



Zürich, July 3, 2019

### MSc Project: Movie Genre Classification Using Convolutional Neural Networks

The task of movie genre classification is mostly done through textual descriptors or with tags. All this information is usually generated by users and can be therefore a tedious and expensive undertaking. Applying machine-learning-based vision classification methods could help in automating the analysis and classification of video content.

There is already a plethora of research dedicated to creating machine learning solutions for the classification of images, mostly for carrying out single-label classification. However, multi-labeling video-based approaches have been proven to be more challenging and not yet thoroughly researched. Furthermore, recent research has demonstrated that the use of convolutional neural networks (CNN) can indeed lead to better results. A convolutional neural network is a class of deep neural networks (see also Figure 1), representing regularized versions of multilayer perceptrons. They are especially useful for finding complex patterns in images, and recognizing objects. One of its advantages is that it can learn directly from the image data, using patterns and thereby eliminating the need for manual feature extraction.

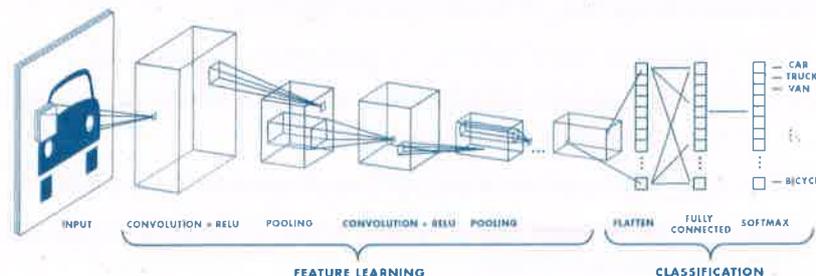


Figure 1. A convolutional neural network (source [3])

This project is about utilizing a deep neural architecture based on convolutional neural networks and perform a multi-label movie-trailer genre classification. The approach encapsulates

an ultra-deep convolutional neural network with residual connections, and it makes use of convolutional layers to extract temporal information from image-based features before performing the mapping of movie trailers to genres. During the project, the steps shown in Figure 2 are followed (more details on specific tasks later). The results will be compared with current state-of-the-art methods for movie classification that employ well-known image descriptors and other low-level handcrafted features.

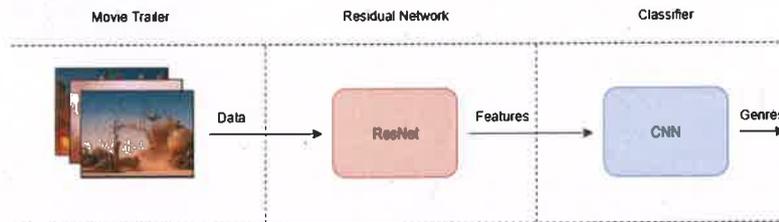


Figure 2. Project pipeline

The project utilizes (and complements) the work done on the database MMTF14, which already consists of more than 13'000 movie trailers. This includes the extraction of keyframes and audio signal spectrograms from the aforementioned database.

The work is structured into the following tasks:

- Studying the state-of-the-art for video and image classification with convolutional neural networks.
- Preparing the dataset, this includes downloading the video clips, extracting keyframes, audio and visual features, and determining/checking the movie genres. One of the most important steps is the preprocessing with the Keras object detection models to identify objects in the keyframes.
- After preparing the input for the neural network, the training phase starts for a single-label genre classification.
- The results are analyzed using a test dataset and are interpreted.
- Depending on the results, several improvements will be developed and implemented. This is going to be an iterative process, switching between testing new features and developing/improving them. This task includes improving
  - the extraction of visual features, i.e., adding new visual features.
  - the use of different pre-trained models besides Keras and possibly even combining different models in the pre-training step.
  - the classification task by moving from single labels to multiple labels.
  - the keyframe extraction by tuning the number of extracted frames and focusing on more meaningful frames.

An optional task is the development of explainable AI techniques for the convolutional neural network, i.e., integrating functionality with which the neural network points out which features of the video were most important for determining its genre. This could, for example, be in the form of highlighting certain keyframes or areas within frames that led to the decision.



### References

- [1] G. Simões, J. Wehrmann, R.C. Barros and D.C. Ruiz. Movie genre classification with Convolutional Neural Networks. 2016 International Joint Conference on Neural Networks (IJCNN), 2016
- [2] J. Wehrmann and R.C. Barros. Movie genre classification: A multi-label approach based on convolutions through time. Applied Soft Computing, 2017
- [3] MathWorks: Deep Learning - Convolutional Neural Network 3 things you need to know, <https://de.mathworks.com/discovery/convolutional-neural-network.html>, accessed 2 July 2019

**Supervisor:** Sven Helmer

Department of Informatics, University of Zurich

Prof. Dr. Michael Böhlen

A handwritten signature in blue ink, appearing to read 'M. Böhlen'.

