



UNIVERSITÀ DELLA CALABRIA

DIPARTIMENTO DI INGEGNERIA INFORMATICA,
MODELLISTICA, ELETTRONICA
E SISTEMISTICA

DIMES

PROGETTO DI ANALISI IMMAGINI E VIDEO

Simonpaolo Lopez, matricola: 216759

Mattia Gatto, matricola: 216649



Introduzione

Il progetto si pone l'obiettivo di classificare la tipologia di un trailer a partire dai keyframe estrapolati.

E' un problema di multi classificazione con 85 possibili etichette assegnabili al singolo trailer.

I dataset messi a disposizione sono 4, di cui solo il primo etichettato.

Il dataset etichettato è composto da 5404 trailer di cui:

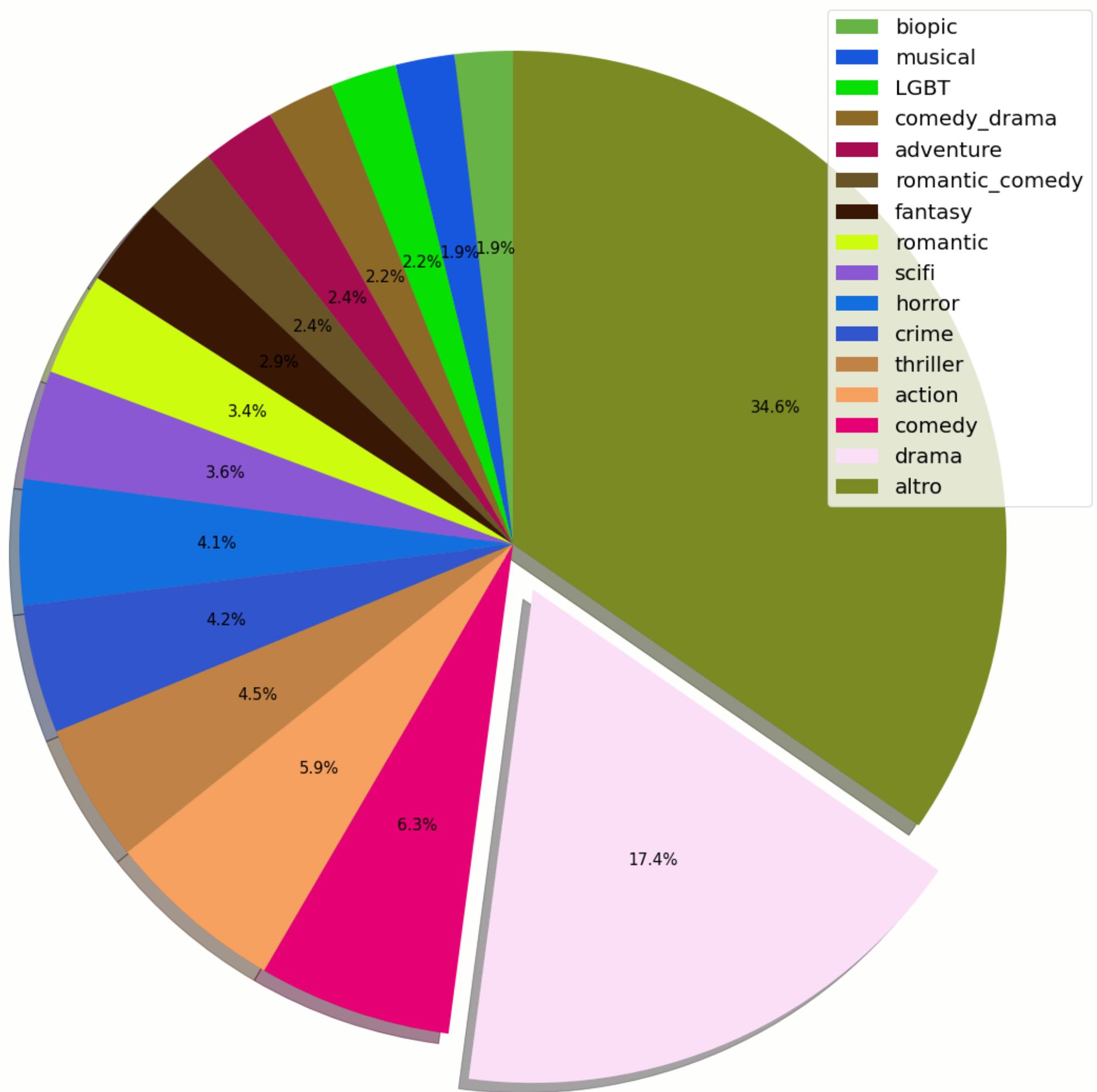
- 4292 appartenenti al training-set
- 1112 appartenenti al test-set

Data analisys

Dal grafico delle 15 maggiori etichette si evince l'alto sbilanciamento dei dati.

Ogni trailer presenta un numero di frame variabile tra [3,80].

Con una media di circa 20 frame a trailer.



Preprocessing

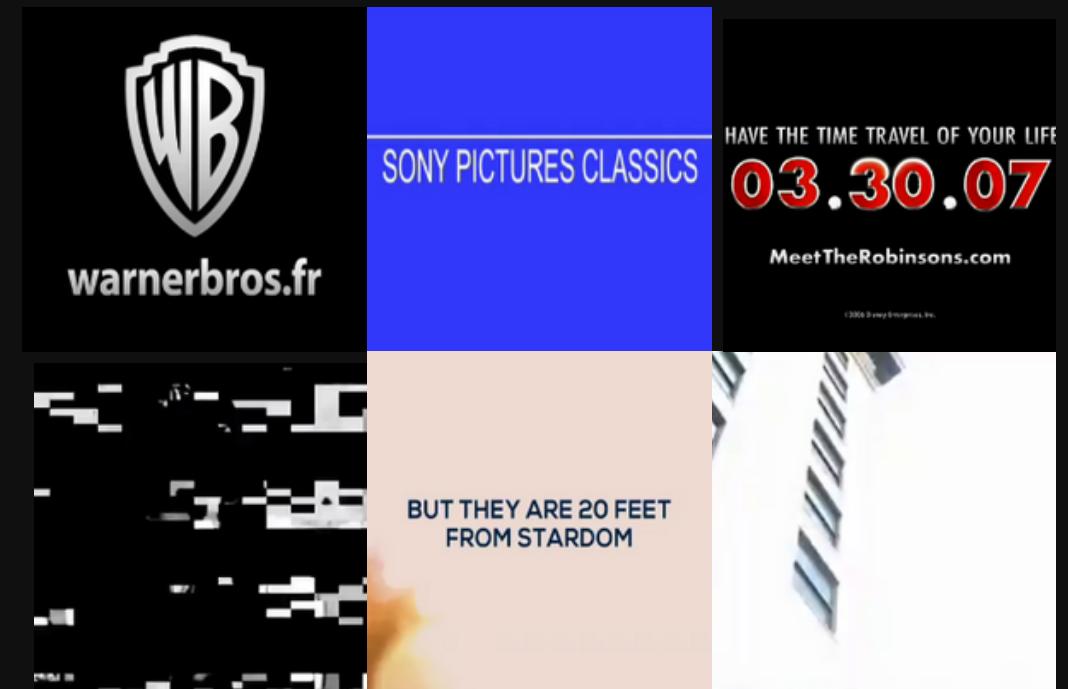
Durante la fase di preprocessamento dei dati abbiamo eseguito diverse operazioni:

- Analizzando i vari trailer abbiamo supposto che per aiutare la classificazione le tonalità di colori delle immagini nei trailer potevano risultare sicuramente più caratteristiche del riempimento con un semplice frame nero.
- Si è quindi fatta una media delle immagini presenti in ogni direcotory.
- Inoltre, abbiamo ricercato le immagini inutili per l'etichettatura di un trailer per un totale di 676 frame e le abbiamo sostituite con l'immagine media del trailer.

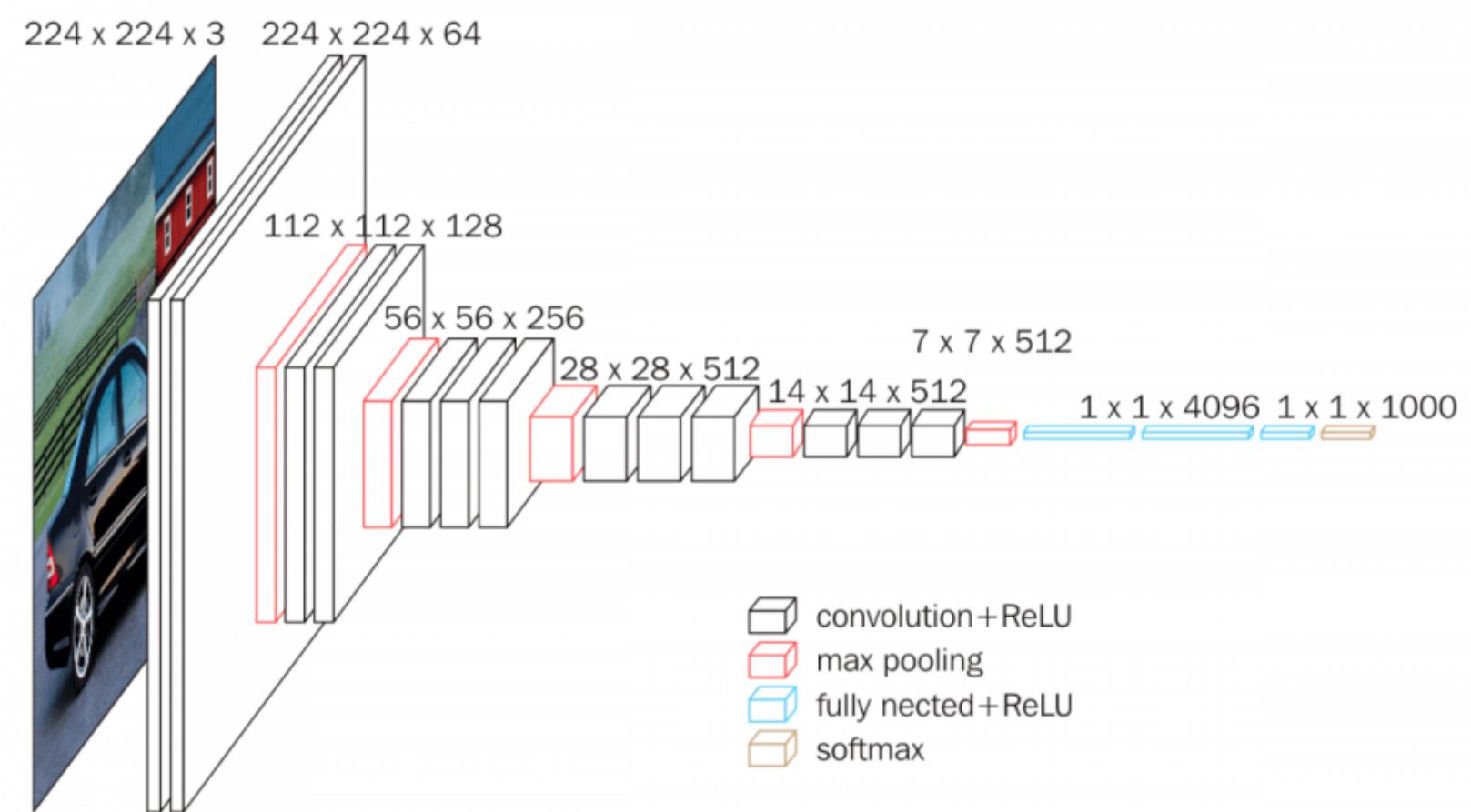
immagini media



immagini inutili



Architettura



Si sono provate varie trasformazioni sulle immagini, tra cui molteplici di data augmentation che però non hanno dato buoni risultati. Quindi si è tenuti soltanto la centerCrop a 224 x 224, e la trasformazione in tensore.

VGG16

I migliori risultati in termini di micro_avg f1_score e n°di classi con f1_score>0 si sono riscontrati utilizzando una vgg16 pretrained, classificando l'intero batch di 16 trailer.
inizialmente prendiamo la singola directory e la passiamo alla vgg intera compresa la parte di fullyconnected, e successivamente i risultati, inseriti in un batch di 16 elementi sono stati classificati attraverso una rete densa.

Ottimizzatore e loss usate

- l'ottimizzatore utilizzato: Adam con un learning rate pari a 0.001
- La loss utilizzata: BCEloss con assenza di pesi inizializzati

Tentativi

Abbiamo provato diverse loss tra le quali:

- BCEwithLogitsLoss()
- PoissonNLLLoss()
- CrossEntropyLoss()

Abbiamo anche provato ad inizializzare i pesi in proporzione alle occorrenze delle singole etichette tramite due differenti formule:

$$1) \frac{1}{numOcc} \quad 2) \frac{occEtichetta}{maxOcc}$$

Valutazione dei risultati

Il modello è stato addestrato per numerose epoche ma il risultato migliore è stato ottenuto alla 26-esima, con un avg f1-score =0.27 e 36 classi etichettate con f1-score>0.

	precision	recall	f1-score	support
LGBT	0.09	0.38	0.15	61
action	0.19	0.97	0.31	191
action_comedies	0.00	0.00	0.00	6
adventure	0.14	0.61	0.22	75
alcohol_addiction	0.00	0.00	0.00	6
alien	0.00	0.00	0.00	12
animation	0.31	0.57	0.40	21
aviation	0.25	0.05	0.08	20
bank_robbery	0.33	0.05	0.09	19
biopic	0.07	0.51	0.12	59
cartoon	0.15	0.38	0.21	8



war	0.10	0.47	0.16	58
wartime	0.09	0.63	0.16	49
western	0.16	0.42	0.23	33
zombies	0.00	0.00	0.00	15
micro avg	0.17	0.61	0.27	3167
macro avg	0.07	0.18	0.08	3167
weighted avg	0.18	0.61	0.26	3167
samples avg	0.17	0.65	0.26	3167

Altre Tecniche utilizzate

Altre Architetture e tecniche provate

VGG19_BN, VGG19, VGG11

queste architetture hanno portato buoni risultati ma non ottimi in termini di numero classi etichettate con $f1_score > 0$ e di $f1\text{-}score$ di micro_avg .

UTILIZZO DI UN BOUND

Durante la fase di training nella selezione del batch da 16 elementi saltavamo un batch se esso conteneva un numero di etichette, che risultavano nelle TOP 35, inferiore ad una data soglia.

RESNET50

Durante le epoche stampando di volta in volta il n° di classi etichettate con $f1_score > 0$ e il valore di $\text{mirco_avg f1_score}$ risultava essere più basso rispetto alle tecniche di VGG.

Autoencoder

UTILIZZO DATI UNLABELED E LABELED

Per le prime 20 epoche abbiamo addestrato l'autoencoder con il dataset etichettato, per poi alternare ogni 10 epoche con i restanti 3 dataset non etichettati.

UTILIZZO DELL'AUTOENCODER

Durante la costruzione del dataset, invece di sostituire con i frame medi , abbiamo inserito i frame della directory ricostruiti con l'autoencoder per identificare più feature.

RISULTATI

In termini di valuzione, i risultati non sono stati abbastanza soddisfacenti.



Grazie per
l'attenzione