Convolutional Neural Networks for Genre Classification of Movie Posters

[1]Shreyams Shetty, [2]Rinkal Suthar, [3]Janhvi Trivedi, [4]Nilesh Rathod

[1,2,3]Student, [4]Asst. Professor, Rajiv Gandhi Institute of Technology, Mumbai, India.

[1]shreyamsshetty@gmail.com, [2]sutharrinkal143@gmail.com, [3]janhvitrivedi13@gmail.com, [4]nilesh.rathod@mctrgit.ac.in

Abstract - **In this paper, we construct a convolutional neural network model to classify movie genre into its respected genre class. Since a movie can belong to multiple genre class, this is a multi-label image classification problem. Convolutional Neural Network is one of the most sought-after deep learning algorithms for solving machine learning problems especially with large image datasets and computer vision. To facilitate our study, we have compiled a large movie poster dataset from IMDB[1] using a web crawler. After applying some image preprocessing techniques on the movie poster dataset, we conclude by demonstrating the effectiveness of this model in classifying the movie dataset into classes such as action, drama, comedy and 11 other prominent movie genres.**

***Keywords* —Convolutional neural networks, Deep Learning, Image Classification, Movie genre classification, Multi-label Classification.**

I. INTRODUCTION

As of late, with the expansion in media content and internet streaming services giving movies in a hurry, there is a requirement for a framework which could group these movies into explicit sorts to diminish the time spent by clients on choosing a substance to watch. Client survey concludes that clients will in general lose enthusiasm within 60 to 90 seconds to peruse on the web for movies on the off chance that they don't locate the correct fit [1]. Movie poster plays an important role in the choice of the movie as it sets the early introduction furnishing the client with the essence of the movie initially.

In film theory, genre is a primary method of classifying movies into different categories. Various presets help in classifying movies into respected genres which helps the consumers select a movie according to their taste and mood. Mood can be defined as an emotion which is carried throughout the movie. Humans can get an idea about the genre of the movie based on low level features such as color, objects, actions and expressions of the actors. If humans are able to predict movie genre only giving a glance at its poster, then we can think what characteristics and features the poster possess which could be learned on by the machine learning algorithms to predict its genre.

Contributions of our work is summarized below:

To facilitate our study, we create a large 40,000 movie posters dataset from IMDB. This dataset will contain movie posters from Hollywood movie and metadata associated with the movie such as movie title, IMDB score and movie genre.

We construct a convolutional neural network model to classify these movie posters into genres. Since, movies can belong to multiple genre, this is a multi-label image classification problem.

Rest of the paper is organized as follows: Related work is reviewed in section II. Section III, describes the proposed methodology. In section IV, we evaluate the model's performance. At last conclusion and future work in section V.

II. RELATED WORK

There have been quite a few works on classifying movies into genres. Sanjay et al. [2] designed a neural network which categorized movie clips into genres using low level audio-visual features. The neural network performed well when the neural network was trained on audio video features together but failed when trained separately. Simoes et al. [3] proposed a deep learning strategy called CNN-MoTion which outperformed all state-of-the-art approaches of classifying movie clips into genres. They performed a comparative study between their novel method and other state-of-the-art feature extraction techniques such as Gist, CENTRIST, w-CENTRIST, and low-level feature extraction.

The aforementioned studies have been done on movie trailers and clips. Not many studies have been carried on

[1] www.imdb.com

 DOI : 10.35291/2454-9150.2020.0163

movie posters. This may be due to the fact that movie posters provide limited information and features to be studied. Marina et al. [4] performed a multi-label movie genre classification based on low-level features. The classifier was tested on a small dataset of 1500 movie posters belonging to 6 genres, i.e. action, animation, comedy, drama, horror, and war. The features used in the classification were low-level features based on color and edge combined with the number of detected faces on posters. Tianmei Guo et al. [5] proposed a low computational cost convolutional neural network on image classification. They analyzed various learning methods and optimization algorithms to find the optimal parameters for image classification. To conclude they verified that shallow network has a relatively good recognition effect. In our work, we propose to implement convolutional neural network on a large movie poster dataset which will perform well in terms of accuracy.

III. PROPOSED METHODOLOGY

Our problem definition requires us to construct a convolutional neural network to classify movie poster images into different genre classes. Since, a movie can belong to multiple genres classes, we are working with a multi-label image classification problem.

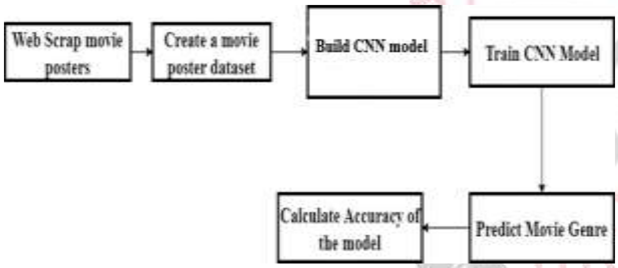


Fig 1. Overview of Proposed Methodology

A. Web Scraping

To facilitate our study, we start by preparing the movie poster dataset. Using the IMDB link of different posters, we web scrap the images along with the metadata and save them in our dataset. Since movie pages on IMDB site has no different structure, through Web scraping we can without much of a stretch get the publication connection of each film basically going on its IMDB page and taking the substance of the src HTML label relating to the notice. When we have all poster links, we add them to our dataset.

B. Collecting Movie Posters

When the Web scraping step is finished, we have additional posters links in our dataset. So now we can download movie pictures utilizing those links. Before doing that, we apply a straightforward advance of information cleaning to the dataset, comprising in dropping all sections without a characterized genre. So, we start downloading all posters from the correspondent links and we spare every one of them utilizing as name in the filesystem the IMDB id of the related movie. Along these lines we keep up the connection among movie and their poster images. It's important to note that some posters could be corrupted during download hence, we check for corrupted images and drop it from our dataset.

C. Data Visualization and Manipulation

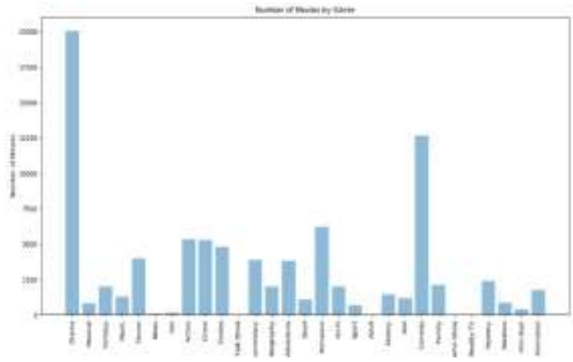


Fig 2. Data Visualization

Instead of directly moving onto creating machine learning model, it is important to have a look at our data and manipulate it to increase our model's performance. If the dataset is heavily imbalanced with respect to genres e.g. the dataset contains many occurrences of genre like "drama" while low number of occurrences for other genres. Through random sampling approach, we can add instances for genres with lesser occurrences to reduce the imbalance. This balancing of genres helps our model to avoid being bias towards a particular genre. Also, all the images are reshaped to match the input size of our model.

D. Model Construction

Convolutional neural network consists of three major layers namely convolutional layer, pooling layer and fully connected layer. Fig. 3 shows the architecture of LeNet-5 introduced by Yann LeCun et al. [6]. Below mentioned are the main components of convolutional neural networks.

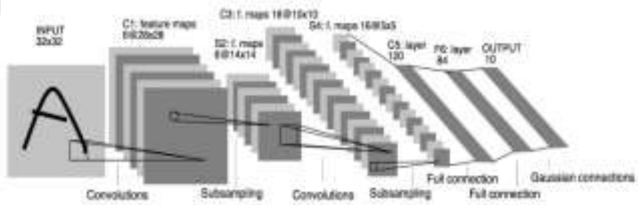


Fig 3. The architecture of LeNET-5 [6]

1) Convolutional Layer

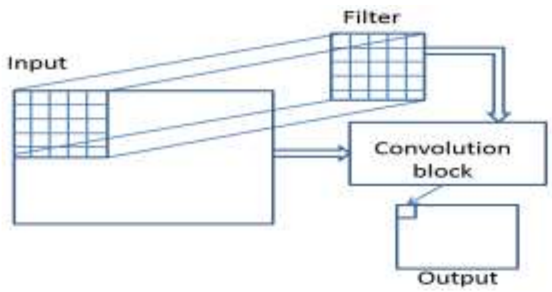


Fig 4. Convolutional Layer [7]

Convolution is the first and the core layer to extract features

 DOI : 10.35291/2454-9150.2020.0163

ISSN : 2454-9150 Vol-05, Issue-12, Mar 2020

from an input image. Convolution preserves the relationship between pixels by learning using small squares of feature data. The image matrix is convolved with a filter matrix to get a feature map as an output. Convolution of image matrix with different filters can perform operations like edge detection, sharpen and blur. In case the filter does not fit the input image completely, padding is performed. Pad the input image with zeros.

2) Pooling Layer

The pooling layer is responsible for dimensionality reduction of the convolved feature. This helps in decreasing the computational power required to process the image but not compromising with the effectiveness of the model as all the important features are retained. Max pooling, average pooling and sum pooling are different types of pooling.

3) Fully Connected Layer

The output matrix from the pooling layer is flattened and feed to the fully connected layer with back propagation applied to every iteration of training. Over a series of epochs, the model is able to distinguish between various high level and low-level features and finally classify them using an activation function. Each output neuron signifies a classification label.

4) Activation Function

An activation function is a capacity that is included into an artificial neural system so as to enable the model to learn complex features in the information. When contrasting with a neuron-based model that is in our brain, the activation function is toward the end choosing what is to be fed to the following neuron. That is actually what an activation function does in an ANN too. It takes in the yield signal from the past cell and changes it into some structure that can be taken as input to the following cell. The activation function used in our CNN model are ReLU and Sigmoid.

i) ReLU Activation Function

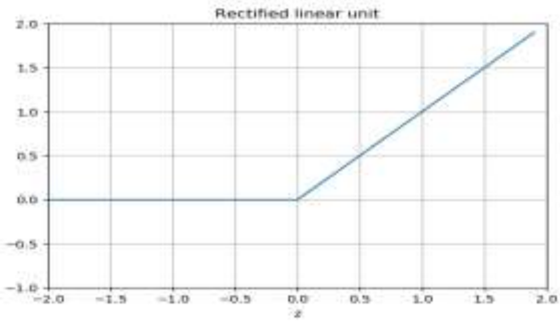


Fig 5. ReLU Activation Function

The rectified linear unit (ReLU) activation function has been the most widely used activation function for deep learning applications with state-of-the-art results [8]. The ReLU activation function performs a threshold operation to each input element where values less than zero are set to zero thus the ReLU is given by:

$$f(x) = max(0, x) = \begin{cases} x_i, if\ x_i \geq 0 \\ 0, if\ x_i \leq 0 \end{cases}$$

ii) *Sigmoid Activation Function*

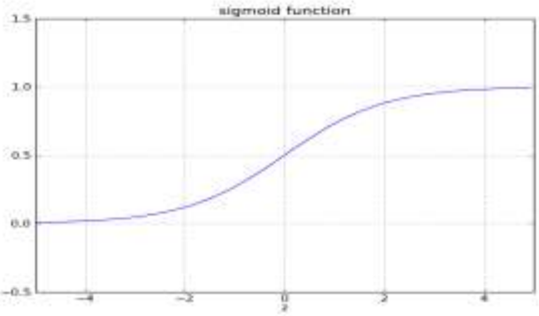


Fig 6. ReLU Activation Function

The Sigmoid is a non-linear AF used mostly in feedforward neural networks. It is a bounded differentiable real function, defined for real input values, with positive derivatives everywhere and some degree of smoothness [8]. Sigmoid can be mathematically represented as:

$$f(x) = \left(\frac{1}{1 + e^{-x}}\right)$$

5) Model Summary

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 182, 268, 32)	896
conv2d_2 (Conv2D)	(None, 180, 266, 32)	9248
max_pooling2d_1 (MaxPooling2	(None, 90, 133, 32)	0
dropout_1 (Dropout)	(None, 90, 133, 32)	0
conv2d_3 (Conv2D)	(None, 90, 133, 64)	18496
conv2d_4 (Conv2D)	(None, 88, 131, 64)	36928
max_pooling2d_2 (MaxPooling2	(None, 44, 65, 64)	0
dropout_2 (Dropout)	(None, 44, 65, 64)	0
conv2d_5 (Conv2D)	(None, 44, 65, 128)	73856
conv2d_6 (Conv2D)	(None, 42, 63, 128)	147584
max_pooling2d_3 (MaxPooling2	(None, 21, 31, 128)	0
dropout_3 (Dropout)	(None, 21, 31, 128)	0
flatten_1 (Flatten)	(None, 83328)	0
dense_1 (Dense)	(None, 128)	10666112
dropout_4 (Dropout)	(None, 128)	0
dense_2 (Dense)	(None, 14)	1806

Fig 7. Model Summary

IV. PERFORMANCE OF MOVIE GENRE CLASSIFIER

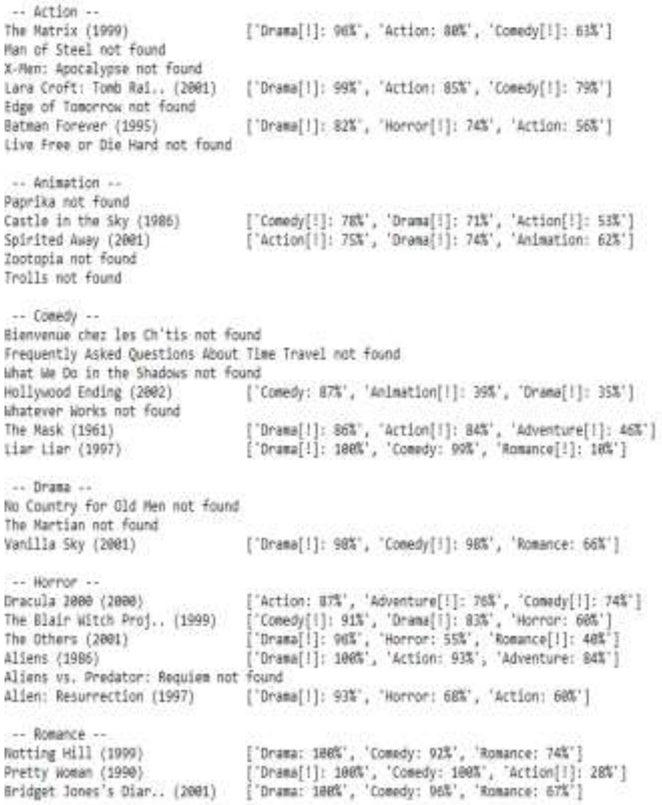
On testing the model on a dataset of over 10,000 movie posters, the test accuracy received was 45%. We were impressed by the performance of our model as the older multi-label classification models were not as accurate. Also, we used the built-in accuracy function of keras to get this

 DOI : 10.35291/2454-9150.2020.0163

number, hence we are not sure whether this is the best method for the same. Our model was most of the times successful in predicting at least one genre correctly but found it difficult in predicting all of the genre classes (3 in most cases) correctly.

Table 1. Model Performance

Name of Movie	Actual Genre	Predicted Genre
Pearl Harbor	Action, Drama, History	Drama, Action, Romance
The Truman Show	Comedy, Drama, Sci-Fi	Drama, Comedy, Romance
Notting Hill	Drama, Comedy, Romance	Drama, Comedy, Romance
Bridget Jones's Diary	Drama, Comedy, Romance	Drama, Comedy, Romance



```
 -- Action --
The Matrix (1999)              ['Drama[!]: 96%', 'Action: 80%', 'Comedy[!]: 63%']
Man of Steel not found
X-Men: Apocalypse not found
Lara Croft: Tomb Rai.. (2001)  ['Drama[!]: 99%', 'Action: 85%', 'Comedy[!]: 79%']
Edge of Tomorrow not found
Batman Forever (1995)          ['Drama[!]: 82%', 'Horror[!]: 74%', 'Action: 56%']
Live Free or Die Hard not found

 -- Animation --
Paprika not found
Castle in the Sky (1986)       ['Comedy[!]: 78%', 'Drama[!]: 71%', 'Action[!]: 53%']
Spirited Away (2001)           ['Action[!]: 75%', 'Drama[!]: 74%', 'Animation: 62%']
Zootopia not found
Trolls not found

 -- Comedy --
Bienvenue chez les Ch'tis not found
Frequently Asked Questions About Time Travel not found
What We Do in the Shadows not found
Hollywood Ending (2002)        ['Comedy: 87%', 'Animation[!]: 39%', 'Drama[!]: 35%']
Whatever Works not found
The Mask (1961)                ['Drama[!]: 86%', 'Action[!]: 84%', 'Adventure[!]: 46%']
Liar Liar (1997)               ['Drama[!]: 100%', 'Comedy: 99%', 'Romance[!]: 10%']

 -- Drama --
No Country for Old Men not found
The Martian not found
Vanilla Sky (2001)             ['Drama[!]: 98%', 'Comedy[!]: 98%', 'Romance: 66%']

 -- Horror --
Dracula 2000 (2000)            ['Action: 87%', 'Adventure[!]: 76%', 'Comedy[!]: 74%']
The Blair Witch Proj.. (1999)  ['Comedy[!]: 91%', 'Drama[!]: 83%', 'Horror: 60%']
The Others (2001)              ['Drama[!]: 96%', 'Horror: 55%', 'Romance[!]: 40%']
Aliens (1986)                  ['Drama[!]: 100%', 'Action: 93%', 'Adventure: 84%']
Aliens vs. Predator: Requiem not found
Alien: Resurrection (1997)     ['Drama[!]: 93%', 'Horror: 68%', 'Action: 60%']

 -- Romance --
Notting Hill (1999)            ['Drama: 100%', 'Comedy: 92%', 'Romance: 74%']
Pretty Woman (1990)            ['Drama[!]: 100%', 'Comedy: 100%', 'Action[!]: 28%']
Bridget Jones's Diar.. (2001)  ['Drama: 100%', 'Comedy: 96%', 'Romance: 67%']
```

Fig 8. Example of classified movies

The model gives an output of 3 genres which it predicts to have the highest probability of being the genre of the movie. The "[!]" indicates that the model predicted the wrong genre of the movie. While the percentage is the probability the model predicted for the genre being of the movie.

V. CONCLUSION AND FUTURE WORK

From above analysis it tends to be reasoned that Convolutional Neural System has a demonstrated reputation of outflanking best in class AI calculations when enormous picture dataset is to be arranged. The current frameworks center to arrange movies utilizing video cuts as opposed to movie poster as it is expected that movie posters have less data from which the CNN model can gain from. Be that as it may, if the CNN model is prepared upon an enormous movie poster dataset, it can gain from it and can precisely group movies into its type. In the present web world, the utilization of mixed media has expanded radically as clients have begun utilizing Advanced stages to peruse movies and consequently a framework is required to assist web based gushing stages with classifying these movies into their regarded kinds so the clients can straightforwardly peruse movie as indicated by their essence of movies. The future extent of the venture is huge as this convolutional neural system model can be fused in different enormous scale movie recommender frameworks for prescribing movies to its clients. Each enormous scale online movie gushing stage use movie recommender frameworks to upgrade their client's understanding by suggesting them movies utilizing NLP on movie outlines and titles. Rather than utilizing content-based arrangement, the utilization of picture order on movie posters can end up being progressively productive as clients are bound to pass judgment on a movie by its spread for example poster as opposed to perusing its summary and concluding whether to watch the movie or not. We are still working on increasing the accuracy of our model and in process of implementing advanced methods.

ACKNOWLEDGMENT

We would like to thank MCT's Rajiv Gandhi Institute of Technology for providing research facilities and to Prof. Nilesh Rathod (Information Technology) for his support and guidance.

REFERENCES

[1] Carlos A. Gomez-Uribe and Neil Hunt. 2015. The Netflix recommender system: Algorithms, business value, and innovation. ACM Trans. Manage. Inf. Syst. 6, 4, Article 13 (December 2015), 19 pages. DOI: http://dx.doi.org/10.1145/2843948.

[2] Jain, Sanjay & Jadon, R.s. (2009). Movies genres classifier using neural network. 575-580. 10.1109/ISCIS.2009.5291884.

[3] Simões, Gabriel & Wehrmann, Jônatas & Barros, Rodrigo & Ruiz, Duncan. (2016). Movie genre classification with Convolutional Neural Networks. 259-266. 10.1109/IJCNN.2016.7727207.

[4] Ivašić-Kos, Marina & Pobar, Miran & Mikec, Luka. (2014). Movie posters classification into genres based on low-level features. 1198-1203. 10.1109/MIPRO.2014.6859750.

[5] Guo, Tianmei & Dong, Jiwen & Li, Henjian & Gao, Yunxing. (2017). Simple convolutional neural network

 DOI : 10.35291/2454-9150.2020.0163

ISSN : 2454-9150 Vol-05, Issue-12, Mar 2020

on image classification. 721-724. 10.1109/ICBDA.2017.8078730.

[6] Lecun, Yann & Bottou, Leon & Bengio, Y. & Haffner, Patrick. (1998). Gradient-Based Learning Applied to Document Recognition. Proceedings of the IEEE. 86. 2278 - 2324. 10.1109/5.726791.

[7] Panwar, Madhuri & Padmini, J. & Venkatasubrahmanian, & Acharyya, Amit & Biswas, Dwaipayan. (2017). Modified distributed arithmetic based low complexity CNN architecture design methodology. 1-4. 10.1109/ECCTD.2017.8093254.

[8] Chigozie Nwankpa and Winifred Ijomah and Anthony Gachagan and Stephen Marshall (2018), Activation Functions: Comparison of trends in Practice and Research for Deep Learning, arXiv:1811.03378.

[9] Wei-TaChu and Hung-JuiGuo (2017). Movie Genre Classification based on Poster Images with Deep Neural Networks. 2017ACM. 978-1-4503-5509-4/17/10...$15.00 DOI:10.1145/3132515.3132516.

[10] Albawi, Saad & Abed Mohammed, Tareq & ALZAWI, Saad. (2017). Understanding of a Convolutional Neural Network. 10.1109/ICEngTechnol.2017.8308186.

[11] Muhammad Zeeshan Khan, Muhammad A. Hassan, Saleet Ul Hassan, Muhammad Usman Ghanni Khan “Semantic Analysis of News Based on the Deep Convolution Neural Network” 2018 14th International Conference on Emerging Technologies (ICET).

[12] Dharti M. Bhoraniya, Prof. Tushar V. Ratanpara “A Survey on Video Genre Classification Techniques” 2017 International Conference on Intelligent Computing and Control (I2C2).

 DOI : 10.35291/2454-9150.2020.0163