Effect of Style Transfer on Movie Genre Classification Based on Poster for INF003 - Spring 2020

Deniz Mert Tecimer
Department of Computer Engineering
Turkish - German University
e160503131@stud.tau.edu.tr
github.com/deniz997/MoviePosterClassification

Abstract

We will classify movies by genres, based on their posters. Beside that, it is also aimed in this project to examine the effect of style transfer on this multi-classification.

1 Introduction

Genre classification for movies, a type of multi-classification, is well used in the industry. Such companies like Netflix, Amazon etc. use this functionality to deliver their contents, to whom it may attract. Beyond static genres, they even produce dynamic(combined) genres. Netflix also generates dynamic posters for their contents, accordingly to consumers interests. Based on this feature, we can say that posters are very important for attracting consumers and also they inform the consumer about the content of the product.[1] Some posters might lead its consumer wrong, and this might cause time waste and disappointment. Dynamic posters of Netflix are mostly based on the objects, and do not represent the style of the genre, movie belongs. This project might lead a future work to see the effect of style transfer on poster based movie genre classifier and by comparing this effect with a survey, with some audience and maybe lead a future work of making this dynamic posters better.

The input to our algorithm is a database of movie poster images and their genres. We then use a neural network to predict movie genres.

2 Database and Features

The database "Movie Genre from its Poster" from Kaggle will be used in this project. It contains information about movies,such as IMDB id, IMDB link, title, IMDB score and the poster of the movie, 3 genres assigned per movie, retrieved from IMDB website. The dataset consists of 36898 256X256 resolution RGB posters. Distribution of examples is as 13550 train set, 1506 val set, 1672 test set. Data is pre-processed similarly as Rahul et al.(2020)[3]'s work. First movies with missing genres or image link to the posters will be removed. Then posters are transferred in to the resolution of 150X150. Twenty-eight different genres will be reduced to top 12 most popular genres.

At the end we have the pre-processed data that the algorithm requires. However, the data is not equally distributed, and this limits the expected accuracy.

Three examples from the database are as follows:

INF003: Neural Networks, Spring 2020, Turkish - German University, Istanbul. (LateX template borrowed from NIPS 2017.)

Genre Names	Number of Posters
Action	4326
Animation	1507
Adventure	3228
Comedy	5860
Family	1559
Drama	9438
Romance	3450
Crime	4124
Thriller	2859
Fantasy	1589
Horror	1988
Mystery	1927

Table 1: Number of posters per genre

Action | Adventure | Crime
Figure 1. Example 1

Action | Adventure | Comedy
Figure 2. Example 2

Action | Adventure | Comedy
Figure 3. Example 3

3 Multi-label Classification

The database consists of movies, that has multiple genres. These genres are independent from each other. Thus we need to run n-classifiers on the data to determine whether the movie belong to a poster. In the multi-class classification problem, it must be determined, which genre movie belongs. In contrast to that in this project, we aim to classify movies for multiple genres, so that the output of the proposed architecture will be n X c, where n is the number of examples and c is the number of genres. As a result of that, we have to come up with an algorithm that calculates the probability of every genre.

4 Deep Neural Network

The neural network architecture is inspired from Davide Iacobelli's Final Project [5] and consists of four convolutional layers, with the number of filters 32, 64, 128, 64 and with a filter size 3x3. After each two convolutional layers, max-pooling is applied, with a filter size 2x2, and followed by a dropout with the probability of 0.25 to avoid overfiting. After that we flatten the last output and fed to a fully-connected layer with 128 nodes. Then a dropout with a probability of 0.5 is applied, before feeding the last fully-connected layer with it. This layer has 12 nodes, each represents a genre. The activation function in each layer except the last layer is ReLu. In the output layer the activation function is Sigmoid. We train our model for 5 epochs with the batch size of 16 examples.

We do not use the Softmax activation function, while it divides the probability among the classes, which is helpfull, when the input belongs to only one class. However we do multi-label classification, not multi-class classification. As a second result of that we use Binary Crossentropy as a loss function. Binary Crossentropy measures how far away the predictions from the true value for each class and averages the class-wise error to calculate the final loss of a prediction. As an optimizer we use Adagrad. Adagrad is a gradient-based optimization algorithm, which performs smaller updates on parameters for frequently occurring features and larger updates on parameters for infrequently occurring features. To generalize, an Adagrad optimizer adapts the learning rate to the parameters based on occurrence of features. It is often used for big databases containing sparse data.

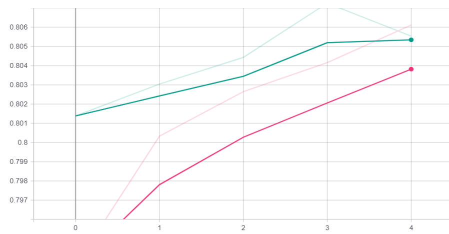

Figure 4. Epoch accuracy

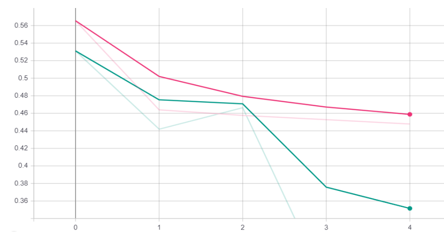

Figure 5. Epoch loss

The accuracy percentage of perfect matches is % 4.36 In the figures 4 and 5 the pink line represents the measurements on test data, while on the other hand the green line represents the measurements on validation data. The lines on the foreground represent the smoothed values and the pale lines on the background prepresent the actual values of the measurements. As shown in Figure 4, the accuracy seems to be constantly increasing. However, as it is shown in the background, the accuracy of the validation data after 3th epoch. As shown in Figure 5, the loss is now converging fast which will be investigated.

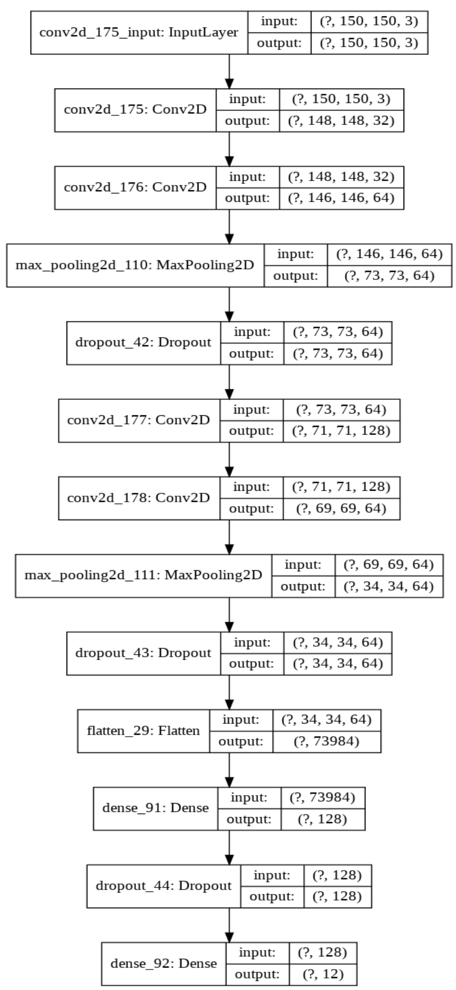

Figure 6. Structure of used deep neural network architecture

5 Future Works

- Other neural network architectures such as VGG-16, ResNet-50 and transfer learning will be tried.
- The amount of imbalanced data will be reduced via various data augmentation methods.
- Trial with a lower number of classes and with an additional Others class is planned.
- While it is hard to compute accuracy of the results of multi-label classification, more metrics will be tried.

References

[1] Amat, F., Chandrashekar, A., Jebara, T., & Basilico, J. (2018). Artwork personalization at Netflix. *Proceedings of the 12th ACM Conference on Recommender Systems.* doi: 10.1145/3240323.3241729

[2] Neha. (2018). Movie Genre from its Poster.Kaggle.

[3]Rahul Chokshi, Samuel Sung. (2020) Classification of Movie Posters to Movie Genres. *CS230 - Stanford University.*

[4] Luan, F., Paris, S., Shechtman, E., & Bala, K. (2017). Deep Photo Style Transfer. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* doi: 10.1109/cvpr.2017.740

[5] Iacobelli, D. (2019). Movie Genres Classification from their Poster Image using CNNs