Abstract. EMOTION DETECTION OF YOUR FAVOURITE TOONS

K. Ramprasad [1] M. Madhav [2] N. Pavan Kumar [3] P Sai Nihal [4] Dr. P. V. Siva Kumar [5]

[1 2 3 4 5] Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering and Technology, Hyderabad, Telangana, India

[1] rpramprasad00@gmail.com [2] mmadhav1234@gmail.com
[3] nimmanapavan@gmail.com [4] sainihalclg@gmail.com
[5] sivakumarpv@vnrvjiet.in

Abstract. Emotion Recognition is a field that computer systems have become eminent at identifying; whether or not it's via pictures, video or audio. Emotion Recognition has proven promising improvements when combined with classifiers and Deep Neural Networks. This includes detecting facial expressions in pictures and videos. While the majority of studies makes use of human faces while trying to apprehend fundamental emotions, there has been little research on whether similar deep learning strategies may be implemented to faces in cartoons. The motive of this project is to know how precise a machine can efficiently perceive emotion from a given set of pictures obtained from animated videos using convolution Image classification Models like Resnet, Inceptionv3, MobileNet and VGG and an object detection Model. This project is about reflecting on their success and discovering if these deep learning strategies may be implemented to research precise statistics. This selection is primarily based on the truth that cartoons are known to express numerous emotions, especially in the characters, and the option being 'animated cartoons' is that we can extract emotions from those characters in movies.

Keywords: Convolution Neural networks, Resnet, Mobilenet, Inceptionv3, VGG, deep learning, Emotion Recognition and object detection.

1 INTRODUCTION

Emotion is a subjective physical and psychological state, composed of related thoughts, feelings, and behavioral responses, all of which are homogeneous. Emotion recognition is very useful in artificial intelligence systems because it enables the system to recognize and predict human emotions to improve performance. Recently, automatic emotion detection has become a popular research area for industry researchers and scholars who specialize in machine consciousness, computer vision, brain computing, physiology, and, more recently, deep learning. Its universality is due to its extensible application area. Ekman and Friesen conducted a popular research on emotion recognition, dividing emotions into six basic expressions of happiness, sadness, disgust, anger and fear. And they have proven to be versatile. His research has become a benchmark for evaluating research in the field of emotion recognition.

The present work is dedicated to identifying emotions from the facial expressions of animated characters. Even emotion recognition on humans has been very successful in the recent times, the recognition of emotions from cartoon images is still a poorly understood field. In order to solve this problem, a new neural network was developed to define the emotions of the cartoon characters, which classifies the cartoon face and recognize the emotions. The extracted cartoon face is used as input to emotion recognition model to recognize the emotion of character. This article considers two cartoon characters Tom and Jerry for analysis. Recognizable emotions are divided into the following categories (sadness, happiness, anger, surprise and unknown).

2 METHODOLOGY

Facial expressions have a significant contribution to interpersonal communication. In this paper, the Resnet model is used to detect the face of a character by separating the foreground and background pixels using the bounding box of the segmented face. Emotion detection of Tom and Jerry cartoon has majorly four implementation parts. The following modules are to be built and an input image will be provided to get the desired output for an image.

- First the data videos need to be collected and frames need to be extracted from at certain frame from a video.
- Now an image classification model is to be built to detect the character of the cartoon. Here in this project, we need to detect whether it is a Tom and jerry image or an unknown image.
- Then an emotion classification model is to be implemented to detect an emotion for certain input image.
- Finally, a Face detection model is to be implemented which uses both character classification model and emotion detection model and certain another algorithm like sliding window and non max suppression

3 IMPLEMENTATION

3.1 IMPORTING LIBRARIES

To start with the initial phase, it is important to examine the dataset. Therefore, we follow the method of Exploratory Data Analysis (EDA) to analyze and identify the characteristics of the dataset. Then import all the libraries required such as NumPy, Pandas, Matplotlib, Seaborn followed by the libraries for evaluating such as Sklearn.metrics, and libraries for preprocessing and model building like ImageDataGenerator, Sequential, MobileNet.

3.2 DATA GATHERING

To implement a tom and jerry character image classification model, the data should be collected and categorized correctly. In this classification model we have two output modules. One is tom and jerry and other is unknown. For example, if we provide an input as image, it will predict one of the above with certain probability.so the data need to be kept in two different folders separately. The data processing techniques need to be followed to feed an input data to image classification models.

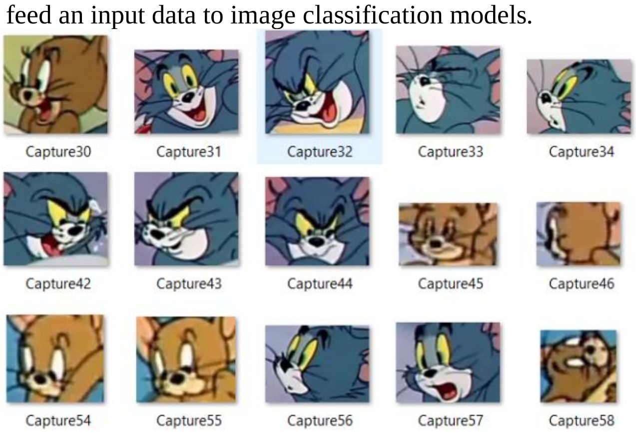

Fig. 1. Dataset

3.3 IMAGE CLASSIFICATION MODELS

Here in this project, we are using four image classification model and considering the model for which we are getting Higher accuracy. The four classification models are VGG19, Resnet50, MobileNet and InceptionV3.We are applying these image classifications for both character classification model and emotion classification model.

RESNET MODEL.

It is a powerful basic model widely used in many computer-vision tasks, ResNet uses transform connections to add output from earlier layers to later layers. This helps alleviate the problem of gradient fading. Keras is used for loading pre-trained Resnet.

```
from tensorflow.keras.applications import ResNet50
base_model=ResNet50(include_top=False, weights='imagenet',input_shape=(80,80,3), pooling='avg')

model=Sequential()
model.add(base_model)

model.add(Dense(256,activation='relu'))
model.add(Dense(2,activation='softmax'))

model.summary()
```

Fig. 2. ResNet model

MOBILENET MODEL.

MobileNet is an optimized architecture that uses deep analytical convolution to build deep and lightweight convolutional neural networks, and provides efficient models for embedded and mobile vision applications.

```
from tensorflow.keras.applications import MobileNet
#from tensorflow.keras.applications.vgg19 import VGG19
#from tensorflow.keras.applications.inception_v3 import InceptionV3

base_model=MobileNet(include_top=False, weights='imagenet',input_shape=(80,80,3), pooling='avg')

model=Sequential()
model.add(base_model)

model.add(Dense(256,activation='relu'))
model.add(Dense(2,activation='softmax'))

model.summary()
```

Fig. 3. MobileNet model

INCEPTION MODEL.

Inception v3 is a neural convolutional network that supports image analysis and object recognition. It was introduced as a module of Googlenet. This is the third edition of Google's Inception Convolutional Neural Network, which was originally shown during the ImageNet Recognition Challenge.

```
from tensorflow.keras.applications.inception_v3 import InceptionV3

base_model=InceptionV3(include_top=False, weights='imagenet',input_shape=(100,100,3), pooling='avg')

Emodel=Sequential()
Emodel.add(base_model)

Emodel.add(Dense(256,activation='relu'))
Emodel.add(Dense(5,activation='softmax'))

Emodel.summary()
```

Fig. 4. Inception model

VGG19 MODEL.

VGG19 is a variant of the VGG model. In short, it consists of 19 layers (16 folded layers, 3 fully linked layers, 5 MaxPool layers and 1 SoftMax layer). It is used as a good classification framework for many other data sets, and they can be used as it is or modified for other similar tasks.

```
from tensorflow.keras.applications.vgg19 import VGG19

base_model=VGG19(include_top=False, weights='imagenet',input_shape=(100,100,3), pooling='avg')

Emodel=Sequential()
Emodel.add(base_model)

Emodel.add(Dense(256,activation='relu'))
Emodel.add(Dense(5,activation='softmax'))

Emodel.summary()
```

Fig. 5. VGG19 model

3.4 FACE DETECTION MODELS

In a Face detect module an image is given as input and the bounding box around tom and jerry faces is shown as output with an emotion of the face. It has some sub function which it implements and they are sliding window, Image pyramids and non-Max suppression.

IMAGE PYRAMID.

Image pyramid function takes an image and scale as an input and gives a bunch of images with resizing the same image without losing image content as an output.

SLIDING WINDOW.

This function takes an image, windows length and breadth and a step to move the window as an input. The is function takes all the windows from image and send to character classification Model to get the probability for each window.

NON-MAX SUPPRESION.

This algorithm takes bunch of images with probability as input and give an output image which is having a maximum probability.

4 RESULTS

At the point when a framework is created, it's trusted that it performs appropriately. In follow, however a few blunders never-endingly happen. the most motivation behind testing an information framework is to look out the blunder and legitimate them.
After applying the deep neural network to a fairly large data set, the data will go through a pre-processing stage. At this stage, the model is used to generate an accurate face from an image. As input to the four deep neural network models, as described above, the emotion confidence value (that is, the probability that the corresponding character recognizes the emotion) is used to identify the emotion of a specific character, and the value ranges from 0 to 1. The classification accuracy of the faces and emotion is listed below:

Image classification Model	Tom & Jerry classification Accuracy	Emotion classification Accuracy
Resnet50	99%	94%
MobileNet	90%	78%
InceptionV3	81%	47%
VGG19	81%	92%

The accuracy and loss graphs of tom and jerry classification and emotion classification are as follows

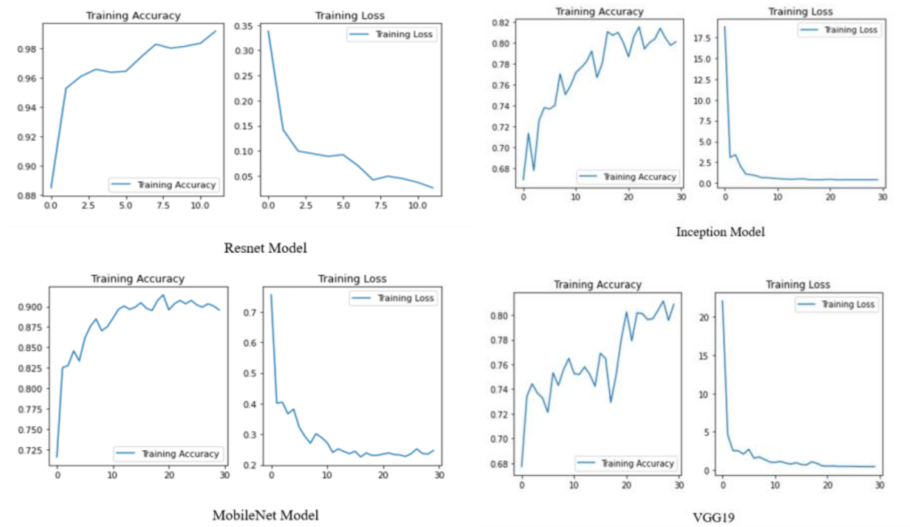

Fig. 6. Training accuracy and loss graphs of Tom and Jerry classification

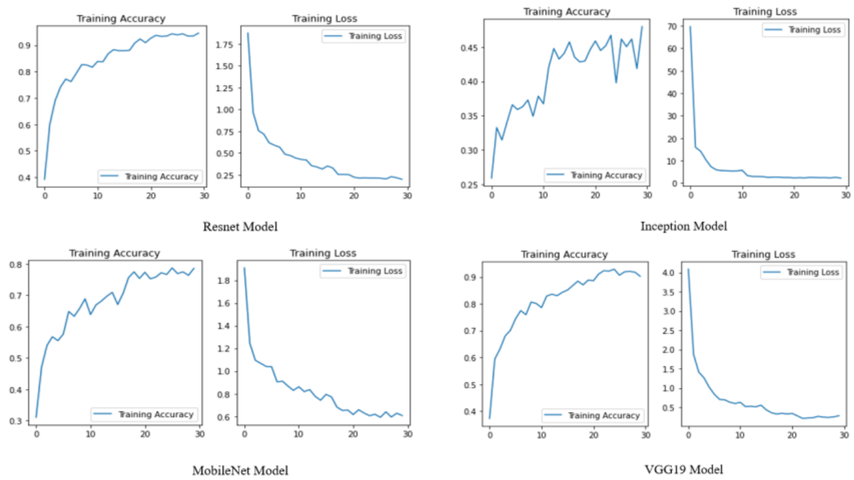

Fig. 7. Training accuracy and loss graphs of emotion classification

Fig. 8. Testcases

- Resnet model iterated for 12 epochs and got an accuracy of around 99% for tom and jerry classification model and in emotion classification, it iterated for 30 epochs for five labeled outputs and the model shows and accuracy of around 94%.
- MobileNet model iterated for 30 epochs and got an accuracy of around 90% for tom and jerry classification model and in emotion classification, it iterated for 30 epochs for five labeled outputs and the model shows and accuracy of around 78%.
- Inception model iterated for 30 epochs and got an accuracy of around 81% for tom and jerry classification model and in emotion classification, it iterated for 30 epochs for five labeled outputs and the model shows and accuracy of around 47%.
- VGG19 model iterated for 30 epochs and got an accuracy of around 81% for tom and jerry classification model and in emotion classification, it is iterated for 30 epochs for five labeled outputs and the model shows and accuracy of around 92%.

5 CONCLUSION

Emotion analysis is showing promising results when applied on human faces. It is an interesting task when same deep neural networks are applied to detect an emotion on cartons. In this project the data is selected from you tube and the data preprocessing techniques are applied to the collected data to feed deep neural networks. Data is separated for both character classification model and emotion classification model. An object detect model is implemented which utilizes the classification models and non max suppression algorithm to detect emotion from an image more accurately. As the project is showing promising results for emotions detection on cartoons. So, it can be concluding that the same deep neural networks can also be applied on cartoon faces by using our own object detection model.

6 REFERENCES

1. Dionysis Goularas,Sani Kamis,Evaluation of Deep Learning Techniques in Sentiment Analysis from Twitter Data, Published in: 2019 IEEE
2. Yongjing Lin;Huosheng XieFace Gender Recognition based on Face Recognition Feature Vectors, Published in :2020 IEEE 3rd International Conference on Information Systems and Computer Aided Education (ICISCAE)
3. Li Xu;Lei Huang;Changping Liu, Raw vs. Processed: How to Use the Raw and Processed Images for Robust Face Recognition under Varying Illumination, Published in :2010 20th International Conference on Pattern Recognition
4. Rajesh Kumar Tripathi;Anand Singh Jalal;Charul Bhatnagar A framework for abandoned object detection from video surveillance,2013 Fourth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)
5. Mansi Mahendru; Sanjay Kumar Dubey, Real Time Object Detection with Audio Feedback using Yolo vs. Yolo_v3,Published in**:** 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)
6. Z. Zhao, Q. Zheng, P. Xu, S. T and X. Wu, "Object detection with deep learning: A review", *IEEE transactions on neural networks and learning systems*, vol. 30, no. 11, pp. 3212-3232, 2019.
7. J. Redmon and A. Farhadi, "YOLO9000: better faster stronger", *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7263-7271, 2017
8. Zhuo Zheng;Hao Zhang;Xinjian Li;Shuai Liu;Yukun Teng, ResNet-Based Model for Cancer Detection, 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE)
9. J. Umamaheswari;A. Akila, An Enhanced Human Speech Emotion Recognition Using Hybrid of PRNN and KNN,2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)
10. Y. Zhang, K. Sohn, R. Villegas, G. Pan, and H. Lee, “Improving object detection with deep convolutional networks via bayesian optimization and structured prediction,” in CVPR, 2019