

Trailer classifier - CV Project



Barbara Vito 214254
Bellusci Pierpaolo 214244

Introduzione

L'obiettivo del progetto è quello di creare un modello in grado di predire i generi di un trailer a partire da un set di keyframes.

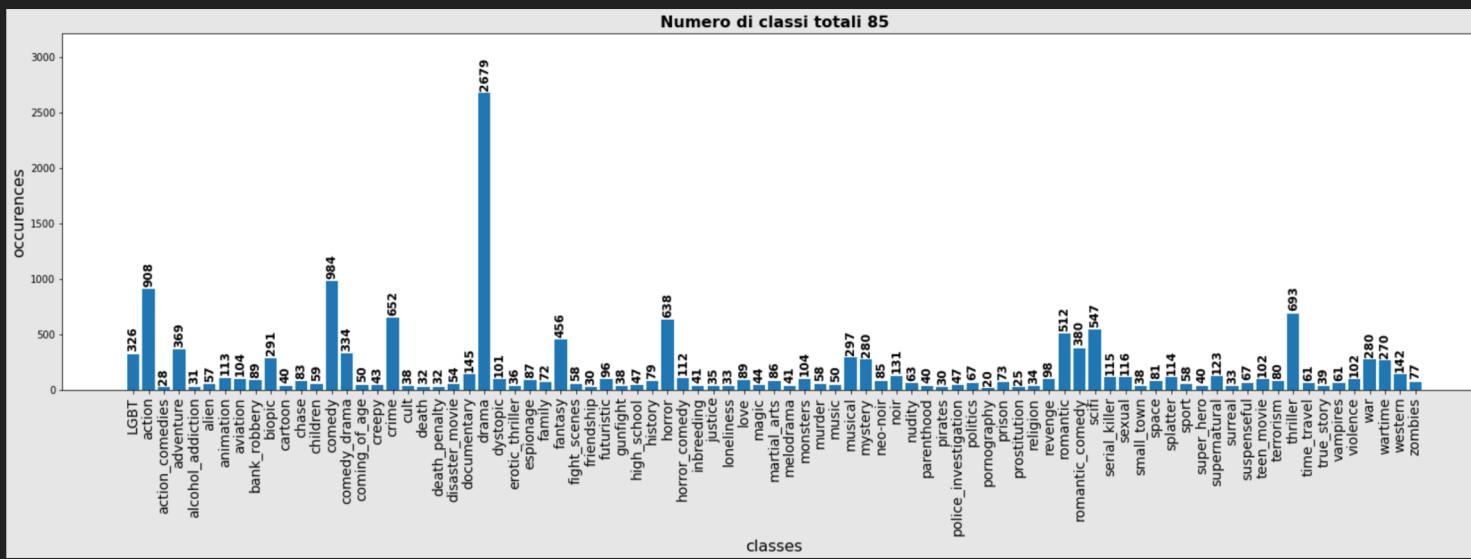
Il problema è di tipo multiclass-multilabel, con 85 possibili generi.

Per lo svolgimento del progetto sono stati utilizzati 2 datasets:

- labeled dataset: 4292 train samples, 1112 test samples
- unlabeled dataset 1: 4944 samples

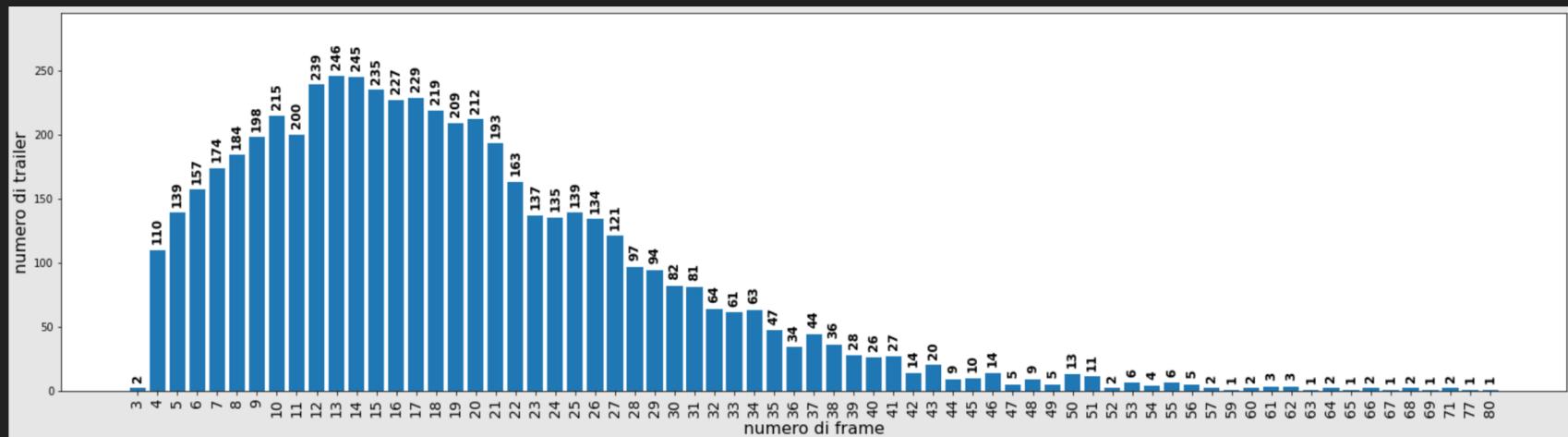
Data Understanding

Il dataset labeled presenta una distribuzione altamente sbilanciata che rende più complesso il task.



Data Understanding

- Ogni trailer presenta un numero variabile di frame [3,80].
- Durante il training sono stati utilizzati al massimo 20 frames per sample, presi in modo casuale ad ogni epoca.

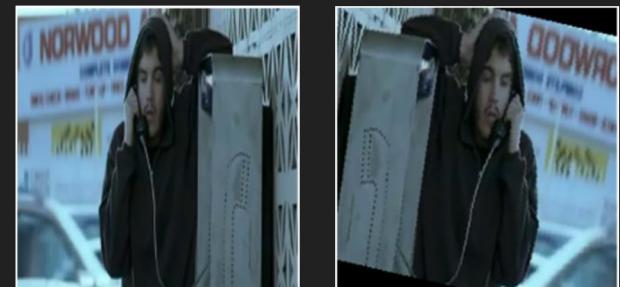


Data Preparation - Labeled

Durante questa fase, le immagini sono state normalizzate e ridimensionate a 224x224 pixel.

Inoltre è stato applicato il processo di Data Augmentation che comprende le seguenti operazioni:

- Rotazione con un valore random tra [-15; 15]
- Mirroring con una probabilità del 50%
- Scale con un offset random tra [0.8,1.0]



Esempio di *Data Augmentation*

Data Preparation - Unlabeled

Come le labeled, le immagini sono state normalizzate e ridimensionate a 224x224 pixel.

Data l'assenza di etichette è stata usata la tecnica self-supervised SESEMI.

- ogni trailer è stato trasformato e associato a una delle seguenti classi:
 - 0° , 90° , 180° , 270° , horizontal flip, vertical flip
- questo ci permette di utilizzare le unlabeled in modo supervisionato per l'estrazione delle features

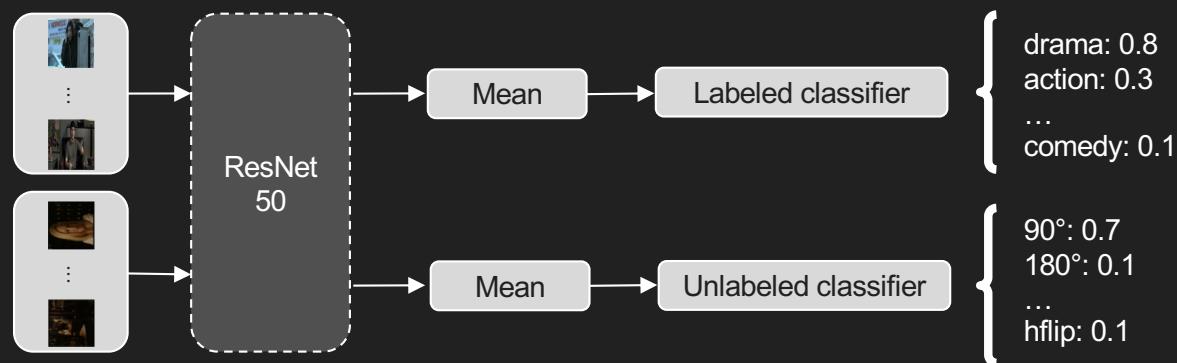


Esempio di classe 90°

Modeling

L'architettura del modello è composta da:

- Un features Extractor comune, Resnet50, con l'ultimo blocco allenabile (fine-tuning)
- Due classificatori distinti, con stessa struttura Linear, Relu, Dropout
 - Labeled classifier ~ 2.8M parametri
 - Unlabeled classifier ~ 1.2M parametri



Modeling - Optimizers

A differenza del framework originale Sesemi, abbiamo utilizzato due ottimizzatori Adam distinti, in modo tale che l'errore di un classificatore non abbia ripercussioni sull'altro.

Modeling - Losses

Due loss distinte:

- Unlabeled Loss: **Categorical Cross-Entropy**
- Labeled Loss: **Asymmetric Loss**

$$ASL = \begin{cases} L_+ = (1-p)^\gamma + \log(p) \\ L_- = (p_m)^\gamma - \log(1-p_m) \end{cases} \quad \text{con} \quad p_m = \max(p - m, 0)$$

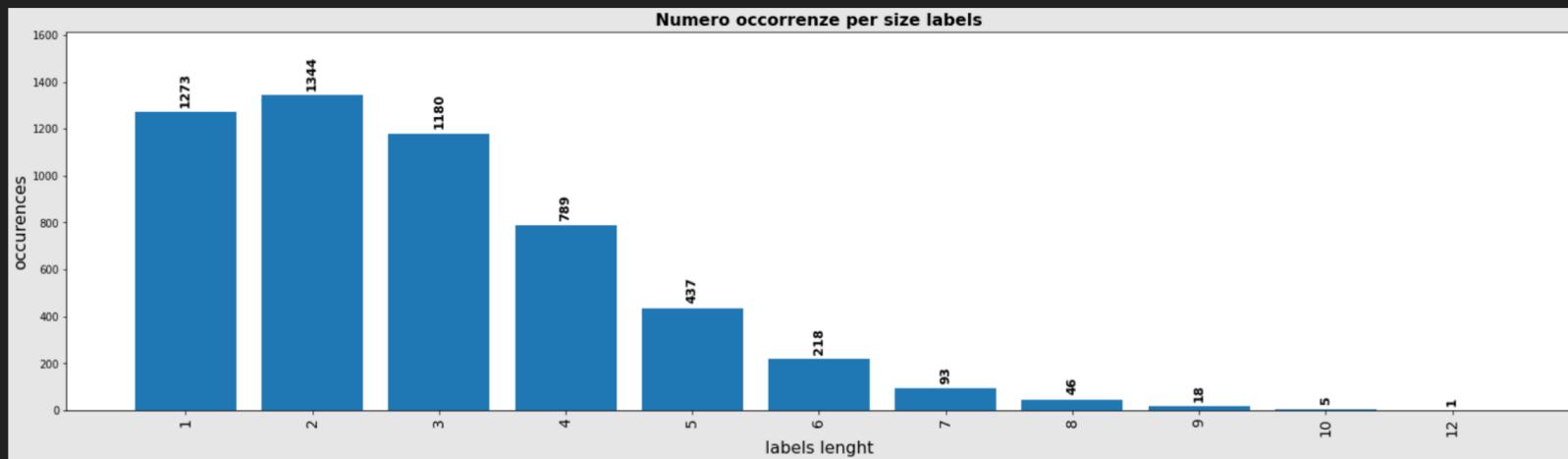
Evaluation

Il modello è stato allenato per molte epoche, ma il risultato migliore è stato ottenuto intorno alla decima.

	precision	recall	f1-score	support
LGBT	0.08	0.75	0.15	61
action	0.23	0.93	0.37	191
comedy	0.19	1.00	0.32	211
...				
drama	0.49	1.00	0.66	546
horror	0.21	0.82	0.33	136
romantic	0.11	0.88	0.20	99
thriller	0.16	0.94	0.28	139
micro avg	0.18	0.62	0.28	3167
macro avg	0.05	0.19	0.08	3167
weighted avg	0.18	0.62	0.26	3167
samples avg	0.18	0.67	0.27	3167

Evaluation

Da un'analisi del dataset, si è notato che la maggior parte dei trailer ha un numero di labels inferiore a 5 e questo ci porta ad avere un gran numero di False Positive utilizzando come top-k un valore alto.



Evaluation

Infatti alla diminuzione del top-k i risultati migliorano notevolmente:

		precision	recall	f1
Top-10:	micro avg	0.18	0.62	0.28
Top-9:	micro avg	0.19	0.60	0.29
Top-8:	micro avg	0.20	0.57	0.30
Top-7:	micro avg	0.22	0.54	0.31
Top-6:	micro avg	0.24	0.50	0.32
Top-5:	micro avg	0.26	0.45	0.33
Top-4:	micro avg	0.29	0.41	0.34

I risultati migliori in f1 li abbiamo ottenuti approssimando le probabilità all'intero più vicino:

round: micro avg **0.32** **0.39** **0.35**

Tentativo SimClr

Un ulteriore approccio provato per le unlabeled è stato **SimClr**

- Due allenamenti separati:
 - Allenamento su dati Unlabeled
 - Fine-Tuning con dati Labeled
- Contrastive Learning per estrazione features

I risultati ottenuti sono stati uguali al modello precedentemente analizzato $f1 = 0.28$.

E' stato scelto l'altro approccio per la sua semplicità e soprattutto per il fatto che SimClr ha richiesto molto più tempo per convergere.

Grazie