



# Riconoscimento e classificazione di gemme tramite convolutional neural network

Francesca Sellarione





# Abstract

Il progetto si pone l'obiettivo di sviluppare un sistema per il **riconoscimento di gemme lavorate**, quindi pietre preziose e semi-preziose, al fine di identificarle.

Questo viene fatto a partire da un dataset di 88 gemme diverse contenente inizialmente circa 3000 diverse immagini, ripulito ed espanso fino a **4400 elementi**.

Viene impiegata una rete neurale convolutiva, che permette di ottenere un sistema in grado di riconoscere le gemme con un'accuratezza di circa il 65%, in congiunzione con altre tecniche.

I risultati sono stati analizzati anche con l'utilizzo della **top-k-accuracy**, nell'ottica di creare un programma che restituisca non il primo risultato predetto, ma i primi k.





# Motivazione

Non tutti sono in grado di identificare una gemma, ed in alcuni casi può persino capitare che una gemma venga dichiarata, ad esempio da un venditore, per ciò che non è.

Sarebbe perciò utile a persone non particolarmente esperte un sistema che, con la giusta quantità di dati a supporto, sia in grado di riconoscere gemme a loro sconosciute, o **disambiguare** situazioni in cui l'identità di una data gemma è dubbia.





# Dataset

Il dataset di base è stato ottenuto da Kaggle e contiene circa 3200 immagini divise in 87 classi diverse, ciascuna rappresentante una gemma. Il tutto è già diviso in training e test set secondo la classica **proporzione 80/20**.

Partendo da questa base, al dataset è stata aggiunta un'ulteriore classe e ulteriori 1200 immagini, normalizzando tutte le classi presenti ed **eliminando le immagini duplicate o di qualità eccessivamente scarsa**. Il risultato è stato poi nuovamente pubblicato su Kaggle, citando l'autrice originale.

Il risultato finale è un dataset di 4400 immagini, 50 per ogni gemma.





# Preparazione e pulizia dei dati

Il dataset iniziale non è totalmente normalizzato e contiene duplicati e immagini di qualità troppo bassa o con dettagli fuorvianti. Oltre ad aggiungere un'ulteriore classe, il numero delle immagini per ciascuna è stato portato da 30-35 a 50, eliminando anche i duplicati.

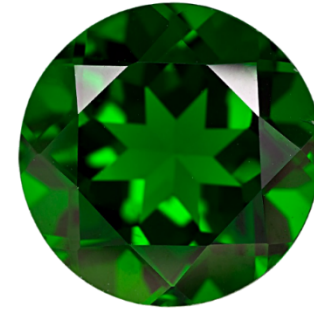
La preparazione dei dati consiste principalmente nel **ridimensionamento** delle immagini e nella **divisione casuale** in train, validation e test set con una proporzione di 80/10/10, per non penalizzare troppo l'addestramento del modello con una dimensione del training set troppo piccola.





# Obiettivo della ricerca

- E' possibile distinguere due gemme di natura diversa, ma con caratteristiche simili fra loro, avendo a disposizione solo delle immagini?
- Con quale accuratezza un sistema di deep learning è in grado di distinguere oggetti lavorati in modi molto simili, ma con colori, proprietà ottiche e composizioni differenti?





# Metodologia – Rete Neurale Convolutiva

Per la classificazione delle gemme è stata implementata una **rete neurale convolutiva** (CNN), in grado di estrarre in maniera automatica le caratteristiche rilevanti di un dato fornito in ingresso e di apprenderne i pattern. Questo la rende un'architettura particolarmente adatta al riconoscimento di immagini.

Le operazioni fondamentali di una CNN sono la **convoluzione** ed il **sottocampionamento** (pooling) e spesso tali reti sono formate da più livelli convoluzionali impilati: nei primi livelli la rete cercherà di apprendere le caratteristiche più elementari, per poi passare a quelle più complesse man mano che ci si scende in profondità nella rete, fino ad arrivare ad identificare l'oggetto nella sua complessità.

La rete implementata consiste di:

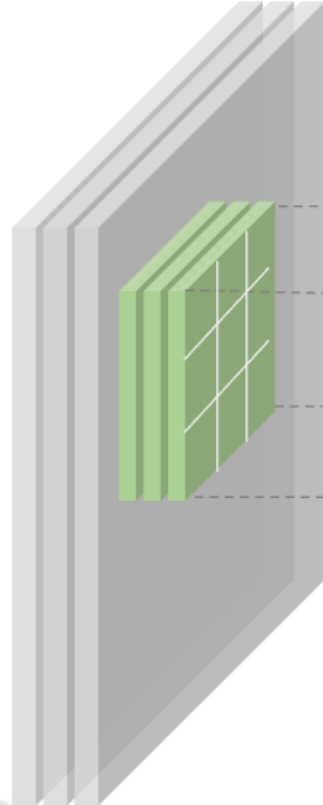
- 6 strati convolutivi, ognuno di  $32 \times k$  filtri con finestre  $3 \times 3$  ed ognuno seguito da uno strato di max pooling con finestre  $2 \times 2$  e passo  $2 \times 2$
- 1 strato di appiattimento (Flatten) seguito da uno strato di Dropout del 30%
- 2 strati Dense con tasso di abbandono del 30%
- 1 strato Dense di uscita di 88 neuroni Softmax



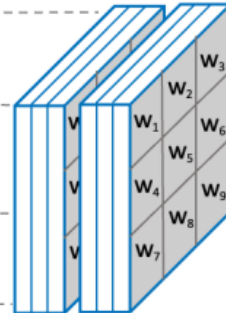


# Metodologia – Rete Neurale Convolutiva

Immagine 220 x 220 x 3



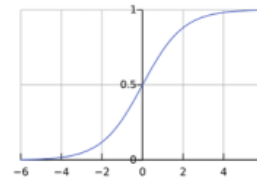
2 filtri con kernels 3 x 3 ciascuno



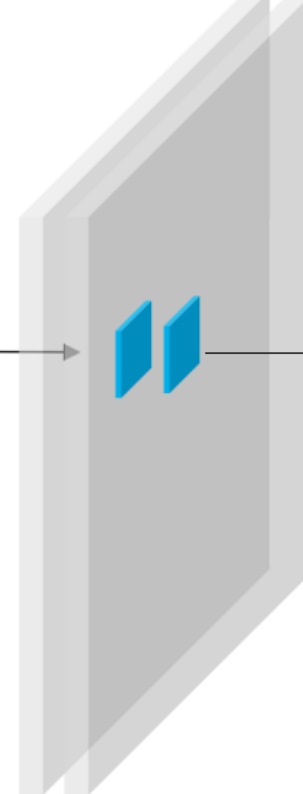
56 pesi totali

Funzione di attivazione (ReLU)

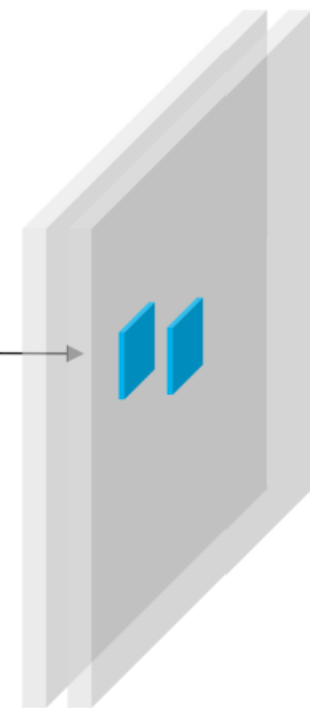
$$f(wx + b)$$



Mappa di attivazione 220 x 220 x 2



Max Pooling 110 x 110 x 2







# Metodologia – altre tecniche utilizzate

Nel tentativo di migliorare le prestazioni del modello sono state utilizzate varie tecniche, tra cui:

- Operazioni di **data augmentation** per l'aumento del numero degli esempi in maniera artificiosa: applicazione di trasformazioni casuali, come ad esempio operazioni di zoom, rotazione e ribaltamento.
- Rimozione dello **sfondo** da tutte le immagini del dataset tramite apposite librerie
- Ritaglio delle immagini lungo i **bordi** della gemma, senza alterarne le proporzioni

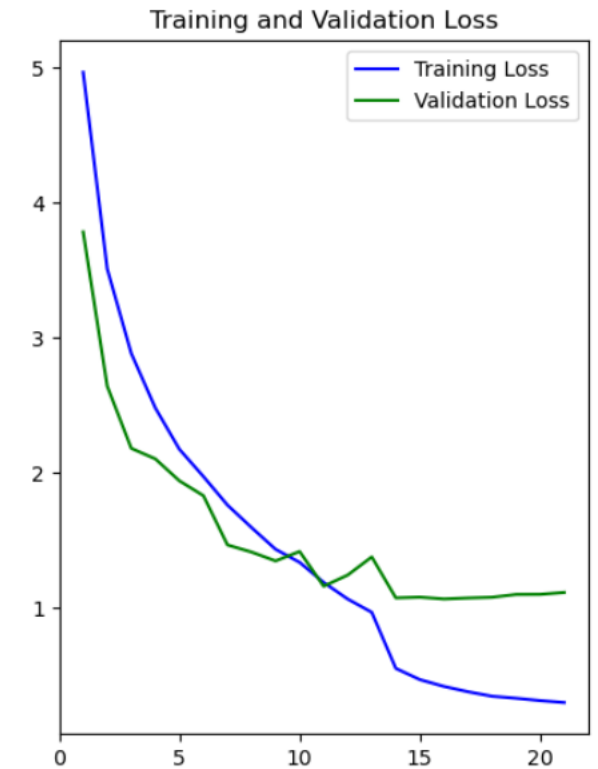
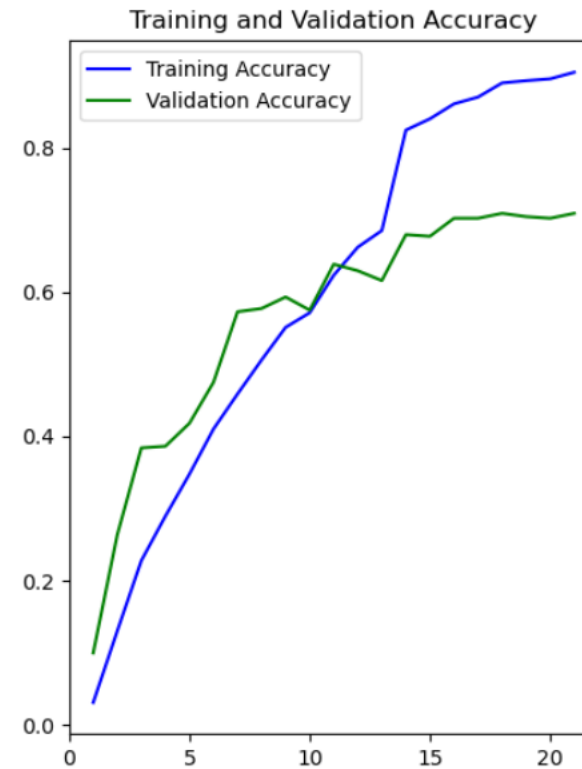




# Risultati

Sia le operazioni di data augmentation, sia le tecniche di pre-elaborazione delle immagini come la rimozione dello sfondo e il ritaglio della gemma lungo i bordi, non hanno apportato miglioramenti, a riprova del fatto che la CNN apprende in modo automatico le feature caratteristiche di un'immagine.

La versione finale della rete raggiunge una accuracy che si attesta intorno al **70%** sul validation set ed al **65%** sul training set.





# Risultati

Considerando, come discusso, che esistono casi di quasi indistinguibilità tra gruppi di 2/3 gemme e che il numero delle classi è elevato, è stata presa in considerazione anche la **top-3 accuracy**, ovvero la possibilità che la classe corretta si trovi nei primi 3 risultati predetti. In questo caso, il valore si alza fino all'85% e sale al 94% in caso di top-5 accuracy.

Questo testimonia nuovamente quanto alcune gemme siano simili sia per l'occhio umano che per la CNN, la quale è disegnata proprio per funzionare come l'occhio umano.

Immagine esaminata:  
Almandine



Predizione 1:  
Pyrope



Predizione 2:  
Garnet Red



Predizione 3:  
Almandine





# Limitazioni

Una prima limitazione del modello è il non riuscire a generalizzare su **pietre grezze** o con **lavorazioni non standard** rispetto a quelle con cui è stato addestrato, come nel caso di gemme incastonate.

Inoltre spesso anche un occhio esperto non è in grado di riconoscere una gemma solo da un'immagine, e sono necessarie informazioni aggiuntive sulla **rifrazione della luce** o sulla **durezza nella scala Mohs**.

Oltre ad usare metriche come quelle discusse, per mitigare il problema si potrebbe ridurre il numero di classi o fare un **clustering** unificando le classi indistinguibili in una superclasse, anche se ciò non porterebbe un concreto vantaggio rispetto al semplice utilizzo della top-3 accuracy.





# Conclusioni

Rispetto agli obiettivi di ricerca posti, possiamo affermare che:

- il sistema è in grado di eseguire una distinzione basata sull'immagine, con un esito che diventa più incerto se ci si trova a classificare una gemma con aspetto troppo simile ad un'altra.
- Considerata la natura del problema e la numerosità delle classi, l'utilizzo della top-3 accuracy si è rivelato necessario ed utile per comprendere le potenzialità del sistema. Tramite la matrice di confusione è poi stato possibile individuare i casi di ambiguità.

In conclusione, le prestazioni sono **in linea con le aspettative** iniziali.

['Chrome Diopside', 'Tsavorite', 'Andradite']



['Tsavorite', 'Emerald', 'Chrome Diopside']





# Riconoscimenti

- **Daria Chemkaeva, autrice del dataset originale, usato come base.**
- **La revisione e l'ampliamento del dataset di partenza sono stati effettuati da me personalmente.**





# Riferimenti bibliografici

- « Transfer Learning with Convolutional Neural Networks for Diabetic Retinopathy Image Classification. A Review »
- « Implementazione di una Rete Neurale Convoluzionale con Dataset Custom per la Classificazione di Veicoli » (Tesi di Laurea — Neri Roberto)
- <https://learnopencv.com/understanding-convolutional-neural-networks-cnn/>

