Politecnico di Torino
Dipartimento di Automatica e Informatica

PhD in Computer and Control Engineering
XXXVI cycle

Supervisors
Tania Cerquitelli
Elena Baralis

Creativity injection into AI-powered Multimedia Storyboards

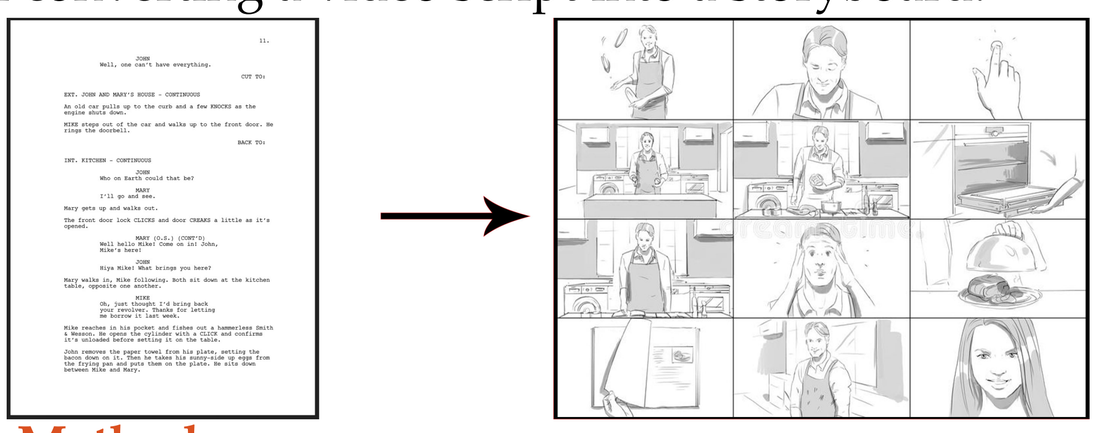
PhD Candidate: Bartolomeo Vacchetti

1. Introduction.

Recently a large variety of advanced applications have been trying to democratize complex activities such as photography, movie editing and filmmaking. While these applications have impressive performances they could be further improved by taking into account creative features that have been considered in the past but not fully exploited.

2. Objectives

The goal of this research is to study and analyze those features in order to address the difficult task of converting a video script into a storyboard.

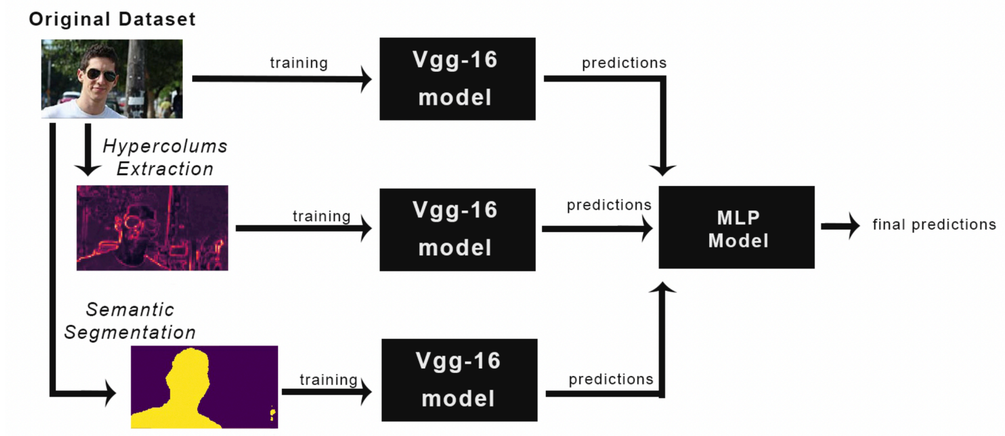


3. Method

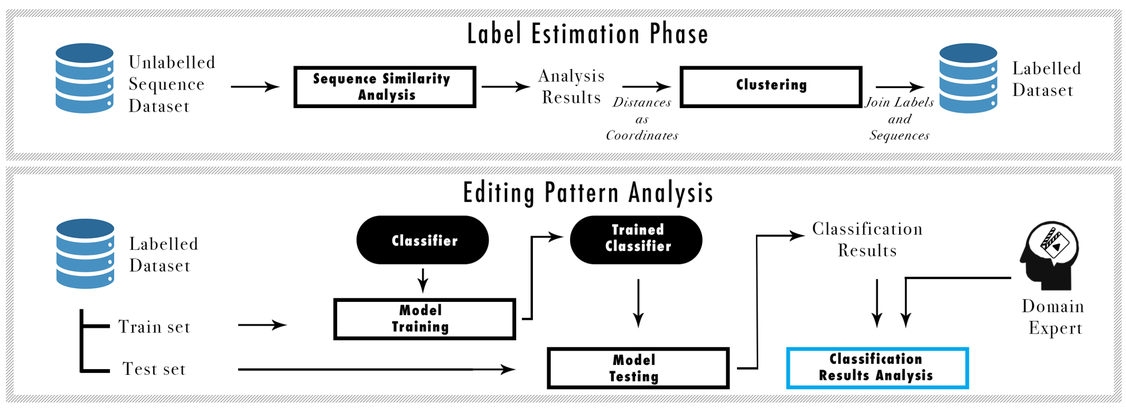
A storyboard is a graphical document that contains sketched frames. Each frame represents a shot to realize. Recently several text to image generators have been presented. However these models would not be fit to perform the task. On one hand they convert text into a single image and not into a sequence of images. Secondly if they were fed dialogues texts, they would represent what the characters are talking about rather than showing them talking.

In order to solve these issues some preliminary steps were necessary.

In [1] we have proposed a novel methodology to address the cinematographic shot classification with 8 classes. Specifically we have fine-tuned three VGG-16 and combined their predictions with the stacking learning technique.



Then in [2] we have analyzed the editing patterns in short movie sequences. In order to do so we have devised a novel approach that relies on the Levenshtein distance, the K-Means clustering algorithm and a multilayer perceptron.



In order to convert script into storyboards we will use two generative models. One will be trained to generate movie editing sequences, i.e. sequences of cinematographic shots, while the other one will be trained to generate sketches of the different shots.

4. Preliminary results

On the left we have the result for [1], while on the right the results from [2].

Labels	precision	recall	f1-score	support
Long Shot (LS)	89%	79%	84%	407
Medium Shot (MS)	70%	77%	73%	381
Full Figure (FF)	79%	76%	77%	324
American Shot (AS)	71%	69%	70%	282
Half Figure (HF)	69%	74%	72%	393
Half Torso (HT)	74%	80%	77%	503
Close Up (CU)	79%	79%	79%	519
Extreme Close Up (ECU)	90%	79%	84%	355
accuracy			77%	3164
macro avg	78%	77%	77%	3164
weighted avg	78%	77%	77%	3164

Number of Classes	Overall Accuracy	Macro Average	Weighted Average
4	93%	93%	93%
8	88%	88%	88%
16	81%	79%	81%
32	77%	72%	77%

5. References

1. Bartolomeo Vacchetti, Tania Cerquitelli (2022)"*Cinematographic Shot Classification with Deep Ensemble Learning*" in ELECTRONICS vol. 11, ISSN 2079-9292.
2. Vacchetti Bartolomeo, Tania Cerquitelli, (in print)"*MOVIELENS: Discovering and Characterizing Editing Patterns in the Analysis of Short Movie Sequences.*" Vacchetti B., Cerquitelli T. in CVEU 2022, Tel Aviv 23-27 October, Tel Aviv.