DYAH E H 'MOVIE POSTER GENRE' PROSIDING_0001

Movie Poster Genre Classification with Convolutional Neural Network

Marcellus[1], Dyah Erny Herwindiati[2], Janson Hendryli[3]

[1,2,3]Faculty of Information Technology, Universitas Tarumanagara, Jakarta, Indonesia

Email: [1]marcellus.535170018@stu.untar.ac.id, [2]dyahh@fti.untar.ac.id, [3]jansonh@fti.untar.ac.id

***Abstract*— Not infrequently we do not have a clear plan of events to do, and chose to do an activity spontaneonsly. One of the activities that is usually done spontaneously is watching a movie in the cinema, and in this case, information about the movie we want to watch is most likely not known to the prospective viewer. Therefore, this paper is expected to help build a software to classify film genres based on image data input in the form of movie posters. By utilizing the intended software, potential viewers can make genre predictions to help choose the films they want to watch. The genre classification process is carried out using the Convolutional Neural Network. MobileNetv2 architecture was chosen due to the minimal computing cost of this architecture, and given computing power of hardware in form of a smart phone is relatively smaller than the computing power of hardware in the form of a computer. The final result produced by this model is the classification is in the form of generic film genres, which are divided into: romance, action, horror, fantasy and comedy.**

Keywords—Multi-label Classification, Movie PosterMovie Genre, Convolutional Neural Network, MobileNetv2

I. INTRODUCTION

Film genre is a form of category based on the similarity of narrative elements, aesthetic approach, and emotional response to the film as well as certain classifications that have similarities/similarities in form, setting, theme, atmosphere and others. Quoting from various sources of literary criticism theory, film genres are usually described by; conventions, iconography, setting, narrative, characters and actors. [1] It is apparent that using the visual cues, in this case, a movie poster could be a powerful tool for genre categorization. [2] Classification could be done with many traditional machine learning methods such as *Naive Bayes, SVM, KNN, decision trees*, and much more. The problem is, that a traditional machine learning method's computing time are comparatively higher than the computing time of one using a Convolutional Neural Network. [3]

The data used is obtained from The Movie Database, in the form of a movie poster. For training purposes, 5000 movie posters are used, with a 25% validation split. For the purpose of re-validation, another 2000 movie posters are used to evaluate the final model. With an additional 450 movie posters to evaluate a specific case, and another 750 movie posters to evaluate a similar but different specific case.

In this paper, movie genres are classified into five generic different genres, namely action, comedy, fantasy, horror and romance. This is due to the countless number of genres, and in order to simplify things for further study it is limited to only those 5 genres mentioned.

The model yielded is expected to be able to be integrated practically with mobile software. Which the mobile software could, later on, help predicting movie genres by utilizing the mobile camera and processing power.

II. Methods

A. Convolutional Neural Network

Convolutional neural network is a deep learning algorithm that can accept input data in the form of image data, and assign weights and biases to various aspects and objects in the image data. The steps in the pre-processing stage required in the convolutional neural network are much less when compared to other classification algorithms, so the pre-processing stage will be faster. Convolutional neural network also has the advantage over the traditional machine learning methods. Whereas on the traditional machine learning methods, filters must be created separately. Convolutional neural network has the ability to study filters and characteristics of image data. [4]

A Convolutional Neural Network could have been implemented with countless different architectures. But due to the goal of this paper is to support ease for prospective movie viewers, which very likely they won't carry around a heavy hardware, the model used for this specific case must be a light-weighted model. The architecture chosen for this case is MobileNetv2, due to the architecture's low computing time. [5] However, by having a relatively minimal computational operations, which makes the computational speed for both training and prediction faster, the model also has a weakness where the model accuracy is relatively lower.[6]

B. Learning Process and Model Evaluation

The activation function used in this particular case is logistic sigmoid, due to the possibility of a movie poster having more than 1 genres. The learning process is done by utilizing the 5000 dataset of movie poster, done with the configurations and parameters mentioned on III.A. The

978-1-6654-3414-0/21/$31.00 ©2021 IEEE
DOI 10.1109/BigMM52142.2021.00020

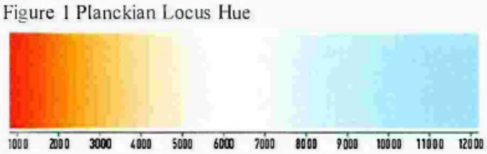
2021 IEEE Seventh International Conference on Multimedia Big Data (BigMM) | 978-1-6654-3414-0/21/$31.00 ©2021 IEEE | DOI: 10.1109/BigMM52142.2021.00020

trained model would hopefully be able to help implementing the creation of movie genre classification software.

C. Pre-processing

The lights used in the cinema lobby are linear fluorescent bulbs, which have a color temperature of 2700K or RGB (255,169,87), while neutral white light has a color temperature of 6500K RGB (255,255,255). Therefore, the image data for prediction, the image data captured by the camera in the cinema lobby will be multiplied by a constant by the value (255/255,255/169,255/87) for each color channel in order to achieve a neutral white color temperature.[7][8]

Figure 1 Planckian Locus Hue



III. Result and Discussion

A. Hyperparameter tuning experiment results

1) Threshold

The method used for choosing threshold here is by looking at the highest accuracy with the same exact configurations used for training, on a given epoch. The result is observable on table **n** below.

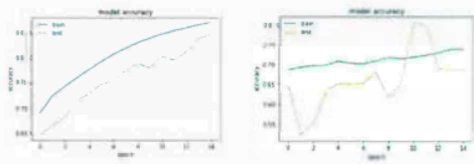
Table 1 Validation result

Threshold	Accuracy
0.5	69.26%
0.6	73.41%
0.7	76.03%
0.8	78.61%
0.9	76.42%

2) Optimizer

The Adam optimizer, as seen on the right side of figure **2**, seems to have problem with having a steady validation accuracy increase per epoch, the accuracy fluctuates a lot, therefore requires more epoch to meet the optimal accuracy. What causes these fluctuations on the Adam optimizer is that the learning rate is relatively too high compared to the validation set, which leads to the big leap when doing the stochastic gradient descent. This can be fixed by having more validation data or readjusting the learning rate but having a slower learning pace. The former option is impossible given the resource we have, so we chose to go with the latter option, while still considering a fast-paced learning.

Figure 2 Model's accuracy



3.) Learning rate

The method used to choose the learning rate is similar to the method used when choosing threshold value. The difference is that we try to get the largest learning rate, while still making sure that the model didn't learn too fast as it is the case with the Adam optimizer. Choosing a low learning rate would mean that the model will learn slower, needing more computing time.

4.) Post result analysis

After the model was trained, we did a separate test for validation. The validation process is done by using 2000 data, with 400 data each genre. The simplified confusion matrix is observable on table 1 down below, with the metrics on table 2.

Table 1 Validation result

Genre	TP	TP	TN	FN
Action	250	65	1300	384
Fantasy	192	85	1348	374
Comedy	567	179	928	325
Romance	487	426	1008	78
Horror	162	19	1432	386
Total	1658	774	6016	1547

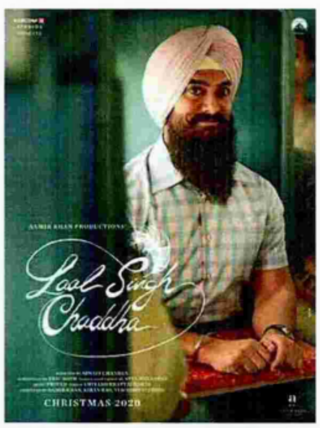
Table 2 Validation result metrics

Genre	Precision	Recall	f1-score	Accuracy
Action	0.7936	0.3943	0.5268	0.7753
Fantasy	0.6931	0.3392	0.4554	0.7703
Comedy	0.7570	0.6356	0.6910	0.7478
Romance	0.5334	0.8619	0.6589	0.7478
Horror	0.8950	0.2967	0.4456	0.7973
Average	0.7344	0.5055	0.5555	0.7677

From the table above, lies a more detailed metrics such as precision, recall and f1-score which could be used for more further detailed analysis. From results shown in the table above, is clear that the precision of the romance and fantasy genres is relatively lower than other genres. This is due to comedy genre posters tend to be predicted as a romance genre. While the other four genres are also sometimes predicted as fantasy genres, which causes the low precision of the fantasy genre to be relatively lower.

Movie posters with multiple genres are also one of the factors which might lead to the unbalanced recall value. For example, the image shown below has a comedy-romance genre, but due to the movie poster lacking comedy visual feature such as relatively bright color tone/palette, and has more of romance-dominant feature. This causes the model to predict the movie poster as just romance, not comedy-romance.

Figure 4 Comedy-Romance Movie Poster



But this doesn't apply for just this specific movie poster. In fact, most of the comedy-romance movie posters have this similar property, where the romance like features are more prominent than the comedy features. Here are some more in-depth analysis for the romance-comedy genre case:

Table 3 In-depth validation result

Genre	TP	TP	TN	FN
Action	0	5	445	0
Fantasy	0	13	437	0
Comedy	306	0	0	144
Romance	390	0	0	60
Horror	0	0	450	0

From the data provided on table above, only 68% of the romance-comedy movie posters were successfully predicted to have a comedy genre, while 86% of the romance-comedy poster were successfully predicted to have a romance genre. This problem also applies to other genres with a low recall value, having a more prominent visual cue of a certain genre than the other.

Some other examples from the multi-labelled fantasy genre, having 2 genres, with 150 sample each:

Table 4 Action-fantasy genre

Genre	TP	TP	TN	FN
Action	59	3	0	88
Fantasy	30	0	3	117
Comedy	0	24	126	144
Romance	0	27	123	60
Horror	0	1	149	0

Table 5 Comedy-fantasy genre

Genre	TP	TP	TN	FN
Action	0	15	135	0
Fantasy	25	0	3	121
Comedy	90	0	4	56
Romance	0	73	77	0
Horror	0	3	147	0

Table 6 Romance-fantasy genre

Genre	TP	TP	TN	FN
Action	0	7	143	0
Fantasy	15	0	2	133
Comedy	0	37	113	0

Romance	101	0	2	47
Horror	0	2	148	0

Table 7 Horror-fantasy genre

Genre	TP	TP	TN	FN
Action	0	19	131	0
Fantasy	21	0	1	128
Comedy	0	26	124	0
Romance	0	52	98	0
Horror	58	0	1	91

From the 4 multi-labelled fantasy tables above, we can observe the number of correct genres predicted, to the total number of samples. We can also get the precision and recall value.

Table 8 Specific case validation metrics

Genre	Predicted Fantasy	Predicted Second-genre	Fantasy Precision	Fantasy Recall
Action	20%	39.33%	100%	60.54%
Comedy	16.67%	60%	100%	17%
Romance	10%	67.33%	100%	10.13%
Horror	14%	38.66%	100%	14.09%

Another contributing factor might be the imbalance of multi-labelled class data dissemination.

V. Conclusion

The model was evaluated for validation with 2000 movie posters, which yields a 76.79% accuracy. It seems that this number is good, but an in-depth analysis found that this seems to not be the case. The recall and precision values that the model yielded is far from what is expected. There are a lot of factors that causes this problem, some of which are the multi-labelled genre movie posters, which has more prominent feature of a genre than the other, or even others. Multi-labelled data imbalance also causes this problem. The 5000 movie posters consist of 1000 posters for each genre, but each of the poster might have more than 1 genre, which causes an imbalance for multi-labelled genres.

The constant mentioned on II.C, the constant which is used to neutralize warm light is not optimal due to the possibility of every cinema having a different light warmth. Therefore, it is proposed that the constant could be changed accordingly. Another factor regarding lighting, is that some cinemas have lightbulbs behind or even beside the movie poster, which causes the color of the poster to differ from the original digital version. Adjusting this is quite a difficult task as there are no known standard lighting rules for lighting a movie poster on the cinema. There are also no known data of real-world movie poster on cinema lobbies dataset to date, making this quite a difficult task to do.

The result yielded could still be improved by changing the MobileNetV2 model to other alternatives such as Inception V2, or even VGG which later on could be integrated with a mobile software that is aimed to help predict movie genres through movie posters

REFERENCES

[1] Grant, Barry Keith (2007). Film Genre: From Iconography to Ideology. Short cuts. 33 (reprint ed.). London: Wallflower Press. p. 2..

[2] Rasheed, Z. et al. "On the use of computable features for film classification." IEEE Transactions on Circuits and Systems for Video Technology 15 (2005): 52-64

[3] M. A. Hedjazi, I. Kourbane and Y. Gene, "On identifying leaves: A comparison of CNN with classical ML methods." 2017 25th Signal Processing and Communications Applications Conference (SIU), 2017

[4] S. Albawi, T. A. Mohammed and S. Al-Zawi, "Understanding of a convolutional neural network," 2017 International Conference on Engineering and Technology (ICET), 2017

[5] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. and Chen, L.C., 2018. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4510-4520).

[6] Sandler, Mark & Howard, Andrew & Zhu, Menglong & Zhmoginov, Andrey & Chen, Liang-Chieh. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks.

[7] warehouse-lighting.com. Movie Theater Lighting, https://www.warehouse-lighting.com/biogs/lighting-application-suggestions/movie-theatre-lighting

[8] Kevin A.G. Smet, 2017, Two Neutral White Illumination Loci Based on Unique White Rating and Degree of Chromatic Adaptation

DYAH E H 'MOVIE POSTER GENRE' PROSIDING_0001

ORIGINALITY REPORT

1%	0%	1%	0%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

1%

★ "Table of Contents", 2021 IEEE Seventh International Conference on Multimedia Big Data (BigMM), 2021

Publication

Exclude quotes	On	Exclude matches	< 1 words
Exclude bibliography	On		