# Movie Posters' Classification into Multiple Genres

Vaibhav Narawade Ramrao Adik Institute of Technology Nerul, Navi Mumbai vnarawade@gmail.com

Vishwaroop Ray Ramrao Adik Institute of Technology Nerul, Navi Mumbai vishwaroopray98@gmail.com Aneesh Potnis

Ramrao Adik Institute of Technology

Nerul, Navi Mumbai

apotnis07@gmail.com

Pratik Rathor

Ramrao Adik Institute of Technology

Nerul, Navi Mumbai

pratikrathor2298@gmail.com

Abstract—Our project intends to classify movies into the three most probable genres that they belong to, from a predefined set of 25 genres, based on only one image i.e the movie poster. We have made use of Convolutional Neural Networks (CNN) to realize this project as we believe it would be of help to extract the features and visual information from the image. Instead of a multi-class classification problem in which the input is classified into any one class, this project would be more correctly described as a multi-label classification problem as a movie belongs to more than one genre. In this project we see a comparative study of different architectures and tune them to yield the best result based on the metric of accuracy. We have applied various techniques such as data augmentation and L2 regularization to comparatively deduce the model that performs best from all the tended models.

Index Terms—Multi-label classification, movie genre classification, neural networks

#### I. INTRODUCTION

In this problem we have used Multi-label classification and Convolutional Neural Networks for Genres based classification of Movie Posters. Multi-label classification originated from the investigation of text categorisation problem, where each document may belong to many pre-defined topics simultaneously. Multi-label classification of textual data is a crucial problem. Examples range from news articles to emails. In multi-label classification, the training set consists of instances that are labelled from a group of labels and the label for a unseen instance needs to be rightly predicted by analyzing the training instances that the model has been trained on with known label sets. In machine learning, Convolutional Neural Networks (CNN or ConvNet) are complex feed forward neural networks. CNNs are used for image classification and recognition due to its high accuracy. It was proposed by scientist Yann LeCun within the late 90s, when he was inspired from the human beholding of recognizing things. The CNN follows a hierarchical model which works on building a network, sort of a funnel, and eventually gives out a fully- connected layer where all the neurons are connected to each other and the output is processed.

# A. Objective

The Objectives of this project is to categorize a movie according to the given poster into predetermined genres with

the use of CNN. This classification would able the user to make informed decisions by providing the genres of the movies based on the features present in the poster. The proper classification of movies will enable seamless process of categorizing the movies according to the previous preferences of the user. Humans are able to recognize certain elements in a movie to classify it into genres and to automate this process CNN can be used to compare the posters present in the dataset and accordingly assign the relevant genres to the movie.

## B. Motivation

The motivation behind this problem statement is to let the user identify if they would be interested in a particular movie by classifying the movie into multiple genres by identifying certain features from the poster of the said movie. This would be helpful to the user as the poster of the movie gets released up-to an year in advance, as opposed to the trailers which are released much later. The model would look for specific features in a movie poster that are predominantly present in a poster belonging to that specific genre, this would serve as a tool to notify the user about an upcoming movie that they might be interested in.

# II. LITERATURE SURVEY

- Binary classification and Multi-Class classification have been highly discussed topics but a lot of newer technology and research has seen the development of Multi-Label classification. This has also highlighted the problems which come along with this classification which were previously not known in binary or multi-class classification. Purvi Prajapati, Amit Thakkar, Amit Ganatra (2012) shine a light on these challenges in their work in 2012. Their paper proposed the two main methods to approach for the solution of multi label problems, namely Problem Transformation Method and Algorithm Adaption Method. [1]
- 2) There have been a few attempts of classifying movie into different genres based on it's poster, the earliest being the work done by Marina Ivasic-Kos, Miran Pobar

- and Ivo Ipsic (2014) which focused on extracting low-level features like dominant colours and colour layout. They also made the use of GIST image descriptor to extract the dominant spatial structures of an image characterized by it's boundary properties. Their classification techniques included Naïve Bayes classifier, a multi-label problem adaptation method called RAndom k-labELsets (RAKEL) and ML-kNN, a variation of the traditional k-Nearest Neighbour algorithm. This was applied on a dataset containing 6,739 movie posters and up-to 18 different genres each poster was classified into. [2]
- 3) This research served as groundwork for a later work done by Wei-Ta Chu and Hung-Jui Guo (2017) which applied deep learning for a modern approach to classification. Their research included a combination of fully connected neural network, convolutional neural network and a pre-trained YOLO v2 network for object detection. They also did a comparative study of the Ivasic-Kos, Pobar and Ipsic model and version versions of their own model by dividing the entire model into three parts and disabling or replacing one or more parts at a time including switching neural networks with a Support Vector Machine (SVM) classifier. Their dataset consisted of 8,191 movie posters belonging to 23 different genres.
- 4) Jonatas Wehrmann and Rodrigo C. Barros (2017) proposed a neural network architecture on the basis of convolutional neural networks for accomplishing the task of genre classification. But what they did differently than the previous works is that they made use of movie trailers (audio and visual) to train their model. They developed a an approach which they called Convolution-Through-Time or CTT, which is an architecture that makes use of trailer frames through time to extract features for the model. [4]
- 5) Speaking of convolutional neural networks, the study by Rahul Chokshi and Samuel Sung (2018) expands on the trend of using deep neural networks to solve image classification problems. Samuel Sung and Rahul Chokshi utilised customized versions of three new and promising deep learning architectures namely: ResNet-50, VGG-16, and DenseNet-169. Their work included a dataset of 36898 posters classifying into 7 genres. [5]
- 6) A recent study on predicting movie genres from movie posters was done by Gabriel Barney and Kris Kaya (2019). Their work acknowledges both the previous attempts at the topic and even cites Ivasic-Kos, Pobar and Ipsic's work as the groundwork they attempted to expand upon. Of all the studies mentioned here, Barney and Kaya used the largest dataset, viz the Full Movie-Lens Dataset from Kaggle, which consists of meta-data collected from TMDB and GroupLens containing entries for 45,466 movies. After pre-processing their dataset, it yielded a total of 35,000 movies belonging to 20 different genres. They too did a comparative study in their research by comparing different techniques namely,

- ML-kNN, One vs. Many multi-label strategy (OvR), a pre-trained version of Residual Network (ResNet34) for transfer learning purposes and a custom neural network architecture. [6]
- 7) Jeong A. Wi, Soojin Jang and Youngbin Kim (2020) proposed a method to extract the optimal information and characteristics from movie posters to aid the classification of movies into genres calling it the Gram layer in a convolutional neural network (CNN). It initially extracts style features by first applying Gram matrix onto the feature map of the movie poster. This is used as style weight, the existing feature map is used along with style information for performing inference. For their study, a total of 20,764 movie posters were collected and 12 genres were defined. [7]
- 8) An extensive survey on modern architectures of Deep Convolutional Neural Networks was done by Asifullah Khan, Anabia Sohail, Umme Zahoora I& Aqsa Saeed Qureshi (2020). While it has become common knowledge in the deep learning community that deeper neural networks result in higher accuracy, this has also given rise to problems like disappearing gradient problems which occur in which more layers cause the loss function gradient to approach zero, making it difficult to train the network. [8]
- 9) The latest approach to solving multi-label classification problems was brought up by the work of Emanuel Ben-Baruch, Tal Ridnik, Nadav Zamir, Asaf Noy, Itamar Friedman, Matan Protter and Lihi Zelnik-Manor (2020). In their study, they have discussed the most common problem associated with multi label classification, the high imbalance in the positive vs negative labels for a single input(picture). This problem, dubbed a the Positive-Negative imbalance, takes up most of the time in optimization process and if not taken care of, can result in under-emphasizing gradients from positive labels which in turn will result in lower accuracy. [9]

## A. Limitations of Existing System

- 1) The paper by Purvi Prajapati, Amit Thakkar, Amit Ganatra (2012) [1] presented a research of problem transformation vs algorithm adaption methods to solve multi label classification problems. One of the challenges stated by them in this field is the preprocessing of data that is required before the machine learning algorithms are applied. These techniques include but are not limited to pruning, feature selection and handling missing values to increase the performance of such problems.
- 2) According to Ivasic-Kos, Pobar and Ipsic's [2] comparative study, the Naïve Bayes classifier yielded the best result when considering the F1 score as the parameter of judgement, in which Naïve Bayes' F1 score was 0.38. Their application of GIST image descriptor along with classic machine learning algorithms can be considered

- outdated as per modern standards as neural network architectures have proved to outperform machine learning algorithms when it comes to large datasets especially the ones which involve image classification.
- 3) Chu and Guo's [3] study showcases the effectiveness of incorporating object information in neural networks using the YOLO v2 pre-trained object detection module which works quite well in detecting animals, which often appear in animation movies. Although their improvement over previous works has been promising, the neural network architecture they used (AlexNet) is an early one (2012) which is not in the trend nowadays.
- 4) Although the work done by Jonatas Wehrmann and Rodrigo C. Barros (2017) [4] showed very promising results especially when compared to low level features and improving accuracy on certain genres, their usage of movie trailers as their data set is not only computationally heavy but also redundant because by the time the trailers are released, the genres of the movies are already made very clear.
- 5) Sung and Chokshi's [5] study was an approach to find the optimum model for identifying genres and led to the comparative study of the THEN promising new models among which:DenseNet-169 was able to generate the best results. They can be still improved. They propose that the improvement can come through various techniques They forecasted that this improvement could come through various techniques, some being, the use of an object identifier algorithm such as YOLO, collection of better data, and application of other convolutional neural network (CNN) architectures to the problem.
- 6) Barney and Kaya (2019) [6] used the largest dataset and after preprocessing, it yielded a total of 35,000 movies belonging to 20 different genres. Possible improvement is using more balanced data. As the dataset that is used for the model is imbalanced, data augmentation can be used to increase the data the model can train from. A mixture of undersampling and oversampling can be applied respectively to the classes that have less data and the classes that have greater number of entries.

#### III. GOALS

- The project aims at classifying the movies into appropriate genre based only on the movie poster of the said movie by performing Multi-Label Classification and Deep Learning to provide an automated way of categorizing the movies into specific genres in order to accommodate the viewing preferences of users consuming the content.
- The Convolutional Neural Network has been identified as the Deep learning model for processing the images of movie posters.
- The dataset of 7242 movies has been used for this problem where these movies are spread across 25 different genres viz. Action, Adventure, Animation, Biography,

- Comedy, Crime, Documentary, Drama, Family, Fantasy, Horror, Music, Musical, Mystery, News, Reality-TV, Romance, Sci-Fi, Short, Sport, Thriller, War, Western.
- We propose to classify the movie poster into specific categories of genres from the 25 genres that are pre-defined. The movie would be categorized into different top three genres based on the probability of them belonging to the said genres. This classification would enable the user to make an informed decision regarding their interest in watching the movie when it releases. These classifications of movie posters into genres would be useful as the posters for various movies are released months before the trailers are released.
- The method incorporated for this project is deep learning. We have used the CNN which is widely used for processing images and recognizing patterns from these images. The first component of CNN called the convolution layer is for feature extraction, specifies the number of filters or feature detectors that make a feature map based on the feature that match with the filters and add a higher number in the feature map in such a place. The second component is the Max Pool where the size of the feature map is reduced while preserving the information about the features. The third component called Flattening transforms a two-dimensional matrix into a vector that can be fed into a fully connected layer. The fourth and final component before the output layer is the fully connected layer. In this layer as the name suggests every neuron is connected to every other neuron from the previous layer. As the genres can be classified into from a total of 25 the output layer has 25 nodes.

# A. Scope

- Movie Poster Classification is a multi-label classification problem with a large movie dataset the aim is to achieve multiple movie genre classification based only on movie posters images.
- For movie-goers, movie posters are among the very first impressions used to get an idea about the content of the movie and its what genres it could possibly be. Humans can get ideas based on things like color, objects, expressions on the faces of objects etc to quickly determine the genre (horror,comedy,animation,etc).
- We can assume that the poster possesses some characteristics which might be utilized in deep learning algorithms to predict its genres.

# IV. DESIGN OF THE SYSTEM

The system has been developed using TensorFlow. The Convolutional Neural Network has been identified as the Deep learning model for processing the images of movie posters. The dataset of 7242 movies has been used for this problem where these movies are spread across 25 different genres viz. Action, Adventure, Animation, Biography, Comedy, Crime,

Documentary, Drama, Family, Fantasy, Horror, Music, Musical, Mystery, News, Reality-TV, Romance, Sci-Fi, Short, Sport, Thriller, War, Western.

The system takes an input in the form of an image of a movie poster and this image is passed through a CNN model to make the appropriate inference of the genre of the movie. [Figure 1]

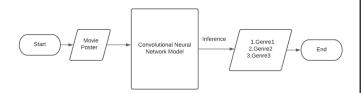


Fig. 1. Design of the system

#### V. IMPLEMENTATION DETAILS

#### A. Models

We have created variation of models to make the inference on the movie poster and the models are described below:

- a) Model 1: The baseline model is a 4 layer CNN model with batch normalization and dropout regularization after each convolutional layer. The baseline model that was created for performing inference on the movie posters.
- b) Model 2: VGG inspired model i.e block based model with each block containing 2 convolutional layers having 3x3 filters. We have 3 blocks with 32, 64 and 128 filters respectively. We have also applied Dropout regularisation after each block. The VGG16 inspired model used has the following layers. [Figure 2] [Figure 3]
- c) Model 3: Same as previous models but with data augmentation. In this iteration of the model we have used data augmentation which randomly performs horizontal and vertical flips with rotations upto 90 degrees.
- d) Model 4: Same model as Model 2 but with L2 regularization in 2 blocks of the model. (each block has 2 CNNs).
- *e) Model 5:* Same model as Model 2 but with L2 regularization AND data augmentation.
- *f) Model 6:* VGG16 architecture was used (using transfer learning).
  - g) Model 7: VGG16 with data augmentation.

Layer (type)	Output	Shape	Param #
conv2d_12 (Conv2D)	(None,	128, 128, 32)	896
conv2d_13 (Conv2D)	(None,	128, 128, 32)	9248
max_pooling2d_6 (MaxPooling2	(None,	64, 64, 32)	0
dropout_8 (Dropout)	(None,	64, 64, 32)	0
conv2d_14 (Conv2D)	(None,	64, 64, 64)	18496
conv2d_15 (Conv2D)	(None,	64, 64, 64)	36928
max_pooling2d_7 (MaxPooling2	(None,	32, 32, 64)	0
dropout_9 (Dropout)	(None,	32, 32, 64)	0
conv2d_16 (Conv2D)	(None,	32, 32, 128)	73856
conv2d_17 (Conv2D)	(None,	32, 32, 128)	147584
max_pooling2d_8 (MaxPooling2	(None,	16, 16, 128)	0
dropout_10 (Dropout)	(None,	16, 16, 128)	0
flatten_2 (Flatten)	(None,	32768)	0
dense_4 (Dense)	(None,	128)	4194432
dropout_11 (Dropout)	(None,	128)	0
dense_5 (Dense)	(None,	25)	3225
Total params: 4,484,665 Trainable params: 4,484,665 Non-trainable params: 0			

Fig. 2. VGG16 inspired Model

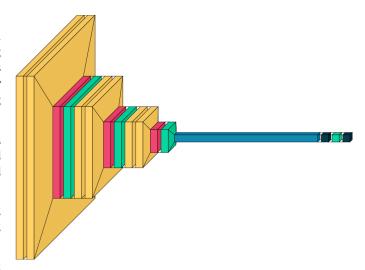


Fig. 3. Visual representation of the layers of Model 2

Yellow = Convolutional layer; Red = Max Pooling layer; Green = Dropout layer; Blue = Flatten layer; Dark blue= Dense layer

#### VI. RESULTS

#### A. Evaluation metric

Accuracy is the fraction of number of correctly predicted outcomes over the total number of predicted outcomes. Informal formula:

Accuracy = Correct Predictions/Total Predictions In classification problems, the formula of Accuracy: Accuracy = (TP + TN)/(TP + FP + TN + FN)

True Positives(TP): These are the outcomes in which the model predicts positive class and is correct. True Negatives(TN): These are the outcomes in which the model predicts negative class and is correct. False Positives(FP): These are the outcomes in which the model predicts positive class and is incorrect. False Negatives(FN): These are the outcomes in which the model predicts negative class and is incorrect. The models are compared on the basis of their accuracy and the cross entropy loss in the table[Table I] below.

TABLE I ACCURACY COMPARISON OF THE MODELS

Sr.	Model Details				
No.	Name	Accuracy	Loss		
1	Model 1	0.279	0.245		
2	Model 2	0.364	0.236		
3	Model 3	0.352	0.239		
4	Model 4	0.323	0.232		
5	Model 5	0.341	0.237		
6	Model 6	0.338	0.261		
7	Model 7	0.329	0.242		

The baseline model, Model 1[Figure 4], and the best performing model, Model 2[Figure 5], are compared through graphs for cross entropy loss and accuracy.

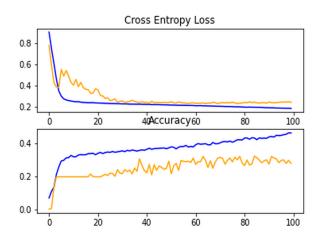


Fig. 4. Baseline model loss and accuracy

The inference from the model is given below where the most probable three genres are predicted. These genres are compared with actual listed genres of the movie.

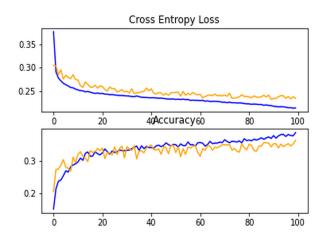


Fig. 5. VGG16 inspired model loss and accuracy



Fig. 6. Alien (1979)

Output: Drama, Thriller, Horror Ground Truth: Sci-Fi, Thriller, Horror



Fig. 7. To All the Boys I've Loved Before (2018)

Output: Comedy, Drama, Romance Ground Truth: Comedy, Drama, Romance ITM Web of Conferences 40, 03048 (2021)



Fig. 8. Green Book (2018)

Output: Drama, Comedy, Adventure Ground Truth: Drama, Comedy, Biography



Fig. 9. The Hitman's Bodyguard (2017)

Output: Comedy, Drama, Action Ground Truth: Comedy, Adventure, Action

Fig. 10. Results

# VII. CONCLUSION

With the results shown and on comparing the accuracies of different models, we can see that deeper and more complex neural network architectures tend to give significantly better results as compared to simpler architectures. After a certain point, the accuracies become stagnant and even different regularization techniques do not give improved results in terms of the metric of accuracy, although the loss is diminished by them. The regularization techniques also helped prevent overfitting of the model. The dataset too, as of now, is not very robust and is heavily partial towards the Drama, Comedy and Romance genres. Especially the genres, Comedy (2900 entries) and Drama (3619 entries) in a dataset of 7242. As is evident by the previous works, the imbalanced dataset problem has been common. Since our problem statement is a multilabel classification problem, classic methods of over-sampling or under-sampling of dataset, which works well for multi-class problems, will not provide us with desired results here. So in future, a well balanced dataset is sure to give much better

results on our model.

# REFERENCES

- [1] P. Prajapati, A. Thakkar, and A. Ganatra, "A survey and current research challenges in multi-label classification methods," International Journal of Soft Computing and Engineering (IJSCE), vol. 2, no. 1, pp. 248-252, 2012.
- [2] M. Ivasic-Kos, M. Pobar, and I. Ipsic, "Automatic movie posters classification into genres," in International Conference on ICT Innovations, pp. 319-328, Springer, 2014.
- W.-T. Chu and H.-J. Guo, "Movie genre classification based on poster images with deep neural networks," in Proceedings of the Workshop on Multimodal Understanding of Social, Affective and Subjective Attributes, pp. 39-45, 2017.
- [4] J. Wehrmann and R. C. Barros, "Movie genre classification: A multi-label approach based on convolutions through time," Applied Soft Computing, vol. 61, pp. 973-982, 2017.
- [5] S. Sung and R. Chokshi, "Classification of movie posters to movie
- [6] G. Barney and K. Kaya, "Predicting genre from movie posters," 2019.
- J. A. Wi, S. Jang, and Y. Kim, "Poster-based multiple movie genre classification using inter-channel features," IEEE Access, vol. 8, pp. 66615-66624, 2020.
- [8] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," Artificial Intelligence Review, vol. 53, no. 8, pp. 5455-5516, 2020.
- [9] E. Ben-Baruch, T. Ridnik, N. Zamir, A. Noy, I. Friedman, M. Protter, and L. Zelnik-Manor, "Asymmetric loss for multi-label classification," arXiv preprint arXiv:2009.14119, 2020.