ICIC Express Letters
Volume **15**, Number **2**, February **2021**

ICIC International ©2021 ISSN 1881-803X
pp. **109–115**

MULTI-LABEL CLASSIFICATION OF GENRE FILM BASED ON POSTER WITH THE CONVOLUTIONAL NEURAL NETWORKS (CNN) METHOD

Jonathan Immanuel* and Sani Muhamad Isa

Computer Science Department, BINUS Graduate Program – Master of Computer Science
Bina Nusantara University
Jl. Kebon Jeruk Raya No. 27, Kebon Jeruk, Jakarta Barat 11530, Indonesia
*Corresponding author: jonathan.immanuel@binus.ac.id

Received May 2020; accepted August 2020

Abstract. *Movie is an important part of our lives and today when we find movie posters, we can quickly discuss elements such as colors, faces, objects and much more to get an accurate understanding of what the movie is like. Inspired by humans, researchers intend to build a Convolutional Neural Network (CNN) that can review visual movie posters to classify them into one or more genres on posters. Because the movie can refute into several genres, researchers also asked for a multi-label classification in this study. To facilitate this research, researchers collected several sets of movie posters documenting more than 500 films from 2010-2015 in 5 different genres. So, the model can predict multi-label genre using AlexNet method and object detection by YOLO v3 with Naive Bayes classification. Accuracy results obtained from the multi-label genre with the AlexNet method on the poster produce Precision of 0.74, Recall of 0.7, and F1 Score of 0.72. Accuracy results obtained from the multi-label genre with the object detection by YOLO v3 with Naive Bayes classification produce Precision of 0.72, Recall of 0.74, and F1 Score of 0.73.*

Keywords: Multi-label classification, Movies classification, Convolutional Neural Network (CNN), AlexNet, Object detection, YOLO v3, Naive Bayes classification

1. **Introduction.** The number of activities or work activities that we do every day, sometimes can make us feel depressed, bored, or stressed. Work that does not stop added deadlines that must be pursued can cause fatigue. For various ways to relieve stress, one of them is looking for entertainment. One such entertainment is watching movies. Almost everyone would love movies, from children to adults though, likes action genre, romantic love, adventure, comedy or even animation. Moreover, with the advancements in technology, watching movies does not have to go to the cinema, and there are many choices offered such as streaming or downloading various films. According to Tirto's research (2017) "*Choice of Means to Watch Gen Z Java-Bali Films*" it is said that 83.01% of respondents decide for themselves what type of film they want to watch, 6.91% of decisions are based on friends, and the rest is influenced by parents.

A film poster is one of the first indicators seen that gives an idea about the contents of the film and the genre that is in it before deciding to watch a film. Everyone makes a personal decision to watch a film according to the circumstances. For example, when he/she is in love, then the main choice in watching the romance genre film or when he/she wants to release the burden and lighten the mood, then the choice is the comedy genre, or when he/she wants to get inspiration or motivation from someone, then the choice of film is the inspiration genre. In its application, humans can understand cues such as colors, expressions on the faces of actors and so can quickly determine the genre of a film in posters such as horror, comedy, and animation. This fact is supported by Charles

DOI: 10.24507/icicel.15.02.109

Darwin's research, namely, "*The Expression of the Emotions in Man and Animals*", in 1872 concerning nonverbal communication, humans can respond to hundreds or even thousands of cues and other nonverbal behaviors, such as facial expressions, postures, movements, and tone of voice through subtle signals that we are not aware of. According to Wikipedia, nonverbal communication is the process of delivering messages not by using words. This shows that the color characteristics of an image such as hue, saturation, brightness, or contour affect human emotions without us knowing it. Certain situations evoke these emotions in humans. If humans are able to classify film genres with a glance at seeing a poster, then we can assume that the color characteristics, texture-based features, structural cues, or objects of posters have several characteristics that can be used in machine learning algorithms to classify them [1,2,12].

In film posters there is usually not only one genre label but can be more than one different genre (*cross-genre*) and it becomes one so that the resulting film becomes more interesting. Therefore, we can also estimate a few labels in the poster, for example, there is a poster titled "*Avengers Infinity War*", we can estimate several genres, namely Science Fiction (*Sci-Fi*) and Action as well. This title is called as "*Multi-Label Classification*". Multi-label classification consists of several classes that are mutually independent of each other where the genre of one another is very different in type. An example above between the Sci-Fi and Action genres is 2 different types of genres [7].

From this background, researchers want to further explore the Convolutional Neural Network (CNN) method in detecting and recognizing objects in images. The dataset used consists of several attributes in the form of images to determine the genre classification of a film later [3,4]. Compared with the previous methods, our methods have improved the multi-label classification. The main contributions of this paper are as follows.

1) We propose 2 methods, such as AlexNet method and object detection by YOLO v3 with Naive Bayes classification, which can better exploit the advantages of CNN model's capabilities to extract deep features of images in multi-label classification.

2) We introduce an automation model that can classify the multi-label genre of a movie poster by the CNN method. We also compare methods that can provide better performance.

The remainder of this paper is organized as follows. Section 2 describes the related work. Section 3 presents the proposed methodology, that is, preprocessing data, AlexNet method, object detection by YOLO v3 with Naive Bayes classification, and evaluation model in detail. Section 4 provides the experimental results. Section 5 presents the conclusions of our research and discusses future work.

2. **Related Work.**

The first reference literature by Chu and Guo used a visual representation by CNN with a combination of existing deep learning methods. They collected a large-scale movie poster dataset and fine-tuned a pretrained convolutional neural network to extract visual representation in posters. Multi-label classification is archived by thresholding the estimate probabilities [1].

The second reference literature was conducted by Ivasic-Kos et al. who probably first discussed the problem of the classification of film posters. In the study, they used 1,500 datasets with 6 genres, namely, action, animation, comedy, drama, horror, and war. The attributes used are low-level features, such as dominant colors, borders, and the number of faces on the poster. The results issued in the form of one or two genres based on the closeness of the existing genre using distance ranking method [3].

Then their research continues in the third literature by adding 6,000 poster datasets with 18 genres. The attributes used are GIST like dominant colors and color moments (feature representation) in posters with the scheme classification method, namely Naive Bayes for multi-label classification [6].

3. **Proposed Methodology.** In this research work, using 2 different methods, AlexNet method and object detection by YOLO v3 with Naive Bayes classification for multi-label genre.

3.1. **Preprocessing data.** We have collected 500 posters of movies from 2010-2015 with 5 genres, namely, Action, Sci-Fi, Romance, Comedy, and Horror. Each genre has 100 examples. Every movie can have several genres such as Thriller, Crime, Fantasy, or Adventure that appear in the data. Genres that have multiple genres are transformed into 5 genre classes. And because of data and genre limitations only take 2 labels out of the 3 labels that are commonly found in film posters. The following are some rules related to the transformation above, as follows (Table 1).

Table 1. Transformation rules

Genre labels original	Genre class label
Action, Adventure	Action
Sci-Fi, Action	Sci-Fi
Comedy	Comedy
Drama, Romance	Romance
Horror, Thriller	Horror

Our goal is to apply this data preprocessing to preventing unstructured data caused by too many genre labels in the dataset. So 500 film posters were collected, of which 450 are for training data, and 50 for testing data. According to Wang et al. use CharCNN-SVM method for text classification to obtain the emotional tendencies of user reviews. Use data augmentation to reduce secondary interference involving low-frequency synonyms appearing in the text [16]. We also implement this method to our data to reduce interference.

3.2. **AlexNet method.** The reason for choosing the AlexNet method is because this method has been trained using a large dataset so that the resulting image value will be much better. The second reason is that AlexNet uses a suitable architecture to be used to handle multi-label classification [5,6]. The characteristics contained in the AlexNet method are as follows.

1) The color layer uses 3 color vectors (RGB), namely Red, Green and Blue. This color layer will begin to appear on the 6th, 7th and 8th layers of the AlexNet method.

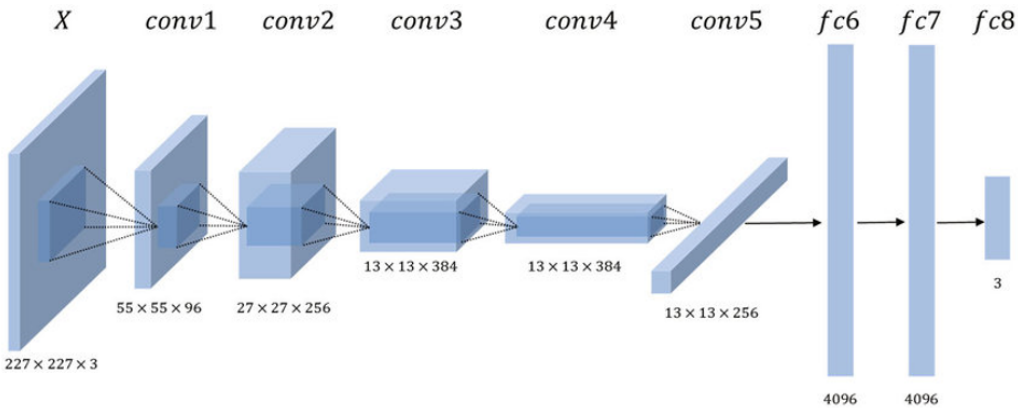

Figure 1. Proposed architecture of AlexNet

2) Color classes used include black, blue, green, brown, pink, purple, red, white, yellow, gray, orange, and others.
3) The variables used are Gaussian with mean zero and standard deviation.

3.3. **Object detection by YOLO v3 with Naive Bayes classification.** Use YOLO, "You Look Only Once", which is a neural network that recognizes images and where stuff is, in one process. YOLO provides a bounding box around the detected object to allow one grid cell to detect multiple objects at the same time. Below is the formula for determining dimension of priority and location prediction of bounding boxes.

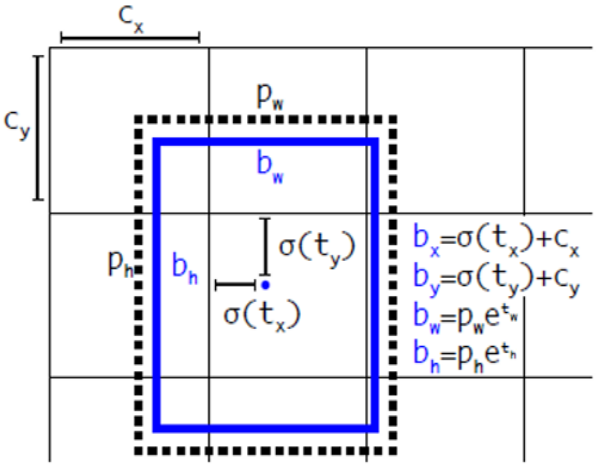

FIGURE 2. Dimension of priority and location prediction of bounding boxes

We use sum of the squared error losses for our training. When calculated from the basic truth box, we use a gradient whose basic truth value is reduced by our prediction $(\hat{}t^* - t^*)$ when the basic truth for some coordinate predictions is $(\hat{}t^*)$. The following are the basic truth values that can be calculated by reversing the equation:

$$\begin{aligned} b_x &= \sigma(t_x) + c_x \\ b_y &= \sigma(t_y) + c_y \\ b_w &= p_w e^{t_w} \\ b_h &= p_h e^{t_h} \end{aligned} \tag{1}$$

where $(\hat{}t^*)$ is some coordinate prediction, (b_x, b_y, b_w, b_h) are the central point coordinates, length, and width of the forecasting frame, σ is a softmax notation symbol, (t_x, t_y) are the offset of the target central point relative to the upper left corner of the grid where the point is located, (c_x, c_y) are the number of lattices with difference between the upper left corner and the upper left corner of the grid where the point is located, (p_w, p_h) are the edge lengths of the anchor box, and (t_w, t_h) are the width and height of the forecasting frame.

The major innovation YOLO is performing the detections in one pass, so that it is quite fast and powerful. It also works with a regression and can predict the bounding boxes and the class probabilities for each with a single network. Other approaches usually use task pipelines, such as forwarding a picture of several classifiers to detect things in different locations and utilizing some other additional methodology [1,11]. By using logistic regression, YOLO v3 can predict objectivity scores in each bounding box. If the boundary box overlaps with the ground truth object more than the previous boundary box, the value will be 1. If the bounding box is not given value but overlaps with the truth object more than a few thresholds, it will ignore the prediction. If the bounding box is not previously set to the ground truth object it will cause damage to the coordinates or prediction of the class [10,11].

After getting the object in the poster, we classify the movie genre of the object using the Naive Bayes classification. The ground truth is taken from the 25M MovieLens which

consists of 1,129 tags in which there are objects that are used to support the needs of this research. Use the Naive Bayes classification method with the following formula [3,13]:

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)} \tag{2}$$

X is data with unknown classes, C is the data hypothesis in a specific class, $P(C|X)$ is hypothesis probability based on conditions, $P(C)$ is hypothesis probability, $P(X|C)$ is probability based on conditions on the hypothesis, and $P(X)$ is probability C.

The advantage of using this method is that it only requires a small amount of training data to determine the estimated parameters needed in the classification process. Because it is assumed to be an independent variable, only the variance of a variable in a class is needed to determine classification.

3.4. **Evaluation.** In multi-label classification, use different evaluations by calculating the average difference between the actual data and the label prediction in the test data [8,9].

Precision is the ratio of the average predicted label to be true with all prediction labels correct for each example.

$$\text{Precision} = (\text{TP})/(\text{TP} + \text{FP}) \tag{3}$$

where TP is True Positive, and FP is False Positive.

Recall is the ratio of the average correctly predicted to all basic truth labels for each example.

$$\text{Recall} = (\text{TP})/(\text{TP} + \text{FN}) \tag{4}$$

where TP is True Positive, and FN is False Negative.

F1 Score is the average between precision and recall.

$$\text{F1 Score} = (2 * (\text{Recall} * \text{Precision}))/(\text{Recall} + \text{Precision}) \tag{5}$$

where Precision is the level of accuracy between the information requested by the user and the answer given by the system and Recall is the level of success of the system in finding back information.

4. **Experimental Results.** In this research work, AlexNet method and object detection by YOLO v3 with Naive Bayes classification are shown below.

Results = {Sci-fi, Action}
(a)

Results = {Sci-fi, Action}
(b)

FIGURE 3. (a) Example of result of AlexNet method in poster; (b) example of result of object detection by YOLO v3 with Naive Bayes classification in poster

Figure 3(a) shows Transformers poster has a clear color scheme of the American flag. The colors come from a head of Transformer Optimus Prime, portrayed on the movie poster. The background behind it looks like a city and cloud, but the color scheme of the background is a smooth grey. It is rare to use large amounts of grey as a metallic nature. This movie is all about robots and metal which is the most logical choice to include a lot of grey in the poster.

Due to the dominant of grey, both the blue and the red act as standout colors, especially as the blue area of the Transformer's helmet is larger than that of the red on his shoulder. The same blue color of the helmet can also be found in the bottom of the poster, giving it all a perfect balance and keeping the blue of the helmet from overpowering the rest of the poster and showing the entire design out of balance. We implement framework based on the combination of ADMM and global consensus optimization which is good to distributed optimization techniques to training CNN and get an efficient poster labeling on GPU. Use different grey scale colors to visualize our labeling results to get faster labeling speed [17].

Figure 3(b) shows that by using multi-label classification, each box will predict classes that may contain bound boxes. We do not use softmax because we do not need to get good performance, so we replace it by using an independent logistics classifier. During training we use categorical cross-entropy loss for multi-label prediction.

From 500 datasets of posters images, we used 10-Fold Cross Validation which makes 450 for training data and 50 for testing data. Our evaluation for AlexNet method with Precision scores were 0.74, Recall scores were 0.7, and F1 Scores were 0.72. And object detection by YOLO v3 with Naive Bayes classification with Precision scores was 0.72, Recall scores was 0.74, and F1 Scores was 0.73. Evaluation results for multi-label classification are shown in Table 2.

Table 2. Evaluation results for multi-label classification

Label-based evaluation	**AlexNet method**	**Object detection by YOLO v3 with Naive Bayes classification**
Precision (P)	0.74	0.72
Recall (R)	0.7	0.74
F1 Score (F1)	0.72	0.73

Table 3 shows comparison results of multi-label classifications between the proposed methods and those of other previous researchers. Our method has the highest validation F1 Score compared to the previous related work, which is 0.72 for AlexNet method and 0.73 for object detection by YOLO v3 with Naive Bayes classification on model with two output genres.

Table 3. Comparison results of multi-label classification

Author	**Method**	**F1 Score**
Chu and Guo [1]	AlexNet	0.14
Ivasic-Kos et al. [3]	Low level features with distance ranking	0.14
Ivasic-Kos et al. [6]	GIST with Naive Bayes classification	0.38
Proposed method (*with 500 datasets*)	AlexNet	0.72
	Object detection with Naive Bayes classification	0.73

5. **Conclusion.** In this paper, we propose CNN architecture for multi-label classification of genre film using AlexNet method and object detection by YOLO v3 with Naive Bayes classification. The experimental results of multi-label classification into 2 genres on datasets with 500 movie posters like Action, Sci-Fi, Comedy, Romance, and Horror

show 0.72 for AlexNet method and 0.73 for object detection by YOLO v3 with Naive Bayes classification. In future aimed at multi-label classification can use more dataset and variety of genres to increase the accuracy and get better results.

Acknowledgment. The authors gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] W.-T. Chu and H.-J. Guo, Movie genre classification based on poster images with deep neural networks, *Proc. of MUSA2'17*, Mountain View, USA, 2017.

[2] J. Wehrmann and R. C. Barros, Movie genre classification: A multi-label approach based on convolutions through time, *Proc. of the Symposium on Applied Computing*, pp.114-119, 2017.

[3] M. Ivasic-Kos, M. Pobar and L. Mikec, Movie posters classification into genres based on low-level features, *The 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, Opatija, pp.1198-1203, 2014.

[4] W.-T. Chu and Y.-L. Wu, Deep correlation features for image style classification, *Proc. of ACM International Conference on Multimedia*, pp.402-406, 2016.

[5] G. S. Simoes, J. Wehrmann, R. C. Barros and D. D. Ruiz, Movie genre classication with convolutional neural networks, *International Joint Conference on Neural Networks*, p.8, 2016.

[6] M. Ivasic-Kos, M. Pobar and I. Ipsic, Automatic movie posters classification into genres, in *ICT Innovations 2014. Advances in Intelligent Systems and Computing*, A. Bogdanova and D. Gjorgjevikj (eds.), Cham, Springer, 2015.

[7] J. Wehrmann, R. C. Barros and R. Cerri, Hierarchical multi-label classication with chained neural networks, *ACM Symposium on Applied Computing*, 2017.

[8] R. S. Perdana, *Accuracy Measurement Using Precision and Recall*, https://rizalespe.com/pengukuran-akurasi-menggunakan-precisiondan-recall-71c04988e6ab, Accessed on May 13, 2019.

[9] Y. Kim, Convolutional neural networks for sentence classification, *Proc. of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, DOI: 10.3115/v1/D14-1181, 2014.

[10] J. Redmon and A. Angelova, Real-time grasp detection using convolutional neural networks, *CoRR*, abs/1412.3128, 2014.

[11] J. Redmon et al., You Only Look Once: Unified, real-time object detection, *arXiv Preprint*, arXiv:1506.02640, 2015.

[12] G. S. Simoes, J. Wehrmann et al., Movie genre classification with convolutional neural networks, *Proc. of International Joint Conference on Neural Networks*, 2016.

[13] M. Pobar and M. Ivasic-Kos, Multi-label poster classification into genres using different problem transformation methods, in *CAIP 2017. LNCS*, M. Felsberg, A. Heyden and N. Krüger (eds.), https://doi.org/10.1007/978-3-319-64698-5_31, Cham, Springer, 2017.

[14] A. Adam, Choice of means to watch Gen Z Java-Bali Films, *Tirto.id*, https://tirto.id/revolusi-gayamenonton-ala-gen-z-ctUd, Accessed on December 12, 2019.

[15] C. Darwin, Concluding remarks and summary, in *The Expression of the Emotions in Man and Animals*, New York, D. Appleton and Company, 1872.

[16] X. Wang, Y. Sheng, H. Deng and Z. Zhao, CharCNN-SVM for Chinese text datasets sentiment classification with data augmentation, *International Journal of Innovative Computing, Information and Control*, vol.15, no.1, pp.227-246, 2019.

[17] J. Fu, Y. Huang, J. Xu and H. Wu, Optimization of distributed convolutional neural network for image labeling on asynchronous GPU model, *International Journal of Innovative Computing, Information and Control*, vol.15, no.3, pp.1145-1156, 2019.