

Tecniche di incremento dei dati per la ricerca di immagini basata su contenuti con CNN

Candidato: Irene Dini irene.dini1@stud.unifi.it
Relatore: Marco Bertini marco.bertini@unifi.it

Abstract

La *computer vision* è quell'ambito dell'apprendimento automatico che si occupa di sviluppare sistemi in grado di comprendere la composizione strutturale e semantica di immagini e video. Una delle sue principali aree di interesse è la ricerca di immagini basata su contenuti il cui scopo è, data un'immagine detta *query*, trovare all'interno di una base di dati di grandi dimensioni, tutte e sole le immagini che raffigurano lo stesso scenario o gli stessi oggetti raffigurati nella query. Il punto focale della ricerca di immagini è la creazione di un descrittore per le immagini: un vettore di numeri reali tale che, data una funzione di distanza, immagini simili abbiano descrittori vicini e immagini diverse abbiano descrittori lontani.

Durante lo sviluppo di questa tesi, la base di dati utilizzata per valutare le prestazioni della ricerca di immagini è *Oxford Buildings*, che contiene 5062 fotografie scattate ad Oxford, in 11 luoghi di interesse. La metrica scelta per la valutazione è invece la *Mean Average Precision* (mAP). Per generare i descrittori delle immagini abbiamo utilizzato *VGG16*, una rete neurale convoluzionale ideata per eseguire la classificazione delle immagini di ImageNet, a cui abbiamo tolto i livelli per la classificazione e su cui abbiamo effettuato un *fine-tuning* utilizzando le immagini di *Paris*, una base di dati contenente foto scattate a Parigi.

Abbiamo poi studiato l'efficienza di tecniche di incremento dei dati. Queste tecniche consistono nell'aumentare il numero delle immagini a disposizione, applicando delle trasformazioni geometriche e sul colore alle immagini che si hanno. La difficoltà nella loro applicazione sta nel trovare quali trasformazioni è conveniente applicare e con quale intensità. In un primo momento abbiamo applicato le trasformazioni individuate dagli sviluppatori dell'algoritmo *AutoAugment*, caratterizzato da molti gradi di libertà ma computazionalmente molto costoso, effettuando l'ottimizzazione su dataset diversi dal nostro. Le prestazioni della ricerca di immagini sono migliorate dell'1.5%. Abbiamo poi ottimizzato l'algoritmo *Randaugment*, con soli 2 parametri ma molto più veloce, direttamente su Oxford. In questo caso l'incremento delle prestazioni è stato circa del 5%. L'ultimo esperimento effettuato è stato quello di combinare descrittori ottenuti ottimizzando la rete neurale con immagini su 2 diverse scale, in modo da catturare 2 livelli diversi di risoluzione, sempre facendo uso di tecniche di incremento dei dati. Rispetto al solo utilizzo di queste ultime c'è stato un ulteriore incremento delle prestazioni del 5%.