Progetto di analisi di immagini e video:
costruzione di un classificatore multi-label di trailer cinematografici

Studenti:

Francesca Senatore matr. 214636

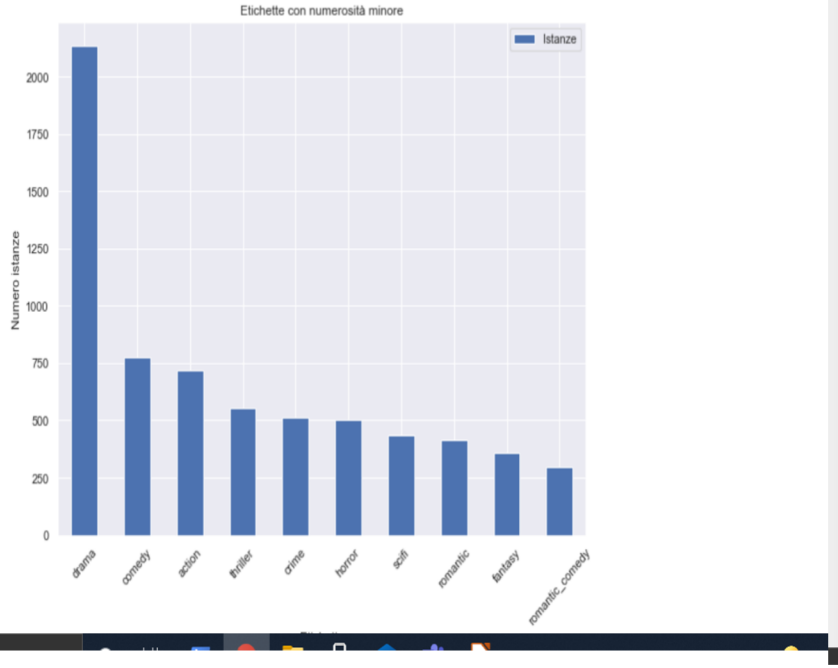
Giuliano Stirparo matr. 214533

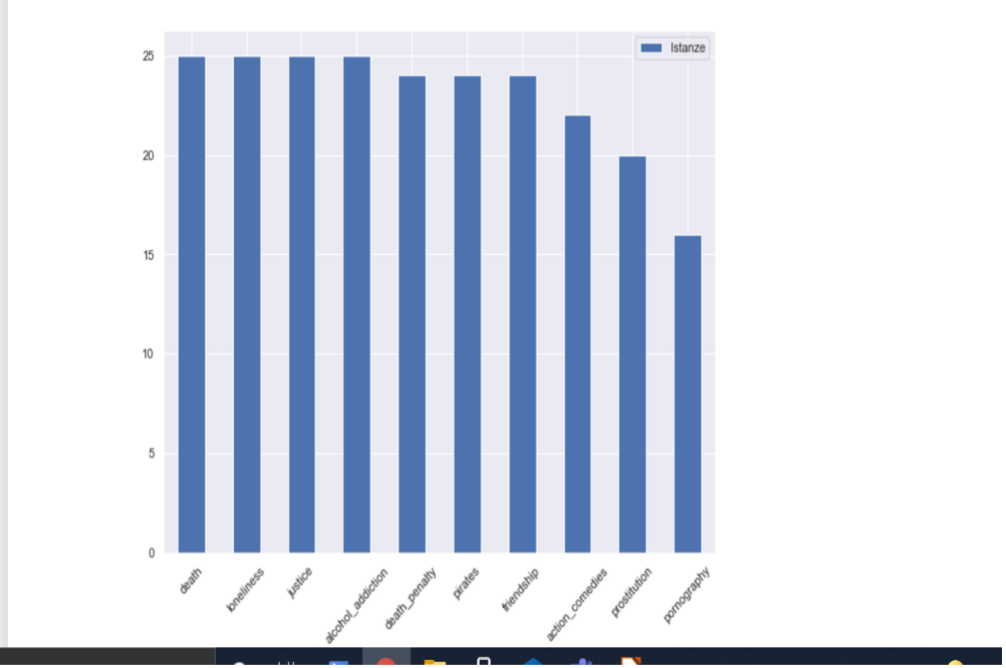
Il dataset

- 4292 oggetti nel training set
- 1112 oggetti nel test set
- Ogni oggetto è una cartella contenente un numero variabile di frame tratti dal film in questione

Il dataset

```
Text(0, 0, 'pirates'),
Text(0, 0, 'friendship'),
Text(0, 0, 'action_comedies'),
Text(0, 0, 'prostitution'),
Text(0, 0, 'pornography')]
```





Gestione sbilanciamento dataset

L’approccio utilizzato consiste nel:

- **Dare peso inferiore alle classi più numerose**
- **Dare peso superiore alle classi meno numerose**

Gestione sbilanciamento dataset

Pesi utilizzati:

- $P_i = 1/f_i$, dove f_i è la numerosità della classe i-esima

Costruzione dei mini-dataset

- Sono stati costruiti un mini-training set e un mini-test set di dimensione 1/10 rispetto alle dimensioni originali che permettessero di fare prove veloci
- Entrambi i dataset ridotti rispecchiano le proporzioni di classe del dataset di partenza

Pre-processing dei dati

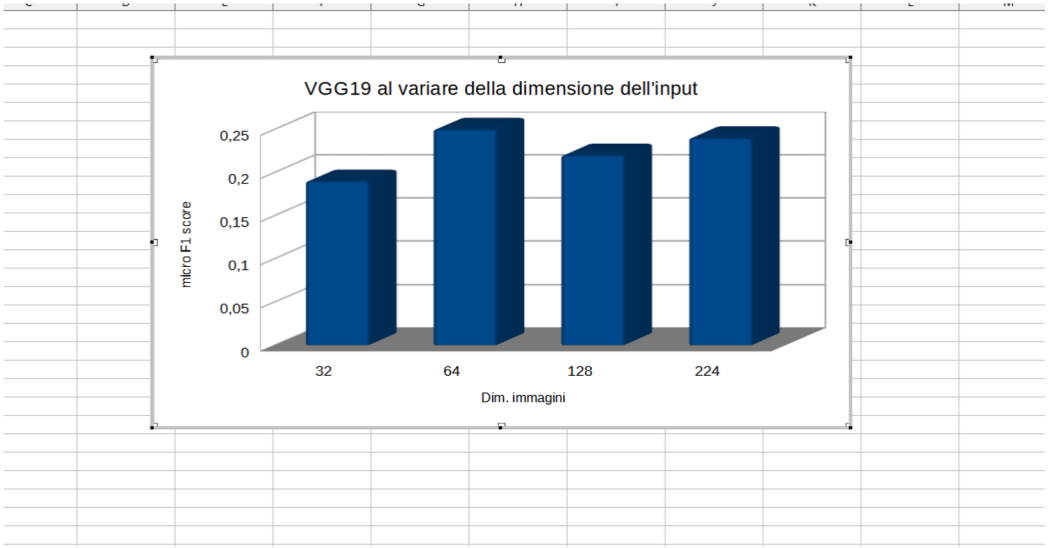
.Detto n il numero di frame presenti in un certo trailer, le immagini effettivamente utilizzate in fase di train e di test sono 20 nel caso in cui il trailer presenti n >= 20, altrimenti esattamente n.

.Center-crop a 224

.Resize a 64x64

Altri tentativi di pre-processing

.Sono state provate anche resize a 32, 128 e 224, ottenendo, a parità di rete, sempre risultati peggiori.



Altri tentativi di pre-processing

- Filtro gaussiano
- ColorJitter
- Grayscale
- RandomAffine

- Nessuno ha prodotto miglioramenti significativi.

Altri tentativi di pre-processing

Sono stati fatti vari tentativi di ridimensionamento del numero di frame e delle modalità di sampling dall'oggetto di partenza:

- **Varie dimensioni (10,15,20,25,30, tutte)**
- **Varie modalità di sampling: primi k frame, ultimi k frame, k frame random**

- **Il risultato migliore è stato ottenuto selezionando sempre i primi 15 frame.**

Reti testate

- VGG
- ResNet
- AlexNet

Rete utilizzata

- **VGG19 pretrained. In particolare la rete è stata riadattata a restituire dalla parte convoluzionale delle feature map di dimensione 512x2x2, che sono passate in input alla parte di classificazione**
- **Parte di classificazione anch'essa riadattata, aggiungendo una serie di livelli densi che in linea di massima dimezzano ogni volta le dimensioni della feature map ricevuta.**
- **Le 20 feature map restituite dal livello convoluzionale sono gestite effettuandone una media, avendo quindi come risultato una feature map unica di dimensione 512x2x2, che è passata in input ai livelli di classificazione.**

Fase di training

- Funzione di loss utilizzata: Binary Cross Entropy
- Learning rate pari a 0.01
- Ottimizzatore SGD con momentum = 0.9

Fase di test

.Sono state rilevate 10 classi

.Gli score ottenuti sono i seguenti:

.Micro F1 score: 0.25

.Micro precision: 0.16

.Micro recall: 0.55