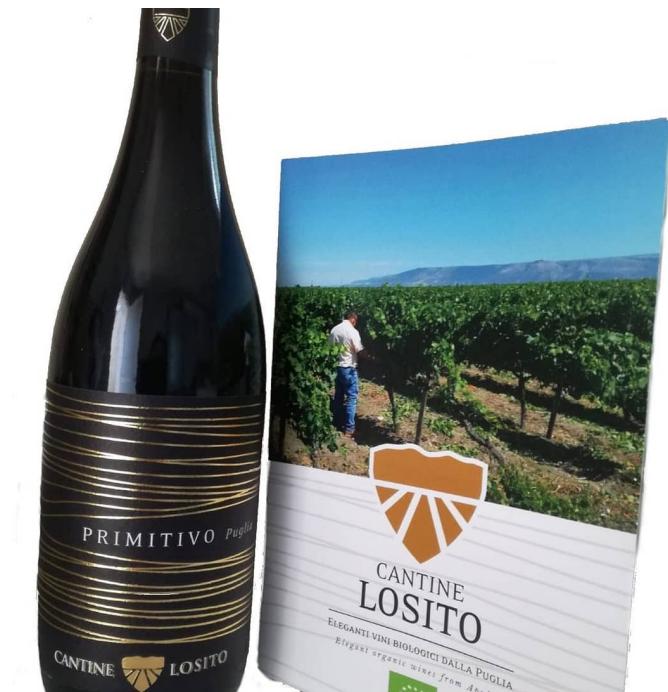


Tabella 4.12: La buona qualità dell’immagine e la ripresa piuttosto ravvicinata hanno permesso un riconoscimento perfetto dell’etichetta e di conseguenza un corretto match con la bottiglia di provenienza. Inoltre, come si vedrà anche più avanti, l’ambiente in cui è inserito la bottiglia talvolta risulta utile per il riconoscimento, presentando scritte e vocaboli che ne aiutano l’identificazione. In questo caso particolare, il dépliant presente accanto al vino presenta le parole “CANTINA” e “LOSITO” che riconducono sempre alla stessa bottiglia in *gt*.

- PRIMITIVO
- Puglia
- CANTINE
- LOSITO
- *altro dall’ambiente*



Tabella 4.13: Nonostante la posizione capovolta della bottiglia le parole presenti sull’etichetta sono state quasi del tutto perfettamente riconosciute, il che ha portato a un match corretto. L’unica parola rinvenuta scorretta è stata “CANTINE” a causa della eccessiva rotazione dell’etichetta verso il basso. Come introdotto nell’immagine precedente, vengono però in aiuto le scritte “CANTINE LOSITO” presenti sullo sfondo.



- PRIMITIVO
- Puglia
- CANTINE
- LOSITO
- *altro dall'ambiente*

Tabella 4.14: Il riconoscimento in questo caso risulta perfetto.



Tabella 4.15: L'insieme di una buona varietà di bottiglie in questo caso potrebbe portare ad associazioni sbagliate, ma la mancanza delle immagini *gt* degli altri vini ha comunque portato un match corretto.

Diversamente accade però nell'immagine in cui viene ritagliata solo la bottiglia di interesse, che, rispetto alla aspettativa, risulta più difficile da riconoscere, non essendo visibile parte dell'etichetta.

- RIMITIVO
- TINE
- LOSITO
- *altro dall'ambiente*



Tabella 4.16: In questo caso la foto ritagliata non consente di leggere correttamente l'etichetta, permettendo di individuare in modo esatto solo la parola “**LOSITO**”, associata anche alle bottiglie di Falesia.

- PRIMITIVO
- Puglia
- CANTINE
- LOSITO
- *altro dall'ambiente*



Tabella 4.17: L'immagine è correttamente associata alla bottiglia di Primitivo. L'ottima qualità della foto ha permesso anche il riconoscimento della scritta “**CANTINE LOSITO**” sul collo della bottiglia.

Diversamente dal precedente set, l'insieme di bottiglie appena analizzato fa notare come in certi casi l'operazione di ritaglio non porti necessariamente a un riconoscimento migliore (se non effettuata in modo minuzioso). Si è notato inoltre come in alcuni di questi lasciare visibile l'ambiente in cui il vino viene posto porti al riconoscimento di ulteriori parole utili alla fase di matching.

4.2.4 Etna Rosso

Per concludere viene presentato un altro vino, l'Etna Rosso, che ha presentato alcuni problemi nel riconoscimento di una specifica immagine.

La bottiglia di Etna Rosso sul sito di Vinitaly viene così presentata:



Figura 4.4: Bottiglia di Etna Rosso

Dal riconoscimento si ottengono le seguenti parole: “Etna”, “Rosso”, “DOC” e “PAPALE”.

Di seguito si mostra una foto, dal dataset *raw*, che presenta una bottiglia di Etna Rosso.

*nessuna parola è stata
riconosciuta*



Tabella 4.18: La bottiglia di vino in questo caso è troppo piccola nella foto e non è stato possibile identificare nulla dall’etichetta.

L’immagine ritagliata in *bottles* risulta pressoché identica e non porta alcun miglioramento.

4.3 Risultati raggiunti

Dall'analisi eseguita si possono definire alcuni punti essenziali per ottenere un buon riconoscimento. Le etichette in particolare vengono facilmente interpretate se:

- presentano parole con un font di semplice lettura in stampato minuscolo o (ancor meglio) maiuscolo;
- sono il meno possibile soggette a rotazioni.

Per quanto riguarda invece l'immagine in generale sono preferibili:

- ambienti chiari o con molta luce;
- buona qualità (per poter individuare anche le scritte più piccole);
- ambienti che presentano ulteriori scritte riguardanti la bottiglia (quali dépliant, fogli illustrativi o casse di vini)
- mancata presenza di bottiglie diversa provenienza.

Inoltre, l'orientamento della bottiglia non risulta essere un problema: come si è potuto vedere sono state efficacemente riconosciute parole disposte in qualsiasi direzione. Rimane solo da definire se il ritaglio sia preferibile o meno: in certi casi ritorna utile ma in altri la presenza di differenti tipi di bottiglie potrebbe portare ad associazioni errate.

Capitolo 5

Classificazione dei prodotti

Una volta terminata la fase di riconoscimento, si sono esaminate le parole ottenute dalle API di Microsoft per determinare il nome della bottiglia riportata nella foto. Per farlo si sono recuperate le parole riconosciute per ogni immagine del dataset *gt* e le si sono comparate a quelle ritrovate nella bottiglia in esame secondo la distanza di Damerau-Levenshtein (DL). Per questa parte è stata utilizzata l'implementazione *pyxDamerauLevenshtein* realizzata da Geoffrey Fairchild ([6]). Nel dettaglio, viene calcolata la distanza tra ogni parola ritrovata nella bottiglia selezionata e tutte le parole di ogni immagine *ground-truth*. La somma finale, ottenuta dall'insieme di tutte le distanze calcolate per ciascun termine riconosciuto nella bottiglia, viene identificata come un punteggio che identifica la somiglianza con quel vino. Una volta ottenuti i punteggi per ogni immagine nel dataset *gt*, quella che ha ottenuto il punteggio minore rappresenta la bottiglia di appartenenza della foto scelta. Per maggiori dettagli riguardo il calcolo dei punteggi si rimanda al capitolo 6.

Ad esempio, data la seguente foto riportante una bottiglia di Falesia:



Figura 5.1: Foto con bottiglia di Falesia

si ottiene, tramite l'esecuzione di *main.py*, la risposta del sistema.

```
Enter the name of an image in 'images_winebottles\test': Falesia.jpg
```

```
Sorting bottles for similarities with the chosen one:
```

```
['Falesia' 'Primitivo' 'LaSala' 'Jorche' 'Valiano' 'TelosWhite' 'Rinaldi'  
'Pinkfluid' 'Colvendra' 'CorteGiaraRipasso' 'CorteGiaraAmarone'  
'CorteMajoli' 'CorteMajoliAmarone' 'TelosRed' 'Antinori' 'Caparzo'  
'PassoDelleMule' 'Fongaro' 'Ferrari' 'Lvnae' 'PraSoave' 'EtnaRosso'  
'LeFontiASanGiorgio' 'Lungarotti' 'MarcheseAntinori' 'Damusa' 'Revi'  
'Gulfi' 'CadelBosco']
```

```
The name of the bottle is: Falesia
```

Figura 5.2: Esecuzione del programma (main.py) con input la foto precedente (5.1)

Capitolo 6

Validazione del sistema

Per studiare l'accuratezza raggiunta dal sistema si è andati a calcolare il numero di immagini presenti nei dataset *bottles* e *raw* correttamente riconosciute, tramite confronto con le rispettive immagini in *gt*. L'insieme *gt* preso in considerazione consiste in un dataset di 41 bottiglie di riferimento di vino. Di seguito si elencano i passi eseguiti per ottenere i riscontri sulle immagini in *raw* (visibili in *validation.py*); similmente vengono raggiunti i risultati per *bottles*, semplicemente modificando il nome del dataset da cui importare le immagini.

1. I nomi delle immagini da analizzare vengono importati dai dataset *gt* e *raw*.
2. Su entrambi i dataset viene applicato il software di riconoscimento che ritorna gli appositi file .json con le parole estratte.
3. L'analisi prosegue esaminando i risultati estratti da ogni foto presente in *raw*, una alla volta. Per ogni file .json:
 - a. l'insieme di parole viene estratto e inserito in un array *wordsBottle*. Quest'ultimo viene passato a una funzione esterna (*getScores*) che si occupa di calcolare la distanza di ogni immagine in *gt* (o meglio delle sue parole) con quelle ritrovate nella foto in esame. Il punteggio viene calcolato tramite uso della funzione *normalized_damerau_levenshtein_distance* che implementa l'algoritmo *Damerau-Levenshtein (DL) edit distance* ([6]);
 - b. calcolato lo score con ogni immagine *gt* si crea un nuovo array (*scores*) contenente l'ordine di queste bottiglie per somiglianza con quella in esame, con la semplice applicazione di un algoritmo di ordinamento (crescente) sugli indici;
 - c. se il nome della bottiglia corretto è ritrovato in posizione *x* dell'array *scores* si aumenta di uno il numero di immagini riconosciute in posizione *x* nel nuovo array *ranks*.
4. Al termine dell'analisi, i valori presenti in *ranks* vengono rappresentati in un apposito grafico. Quest'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.

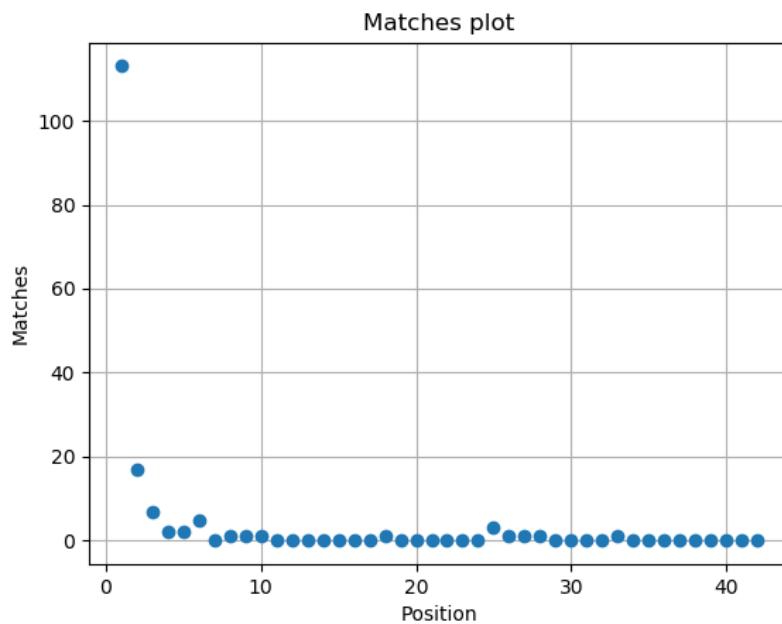


Figura 6.1: Grafico dei match per il dataset *raw*.

5. In seguito, si ricava il grafico relativo alla CMC (*Cumulative Match Curve*), ottenuto dalla somma cumulativa dei valori in *ranks*, precedentemente divisi per il numero di bottiglie in *gt*.

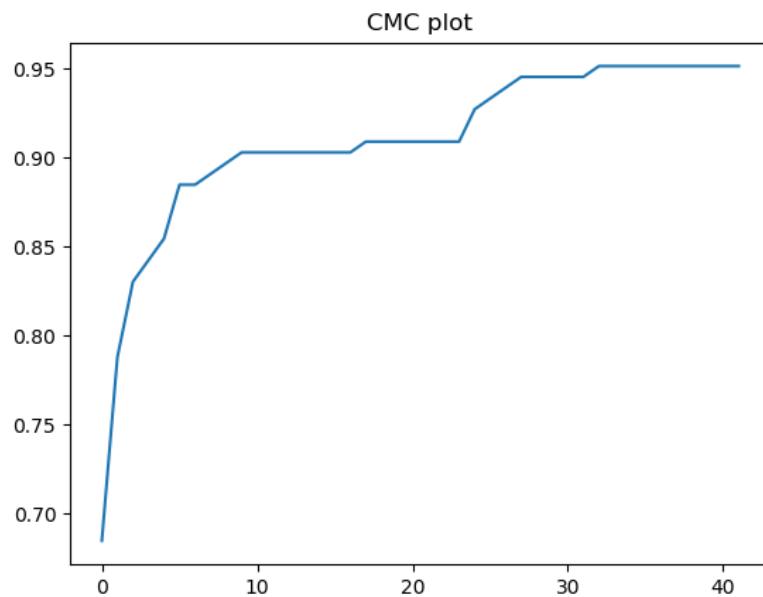


Figura 6.2: Grafico della CMC per il dataset *raw*.

6. Per concludere, viene stampato il valore che rappresenta l'accuratezza del sistema, ottenuto calcolando l'area sotto la CMC (AUC, *Area Under the Curve*).

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

	<i>Bottles</i>	<i>Raw</i>
Campioni	90	165
AUC	0.9604	0.9204

Tabella 6.1: Tabella con valori di accuratezza

Capitolo 7

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 all’89% (addirittura del 96% se posta 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’aggiunta una serie di documenti da cui poter estrarre tutte le informazioni riguardanti la bottiglia richiesta (quali il tipo di vino, la regione di provenienza, la gradazione alcolica e così via). Inoltre, per poter soddisfare al meglio le richieste dell’utente sarebbe necessario ampliare il dataset di immagini *gt* (al momento piuttosto ristretto), ed aggiungere un’interfaccia utente adeguata in modo da rendere il programma facilmente consultabile. Infine, una futura ricerca potrebbe permettere di passare dall’utilizzo di un software a pagamento per il riconoscimento (come quello messo a disposizione da Microsoft) a un nuovo e (possibilmente) miglior software che non necessiti l’uso di licenze.

Bibliografia

- [1] A. Bodini, “Wine bottles recognition.”
- [2] R. Hemming, “Best wine label scanning apps 2016,” *JancisRobinson.com*, 04 2016.
- [3] J. Lim, S. Kim, J. Park, G. Lee, H. Yang, and C. Lee, “Recognition of text in wine label images,” pp. 1–5, Nov 2009.
- [4] A. M. Arboleda and C. Arce-Lopera, “Quantitative analysis of product categorization in soft drinks using bottle silhouettes,” *Food Quality and Preference*, vol. 45, pp. 1 – 10, 2015.
- [5] A. C. Tao Wang, David J. Wu and A. Y. Ng., “End-to-end text recognition with convolutional neural networks,” 2012.
- [6] G. Fairchild, “pyxdameraulevenshtein.”