

Progetto di Machine Learning

Anno Accademico 2024-2025

Matteo Galletta, Marco Gionfriddo, Kevin Speranza

Indice

1	Problema	1
2	Soluzione	2
3	Dataset	3
4	Metodi	4
5	Valutazione	5
6	Esperimenti	6
7	Demo	7
8	Codice	8
9	Conclusione	9

Capitolo 1

Problema

Da anni ormai si tenta di risolvere il problema del riconoscimento di testi contenuti in immagine. Il problema è noto come Optical Character Recognition (OCR) e consiste nel riconoscere i caratteri di un testo contenuto in un'immagine. Il problema è complesso e presenta una serie di insidie che non sono di immediata risoluzione. Nonostante ciò, allo stato dell'arte esistono diversi algoritmi che consentono di ottenere risultati soddisfacenti. Quello che viene presentato in questo documento è un modello che mira a semplificare il problema a una sottoclasse di immagini, avendo il vantaggio di ottenere un algoritmo più leggero ed efficiente, a discapito della sua versatilità.

Capita spesso che le immagini da cui è utile estrarre il testo siano screenshot. L'algoritmo presentato si occupa di estrarre il testo contenuto in uno screenshot, indipendentemente dal font e dai colori utilizzati. In realtà, viene inizialmente affrontato il problema assumendo che il testo presenti una sola parola. Tramite l'uso di euristiche, si può estendere facilmente l'implementazione comprendendo frasi (purché non siano divise su più righe).

TODO: immagini

Capitolo 2

Soluzione

TODO: spostare sto capitolo?

La soluzione proposta è realizzata tramite una pipeline composta da due fasi principali:

- **Fase 1:** Suddivisione della parola in caratteri
- **Fase 2:** Classificazione del singolo carattere

Per la prima fase vengono utilizzare euristiche e algoritmi di image processing per suddividere la parola in caratteri. Per la seconda fase viene utilizzato un modello di deep learning per classificare i singoli caratteri, che viene approfondito nei capitoli successivi.

TODO: mettere schema pipeline

Capitolo 3

Dataset

Essendo il problema dell'OCR uno dei più studiati in ambito di Computer Vision, esistono diversi dataset pubblici che possono essere utilizzati per addestrare e testare i modelli. Tuttavia, la maggior parte di questi dataset sono stati creati per risolvere problemi generali e non sono specifici per il riconoscimento di testi contenuti in screenshot. Per questo motivo, è stato necessario creare una coppia di dataset ad hoc per il problema in questione.

In particolare, essendo l'algoritmo diviso in due fasi, avere due dataset distinti consente di poter valutare in modo individuale ognuna delle due fasi, consentendo di valutare l'accuratezza del modello in modo più preciso. Il primo dataset è composto da immagini di screenshot contenenti una sola parola. Questo consente di testare la pipeline nella sua interezza, comprendendo sia la suddivisione in caratteri che la classificazione del singolo carattere. Il secondo dataset è composto da immagini di singoli caratteri, che consente di testare esclusivamente la fase di classificazione del singolo simbolo. I dettagli sulla sintetizzazione vengono approfonditi nel capitolo ??.

Capitolo 4

Metodi

Capitolo 5

Valutazione

Capitolo 6

Esperimenti

Capitolo 7

Demo

Capitolo 8

Codice

Capitolo 9

Conclusione