Unveiling Emotions in Cartoons: A Deep Learning Approach for Accurate Emotion Recognition in Animated Characters

Ms. Preeti Sehrawat
(Assistant Professor, MSIT)

Shubham Kumar • Abhishek Sharma • Arjun Sharma

Abstract

Emotion is a spontaneous or intuitive feeling as distinguished from logic or knowledge. It varies over time, since it's a natural spontaneous state of mind inferring from one's circumstances, mood, or connections with others. Since feelings vary over time, it's important to understand and dissect them meetly. Being workshop have substantially concentrated well on feting introductory feelings from mortal faces. still, the emotion recognition from cartoon images has not been considerably covered. thus, in this paper, we present an intertwined Deep Neural Network (DNN) approach that deals with feting feelings from cartoon images. Since state- of- workshop don't have large quantum of data, we collected a dataset of size 8 K from two cartoon characters 'Tom' & 'Jerry' with four different feelings, videlicet happy, sad, angry, and surprise. The proposed integrated DNN approach, trained on a large dataset conforming of robustness for both the characters (Tom and Jerry), rightly identifies the character, parts their face masks, and recognizes the consequent feelings with a delicacy score of 0.96. The approach utilizes Mask R- CNN for character discovery and state- of- the- art deep literacy models, videlicet ResNet- 50, MobileNetV2, InceptionV3, and VGG 16 for emotion bracket. In our study, to classify feelings, VGG 16 outperforms others with a delicacy of 96 and F1 score of 0.85. The proposed intertwined DNN outperforms the state- of- the- art approaches.

Keywords Animation, Cartoon, Character Detection, Convolutional Neural Network, Emotion, Face Segmentation, Mask R-CNN, VGG16

To the best of our knowledge, these are the largest data for cartoon emotion classification and are available for research purpose.

& Ms. Preeti Sehrawat
(Assistant Professor, MSIT)
preetisehrawatece@msit.in

Shubham Kumar
shubhamzxv@gmail.com

Abhishek Sharma
abhishekshrma6923@gmail.com

Arjun Sharma
ar29061999@gmail.com

1 Introduction

An emotion is a physiological state of mind that's private and is constituted by associated studies, passions, and behavioural responses that are homogeneous. Feting feelings is largely usable in instinctively intelligent systems, to enable similar systems in feting and prognosticating mortal feelings to enhance productivity and effectiveness of working with computers. lately, automatic recognition of emotion has come a popular exploration field involving experimenters from the assiduity, as well as academia, specializing in artificial knowledge, computer vision, brain computing, physiology, and, more lately, deep literacy. Its ubiquity emerges from extensible regions of implicit operations. Ekman and Friesen 1) has conducted one popular exploration work in the field of Emotion Recognition, who have classified the feelings into six introductory expressions of happiness, sadness, nausea, wrathfulness, and fear and stated they're universal. Their exploration has come a standard for the evaluation of studies conducted in the field of emotion discovery.

Feting feelings is gaining thrust since the once many times. It can be performed by assaying data in any medium textbook [2 – 5], audio or speech [6 – 9], videotape, or images [10 – 13]. The exploration in the field of emotion recognition has handed precious data giving the emotional state of cases [14], response to an announcement, and indeed in times of extremity like coronavirus [15]. Emotion recognition from images generally involves facial expressions or gestures. Facial feelings are a form of on-verbal communication that conveys both the emotional state and behavioural intentions of an existent. The task of recognizing similar feelings can be performed over the facial image data of mortal beings or creatures or any real-world reality. There are several operations of this task in a variety of fields ranging from drug [16 – 18] and e-learning [19 – 22] to entertainment [23] and marketing [24 – 26] and indeed bar [27].

The operation of facial emotion recognition isn't limited to reading mortal face feelings. It can also be enforced to descry the feelings of animated characters or cartoons. Cartoons are substantially made, keeping in mind the entertainment and felicity of observers (especially children). They're frequently filled with colourful kinds of feelings that are portrayed in multiple forms by the same character.

The provocation behind current exploration lies in the fact that there is plenitude of feelings portrayed in cartoons, indeed by the same character and amped cartoons give an occasion, where one can prize feelings from these characters (from one or further vids). This idea of relating feelings is useful in cases where parents or guardians frequently want to choose an order of the cartoon (sci- fi, ridiculous, humor, riddle, and horror) grounded on their child's interest or according to their felicity.

To identify mortal faces, an image can be segmented using the OpenCV library[1] [28]. still, the employed algorithm misses detecting any other real- world reality [29] cartoon, in this paper). thus, we propose a general approach that uses a popular system Mask R- CNN to efficiently member objects. likewise, it was questionable to find an being dataset online owing to the time-consuming nature of data medication from vids in this task, encompassing character identification, as well as emotion identification. thus, it's in this regard that a dataset is erected (with two cartoon characters Tom & Jerry, presently) that can indeed be employed in different operations if intimately released on the web. also, the dataset is extensible for any number of cartoon characters keeping the general approach perpetration same.

1 https://opencv.org/.

The current work deals with feting feelings from facial expressions of cartoon characters. The ideal is to find out if DNNs can be stationed to prize and fete feelings from cartoons. Indeed, though emotion recognition has been considerably performed over mortal facial images; yet, feting feelings from cartoon images is still an under-explored area. To handle the same, a new integrated DNN approach has been developed to identify feelings from cartoon characters wherein the faces of characters are segmented into masks using the Mask R- CNN fashion. These generated masks are further used as input to the emotion recognition model to fete the feelings of the character. For the analysis conducted in this paper, two cartoon characters Tom and Jerry have been taken into account. The honoured feelings fall into the following orders (Sad, Happy, Angry, and Surprise). The stationed approach gives an F- score of0.85 when enforced on a created dataset of two characters used then viz. Tom and Jerry. The proposed approach is general and scalable to fete feelings.

The rest of the paper has been organized as follows sect, 2 puts forward the being literature and datasets for emotion recognition from animated images, including cartoons. Section 3 presents the figure of the benefactions done in this paper. Section 4 discusses accoutrements used in this work dataset collection, medication and the deep neural networks in detail. Section 5 presents the methodology of the proposed work. Section 6 shows the experimental analysis, results and comparison with the state of the art. Section 7 concludes the paper with an explanation of the attained results and their evaluation with the birth styles with farther compass of improvement.

2 Related works

Being exploration on emotion recognition from facial images is expansive. Facial expressions give accurate information allowing the bystander to separate between colourful negative and positive feelings; still, the substantiation relates to posed feelings only [30].

For emotion recognition, one of the workshops that has been conducted by [31] where the dataset prepared from several print spots similar as Flickr, Tumblr, and Twitter has been classified into five feelings viz. Love, Happiness, Violence, Fear, and Sadness. The authors have tested colourful pre-trained Convolutional Neural Network (CNN) models like VGG- Image Net, VGG- Places205, and ResNet- 50, out of which ResNet- 50 has performed the stylish, giving a delicacy of 73 after fine- tuning. Another donation by 32) where mortal feelings learned from 2D images have been transferred to animated cartoon 3D amped character for farther bracket into seven feelings (joy, sad, wrathfulness, fear, nausea, surprise, and neutral). The authors have proposed a fused CNN armature (f-CNN), giving a total recognition rate of75.5, having

originally a CNN trained on mortal expression dataset followed by a transfer literacy- grounded bracket to dissect the relationship between emotion transfer from 2D mortal images to 3D cartoon images. The recognition rate then's a parameter that shows how well an animated character can pretend a mortal face emotion.

Grounded on the emotion transfer conception, as mentioned in the former paragraph, authors in [33] have proposed a mortal animated face emotion bracket where mortal expressions have been dissembled using an animated face. The trials have been performed on a dataset of annotated (cartoonish) face images of several mortal stylized characters. Using a modified CNN armature, the trials attained different expression recognition for all feelings mentioned (as compared in Section Conclusion). still, a recent composition [34, 35] has argued that feting facial emotion of specific cartoon characters adds another challenge to descry emotion since cartoons generally depict extreme situations of feelings that aren't else seen and captured from mortal faces. The authors have also shown through interview-based trials that specific cartoon face emotion recognition requires a advanced processing intensity and speed than real faces during the early processing stage.

The author in [29] gives another colonist donation specific to cartoon character emotion recognition using Haar Cascade [28] and modified CNN armature for character discovery and emotion bracket, independently, from cartoon movie vids. Classifying three feelings (Happy, Angry, and Surprise), the trials achieved a bracket delicacy of 80. still, the authors claim that an enhancement in delicacy can be achieved by transfer literacy and hence proposes it as an open problem, thereby contributing a public dataset of emotion labelled cartoon character images.

Lately, several datasets [36, 37] have been contributed intended for trial of cartoon face discovery to the state of the art. These datasets, still, only contain the character markers and don't give any information about emotion markers for those characters. Hence, contributing a dataset having emotion labelled cartoon faces becomes another significant donation of this work. Such a dataset can be used for prospective exploration if the dataset is released for the public. operations in emotion recognition, as [38] points out, include avoidance, waking, product, training, and entertainment. The focus of this composition is to address the operations of training and entertainment. This can allow a computer to fete emotion in an animated cartoon automatically. It could induce mottoes (textbook or audio) explaining and tutoring the feelings of the characters throughout the videotape to children.

The ultimate relies on the fact that cartoons are a form of entertainment, to grown-ups and especially children. For illustration, a recommendation system can be designed where an animated cartoon has an emotion standing outlining which characters retain colourful feelings in an occasion.

Although there has been important work over mortal facial emotion recognition [39, 40], being literature on emotion recognition from cartoons is limited and has compass for expansive work. The mentioned benefactions don't give any cartoon emotion labelled dataset (attained from cartoon vids) and rather propose specific mortal facial expressions dissembled amped data. still, contouring of features in non-human faces (similar as cartoon characters) is different from mortal faces, which requires specific discovery styles. Also, the being character discovery algorithms that enable effective emotion bracket substantially calculate on dereliction libraries and modified CNN infrastructures helping in point birth. similar styles miss detecting a real- world reality (then a specific cartoon face) thereby giving low emotion recognition delicacy (ref. Section Results). The major benefactions of this paper are presented in side 3 as follows.

3 Contribution outline

Integrating DNN and validating it on a fairly (with respect to state- of- the- art workshop) large quantum of data to understand and dissect feelings are the primary end of the study. This allows us to have multiple objects:

(a) The proposed integrated DNN includes Mask R-CNN for cartoon character discovery and star deep literacy infrastructures models, videlicet VGG16, InceptionV3, ResNet- 50, and MobileNetV2 for emotion bracket. Compared to the state- of the- art workshop, the use of Mask R- CNN makes a difference in terms of performance (ref. Section Results). Further, employing multiple deep literacy infrastructures models provides a fair comparison among them.

(b) As no state- of- the- art workshop give large quantum of data for confirmation, we created a dataset2 of 8,113 images and annotated them for emotion bracket. This brings us a result to quantify test DNN meetly and is available for exploration purpose upon request).

2 https://github.com/emotionRecog.

The proposed approach, as depicted in Fig. 1, works by collecting and preparing dataset by downloading vids from a popular YouTube channel (ref. Sect.4.1) followed by character discovery and consequent emotion bracket through an intertwined DNN.

4 Materials

In this paper, a custom dataset conforming of the images uprooted from Tom and Jerry occurrences was created. The images demanded pre-labelled emotion classes for supervised bracket. The labelling process is defined in sect. 4.1. presently, there's no active dataset in this regard available. To rightly fete cartoon emotion from a given input videotape or image frame, the first and the foremost step includes character discovery followed by accurate face segmentation procedures.

4.1 Dataset collection and preparation

The occurrences for rooting the images have been linked and also downloaded from the Jonni Valentayn channel3 on YouTube using a downloading tool named Videoder4 in MP4 format. The frames were attained at a rate of 115 in the JPG format using OpenCV from the downloaded vids. From the frames uprooted, a training dataset has been generated for the Mask R- CNN model, which is latterly used to classify and member Tom & Jerry's faces from the input frames. Regarding the dataset prepared for training the Mask R- CNN model, the frames uprooted were classified into two classes, Tom and Jerry. Each frame has an associated class, which specifies that the cartoon character's face is present in the frame. A frame can have Tom's face, Jerry's face, or both.

For preprocessing, each data frame is stoked with a JSON train that stores the frame name, cartoon character name, and the X – Y equals of the corresponding cartoon face. The X – Y equals of the face were pronounced using a labelling tool, i.e., VGG Image Commentator (VIA).5 The pronounced regions were Tom's face and Jerry's face. The Mask R- CNN model learns this Region of Interests through the X – Y equals of cartoon character's faces marked through VIA tools. Frames with unknown faces were left unmarked. Figure 2 shows the screenshot of the affair given by the VGG commentator tool via where Tom's face is pronounced.

The dataset consists of 10 k images, i.e., 28 robustness of Tom and Jerry. The dataset generated from the Mask R- CNN model is annotated into four feelings. These feelings are Happy, Angry, Sad, and Surprise. For both cartoon characters, around 1000 images depicting each out of the four feelings were manually insulated. In total, images (or masks) of size 256 x 256 were used for the purpose of training after discarding the poor- quality images.

4.2 Mask Annotation for obtaining emotion labels

Three independent evaluators were employed to annotate the masked faces manually. These evaluators are professed and well clued in the field of vitality and multimedia.

The evaluators were asked to interpret each masked face with one of the linked markers in the dataset. The quality of reflection was measured by two standard agreement parameter sinter-Indexer thickness (IIC) [41] and Cohen's Kappa [42], with values of92.06 and84.9, independently. The feelings of both the cartoon character masked faces, videlicet Tom and Jerry, were annotated and also labelled in the following order of Sad, Happy, Angry, and Surprise, independently, in which the feelings of Jerry have recorded first and also feelings of Tom next in a analogous manner. The masked face images with their separate markers so attained after reflection (as described over) were used to train the emotion recognition model. Figure 3 shows the distribution of emotion markers used for training purposes. A fair distribution of feelings was used to have balanced supervised bracket training. The farther process to gain the masked faces from the character images (as set) is explained in Sect. 4.3.

4.3 Basics of CNN

This section provides an explanation on the working of CNN along with the other abecedarian generalities used.

(i) Convolutional Operation The main reason for performing complication is point birth. Applying convolutional operation on an image of size H x W x D, where H; W; D represent height, range, depth of image, independently, with complication sludge of size f x f x D, taking stride size equal to s with padding analogous to p gives an affair of size where n_f is the number of pollutants. The operation of complication between image matrix P of size (I, J) and kernel matrix Q of size (M, N) gives the affair O given by:

3 https://www.youtube.com/user/TomAndJerryWarner.
4 https://www.videoder.com/.
5 http://www.robots.ox.ac.uk/*vgg/software/via/.

Fig. 1 Workflow of the proposed approach

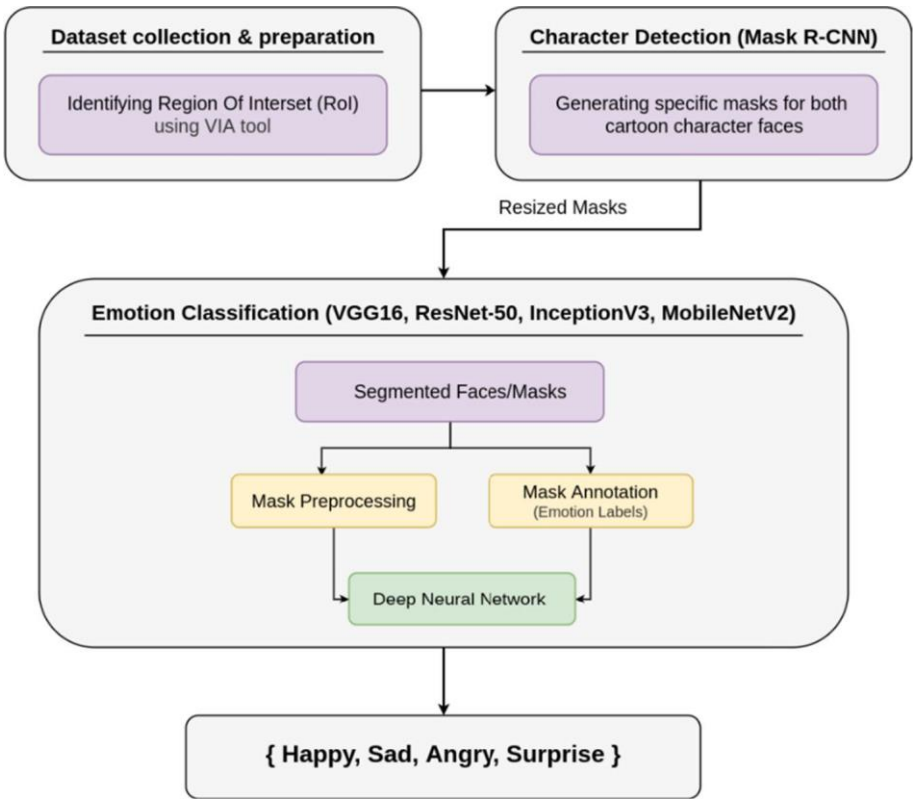
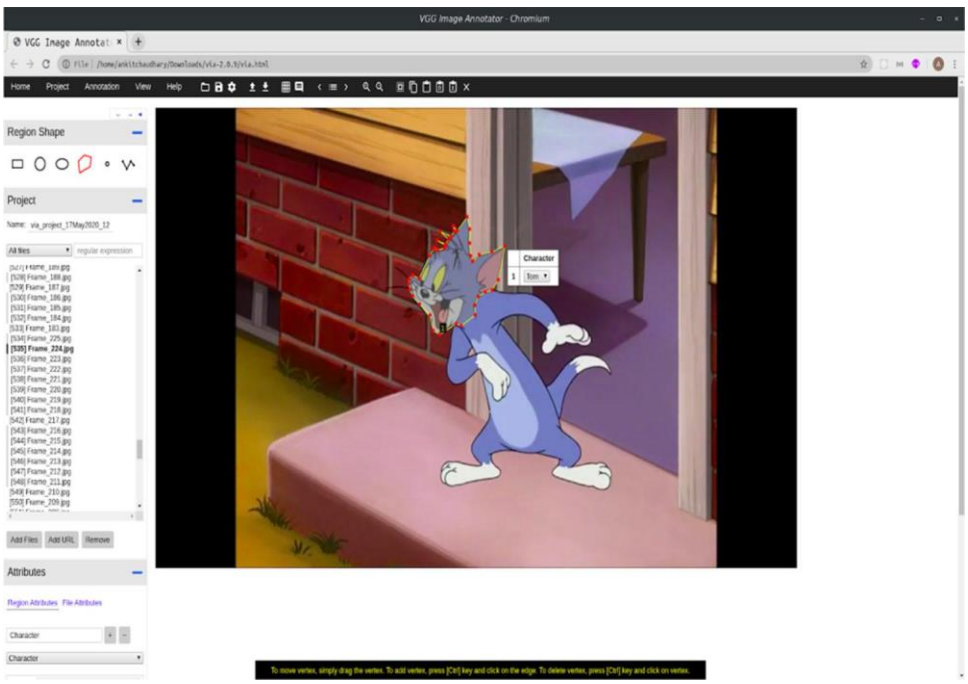


Fig. 2 Screenshot of the VGG Image Annotator



$$O(a,b) = (P * Q)[a,b] = \sum_{i=0}^{I} \sum_{j=0}^{J} P(i,j) * Q(a-i, b-j) \qquad (1)$$

where $0 \le a \le I + M + 1$ and $0 \le b \le J + N + 1$.

(ii) Pooling Operation Along with complications layers, CNN's use pooling layers like maximum-pooling and average pooling. therefore, for an image of size H x W with a sludge size of k and stride size s, the size of the affair is:

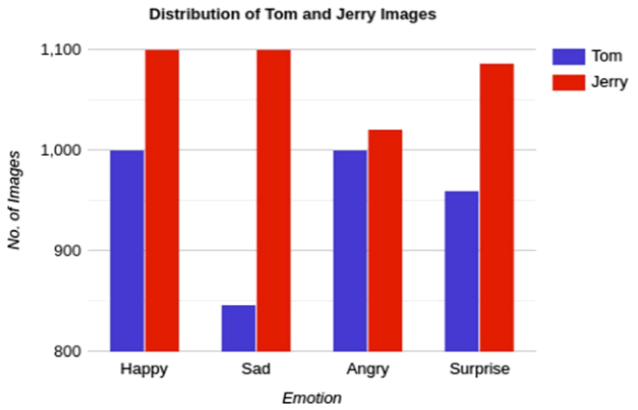


Fig. 3 Distribution of emotion labels after annotation, for the dataset

(iii) Bracket For the bracket, the completely connected layers serve as classifiers on top of the uprooted features given by retired layers. This subcaste also generates the final chances for determining the class for the input image, after applying the weights over the affair generated by point sensors.

(iv) Powerhouse Regularization This fashion refers to dropping out neurons (both retired and visible) in a neural network aimlessly.

(v) Activation Function ReLU is the most regularly used nonlinear function since it provides better performance than its druthers. Some of the generally used activation functions are mathematically expressed as follows:

$$\text{Simple ReLU} : f(x) = \max(0, x) \tag{2}$$

$$\text{Leaky ReLU} : f(x) = \begin{cases} x & x \geq 0 \\ 0.01x & \text{otherwise} \end{cases} \tag{3}$$

$$\text{Parameteric ReLU} : f(\alpha, x) = \begin{cases} \alpha x \text{ for } x < 0 \\ x \text{ for } x \geq 0 \end{cases} \tag{4}$$

4.3.1 Mask R-CNN

Mask R- CNN is an extension of Faster R- CNN. Faster RCNN gives two labours for every object — a class marker and the bounding box equals. In Mask R- CNN, a third branch for the affair of the object mask is added, which enables it to perform image case segmentation. The third added branch also shows the vaticination of the object mask in resemblant with being branches performing bracket and localization. For applicable case segmentation, pixel- position segmentation is performed, which requires precise alignment as compared to just the boundary boxes. Hence, Mask R- CNN uses RoI (Region of Interest) pooling subcaste known as RoIAlign Layer, so important more precise regions can be counterplotted for segmentation. The backbone of Mask R- CNN is a standard convolutional neural network (like ResNet- 50 or Resnet- 101), and it helps in the birth of features. These early layers descry features like edges and corners. After passing through the backbone network, the image is converted to a 32 x 32 x 2048-point chart from the given image. Figure 4 is the visual model for the armature of Mask RCNN. The Region Proposed Network (RPN) is a featherlight neural network that checks the image in a sliding door fashion to find the regions that contain the objects. RPN reviews over these regions (also known as anchors) using the backbone point chart rather of directly surveying over the image, enabling it to run briskly and more efficiently. Accordingly, it avoids indistinguishable computations by reusing uprooted features. The use of RoIAlign Layer fixes the position misalignment caused due to quantization in the case of RoIPool used in Faster R- CNN. Figure 5 shows the use of spatial motor and bilinear slice kernel. Bilinear interpolation is used to cipher the exact floating- point position values of input features at four regular tried locales in each RoI bin and also summations the result. Figure 6 shows the use of spatial motor and bilinear slice kernel in RoIAlign operation. The ensuing affair, i.e., Target point value at position i in channel c, is attained from the use of a slice kernel from the sample:

$$Q_i^c = \sum_b^H \sum_a^W P_{ba}^c k(x_i^s - a; z_x) k(y_i^s - b; z_y) \forall i \in [1 \ldots H'W'] \forall c \in [1 \ldots C] \tag{5}$$

where x_i^s and y_i^s are sampling coordinates at location i.

After applying the bilinear sampling kernel to the above output, the equation transforms to:

$$Q_i^c = \sum_b^H \sum_a^W P_{ba}^c \max(0, 1 - |x_i^s - a|) \max(0.1 - |y_i^s - b|) \tag{6}$$

where the bilinear slice kernel clones the value at the nearest pixel to (x_i^s, y_i^s) to the affair location (x_i^o, y_i^o). The objective multitasking function, which includes bracket lossL_{cls}, bounding box position loss L_{bbox}, and the segmentation loss for mask L_{mask} is defined. The loss equation can represent this:

$$L = L_{cls} + L_{bbox} + L_{mask} \tag{7}$$

or it can also be written as:

$$L(c, g, b^g, z) = L_{cls}(c, g) + \tau[g \geq 1] L_{loc}(b^g, z) \tag{8}$$

where c is the predicted class, g is GT (ground truth) class, b^g is predicted bounding box for class g, z is the GT bounding box, the classification loss is be defined as:

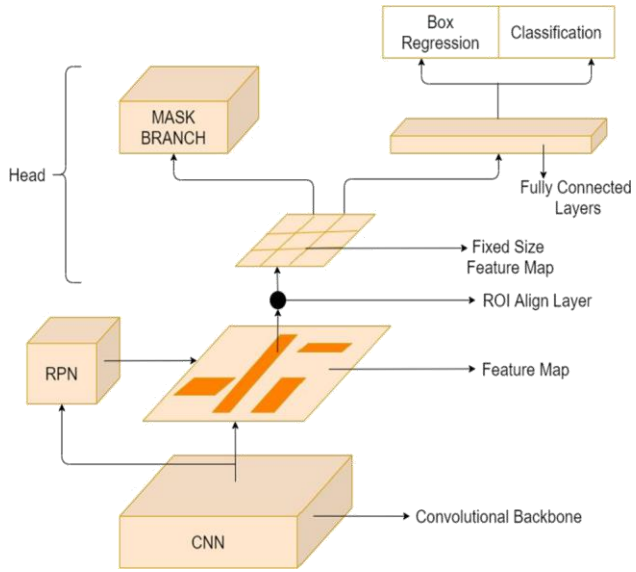


Fig. 4 Architecture of Mask R-CNN

$$L_{cls}(c,g) = -logc_g \tag{9}$$

$$L_{loc}(b^g, z) = \sum_{i\in(x,y,w,h)} smooth_{L1}(b_i^g - z_i) \tag{10}$$

$$smooth_{L1}(a) = \begin{cases} 0.5(a^2) & if|a|<1 \\ |a| - 0.5 & otherwise \end{cases} \tag{11}$$

L_{mask} is the mean binary cross-entropy k (a x a) the sigmoid output helps in pixel-wise binary classification and allows one mask for each class and hence eliminates competition

This definition of L_{mask} allows Mask R-CNN to generate masks for every class without competition between the classes; only the classification branch is used to predict the class label of the output mask. This process disconnects

Fig. 5 RoIAlign Layer

mask and class prediction. Since the considered case includes a per-pixel sigmoid and a binary loss, the masks across classes do not compete and hence provide good instance segmentation results.

4.3.2 Popular DNNs

(a) VGG16- VGG16 is a deep convolutional neural network which is 16 layers deep as its name suggests. It's trained on ImageNet database and takes an input of size 224 x 224. An image is passed through a group of convolutional layers with small open field, i.e., it uses 3 x 3 kernel with stride size 1. It uses three completely connected layers after the convolutional layers to perform bracket. Figure 7 shows the complete armature of VGG16 with all the layers.

(b) InceptionV3- InceptionV3 is made with computational power effectiveness in mind and therefore a smaller number of parameters are generated. It uses ways like factorized complications, lower complications layers, resemblant calculations, reduction in confines, and regularization to make the network more effective. This network is 48 layers deep.

(c) ResNet- 50- The full form of ResNet is Residual Networks. rather of counting on depth of network to learn further features, the residual networks try to lean features from the residual of former subcaste which helps in perfecting delicacy and also helps to break the problem of evaporating grade. ResNet has numerous variants out of which ResNet- 50 has been used which is 50 layers deep.

(d) MobileNetV2- MobileNetV2 farther improves on MobileNetV1 by using feather light depth-wise divisible complications along with direct backups

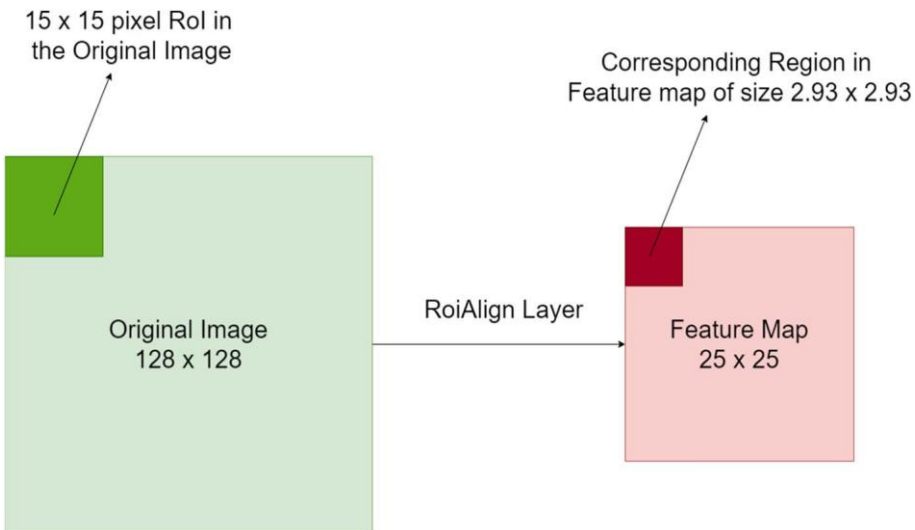
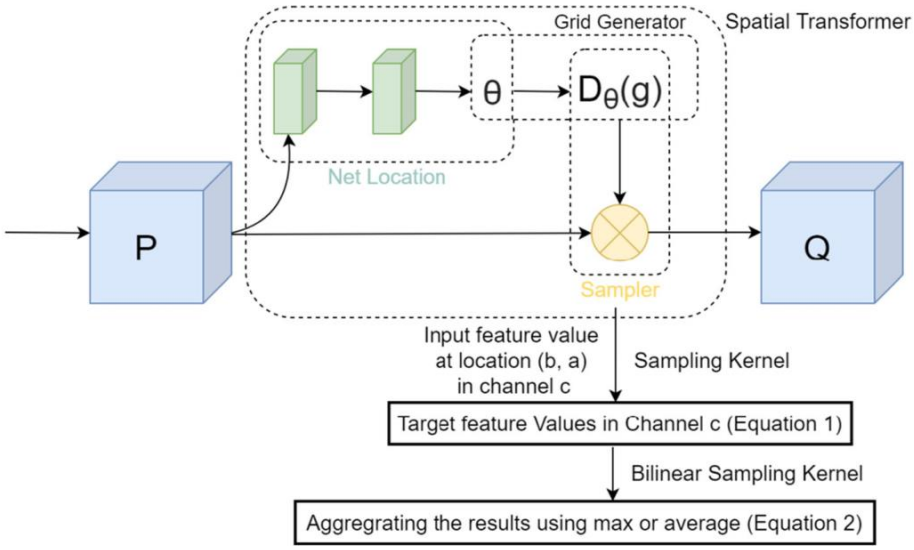


Fig. 6 Use of spatial transformer and bilinear sampling kernel



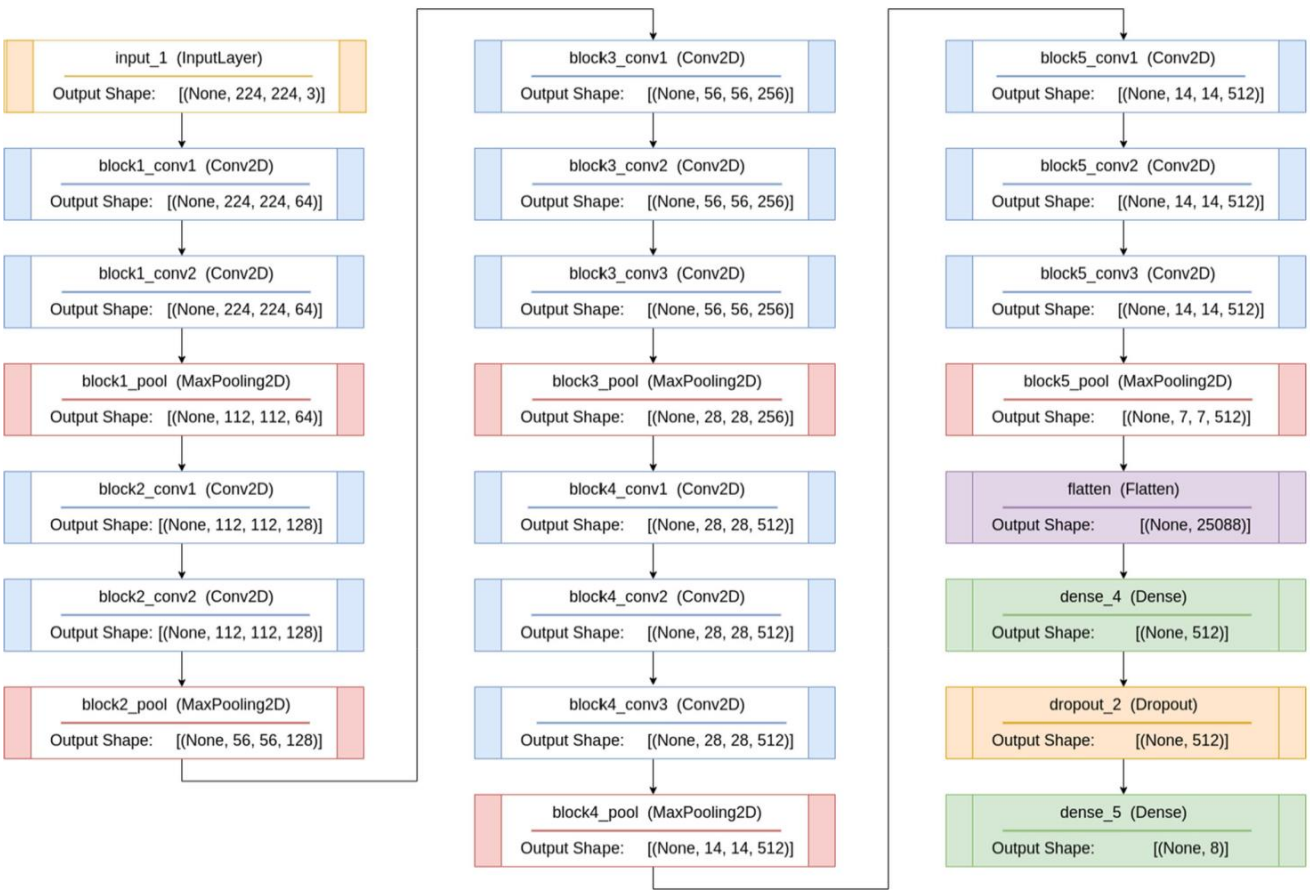


Fig. 7 Layers involved in VGG16

and short connection between these backups. This allows this network to be enforced indeed on mobile bias. It's 53 layers deep and has also been trained on ImageNet database.

Table 1 describes the number of subcaste, total parameters, and trainable parameters for different CNN infrastructures used in this paper, i.e., VGG16, InceptionV3, ResNet50, and MobileNetV2.

Table 1 Summary for CNN architectures used in this paper

Model name	# of Layers	Total parameters	Trainable parameters
VGG16	16	19,175,207	13,863,719
InceptionV3	48	7,317,608	5,688,975
ResNet50	50	20,942,054	16,848,092
MobileNetV2	53	5,440,729	2,804,612

5 Methodology

5.1 Character face detection using mask R-CNN

Facial expressions are the significant contributor to interpersonal communication [43]. In this paper, Mask R-CNN is used for character face discovery, which separates the focus – background pixels from each other using a bounding box that parts the face [44]. The model takes images or vids as input to prize masks of Tom's face or Jerry's face out of it. Algorithm 1 specifies the step-wise methodology espoused for the character face discovery Mask R- CNN, also explained in posterior sections: -

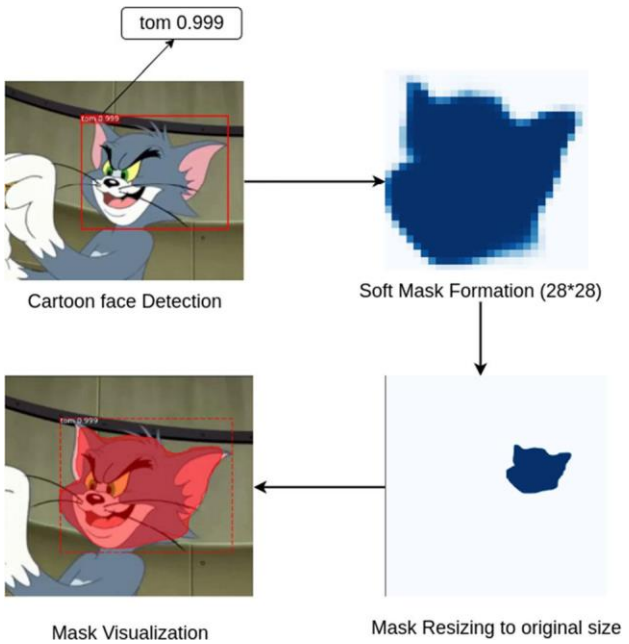


Fig. 8 Mask formation steps

Fig. 9 Masks generated by Mask R-CNN

(i) Frame extraction

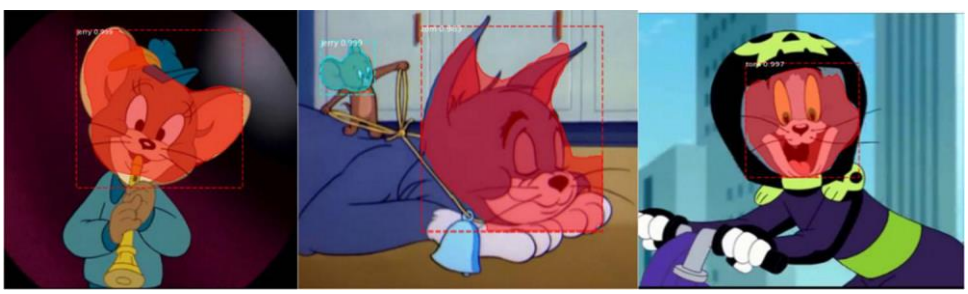
The multi-coloured frames (from vids) are uprooted using OpenCV and are also converted from BGR (Blue, Green, Red) to RGB (Red, Green, Blue) colour order. Frame birth process used then explained in side 4. Mask R-CNN takes input images of same size. Further, all the images are resized to a fixed dimension of (1280 x 1280) and the aspect rate is maintained using padding.

(ii) Mask generation and image storage

Then, the resized image (H' x W') is given to Mask R-CNN as input to detect faces of Tom and Jerry. It then returns a dictionary for each image with four key-value pairs (parameters) concerning the face detected:

(a) boundary box coordinates (y1, x 1, y2, x 2), which are generated around the cartoon's face,
(b) class id for both cartoon characters (1 for Tom and 2 for Jerry),
(c) binary masks.
(d) mask confidence score.

After carrying the bounding boxes and enriching them, the case segmentation model generates the masks for each detected object. The masks are soft masks (with pier pixel values) and of size 28 x 28 during training. Eventually, the prognosticated masks are rescaled to the bounding box confines using padding and scaling factors to induce double masks (0 and 1). Then, 1 denotes the region of the face detected and 0 denotes the rest of the image features. The face detected is marked with a mask confidence score that signifies the confidence in feting the mentioned character. As shown in Fig. 8, the score of detecting Tom's face is 0.999. These masks can be overlaid on the original image to fantasize the final affair. The reused masks generated are shown in Fig. 9. As can be observed from the



figure, the employed Mask R- CNN model detects straight faces, or listed faces or indeed the faces girdled by any object, similar as helmet (ref. Fig. 9).

Using boundary box coordinates or pixel value (y1, x 1, y2, x 2) of the detected face, images ((y2- y1) x (x2- x1)) containing only faces of Tom and Jerry were cropped from the original image. Since the image size of both the original image and double mask was the same, each RGB pixel value of the image is multiplied by the corresponding double value present in the mask. The double value 0, when multiplied with any pixel value, results in 0 (representing black colour). This process removes redundant features from the image. While double value 1 in the mask, when multiplied with any pixel value, results in the same value. Hence, only the pixel values of the face were included in this mask henceforth, will be called as segmented masks). After this, the cropped image and segmented masks are resized into dimension 256 x 256. The cropped images contained redundant features (background regions), whereas segmented masks contained the needed features for emotion bracket.

ALGORITHM 1: Image segmentation using Mask R-CNN

```
This algorithm takes the extracted images
if (videos)
      images = ExtractFrames(videos)
end if.
for each image
      NewImage = Preprocessing(image)
      function Resize(NewImage, MinSize, MaxSize, Padding)
         return ResizedImage, Window, ScalingFactor, Padding
      end function
    function Mask RCNN (ResizedImage)
        return BoundaryBox, ClassID, Score, BinaryMask
   end function
   Dictionary = GetClassID_Score(classID, Score)
   function getSpecificMask(BinaryMask, Dictionary, ResizedImage)
     if detected = Tom
       Mask = GenerateMask(BinaryMask, ResizedImage)
     else if detected = Jerry
       Mask = GenerateMask(BinaryMask, ResizedImage)
    endif
    function Resize(Image,MinSize, MaxSize, Padding)
       resized_imge = cv.resize(img, (256, 256))
    return Mask
end For
function ExtractFrames(videos){ //Frames get extracted at a frame ratio of 1:15
end function
      function Preprocessing(image){
      OpenCV.ImageRead(image)
       OpenCV.convert(image)
return ProcessedImage
function Resize(Image, MinSize, MaxSize, Padding){
   if Padding is true
     Image resized to MaxSize x MaxSize
   else
      Image resized to MinSize x MaxSize
   endif
returns Image
function GetClassID_Score(ClassID, Mask_Confidence_Score){
    Separate ClassID and Score of Tom and Jerry
return ClassID , respective Mask_Confidence_Score
end function
function GenerateMask(Mask_Array, Image){
   for j in range(3):
      Temp[:, :, j] = Temp[:, :, j] * Mask_Array
   end for
 return Image
```

5.2 Emotion classification using transfer learning and fine-tuning

Transfer literacy uses the weights and knowledge gained from working a specific problem and applying that knowledge to break other analogous tasks. It helps in using the weights and impulses of different state- of- the- art algorithms and hence uses it as an advantage without it being necessary to have vast quantities of data or expansive calculation capabilities. The final step includes fine- tuning the model by unfreezing the specific corridor of the model and re-training it on the new data with a small literacy rate. The general channel followed for emotion bracket is as follows:

(a) Preprocessing of segmented masks

Preprocessing of segmented masks the segmented masks (of size 256 x 256) are entered as an affair from character discovery stage using Mask R- CNN (ref. Sect. 5.1). These masks were resized to 224 x 224 for the bracket of feelings using four birth deep neural networks. These images are also converted into tensors. The value of these tensors is in the range 0 to 255, regularized to the range of 0 to 1. subsequently, data batches are created, each having 32 filmland for input into the emotion bracket model. Figure 10 shows an illustration of a data batch created.

(b) Training and Classification

Training and Bracket Four deep neural networks are trained using transfer literacy from the data batches created. Grounded on the results ref. Sect. 6), the best-trained model for every deep neural network is used for the bracket of feelings. A shot of the results attained by this proposed end- to end approach is shown in Fig. 11. For case, 0.96 is an emotion confidence score depicting an angry 'Tom'.

6 Experiments

After applying deep neural network on fairly large data collection, the data go through a preprocessing phase. This phase which uses the Mask R- CNN model produces

Fig. 10 Visualizing the data batch created

Fig. 11 Visualization of emotion recognition using the proposed approach

accurate character masks as per the Algorithm 1 (ref. side5.1). Further, these masks are given as input to four deep neural network models, as described in side4.3. also, the emotion of a particular character is honoured with an emotion confidence score (which is honoured emotion probability of the separate character) gauged in the range of 0 to 1.

6.1 Results

Classifying Feelings on a large dataset is amulti-class bracket problem. In this paper, each sample in the set dataset has been distributed into one of the eight 4 9 2) different classes. Standard criteria — perfection, recall, F1-score, and delicacy score have been reckoned for each class for the purpose of evaluation. The perfection score for ‘happy jerry’ marker is the number of rightly honoured ‘Jerry’ images with happy emotion out of total ‘Jerry’ images with factual happy emotion. From Table 5 given in excursus – I, the number of directly detected ‘happy jerry’ marker is 198 out of 221 total honoured ‘happy jerry’ marker, performing into perfection score of0.90 198/221) for the proposed approach, whereas the perfection score for other three models, i.e., InceptionV3, MobileNetV2, and ResNet- 50 is 0.71, 0.63, and 0.74, independently, for ‘happy jerry’ marker. Next, the recall for ‘happy jerry’ marker is the number of rightly honoured ‘Jerry’ with happy emotion out of the number of factual ‘Jerry’ images with happy emotion. As inferred from Table 5 given in excursus — I, the number of rightly

Table 2 Classification report for all the models with Mask R-CNN

Character	Emotion	VGG16				InceptionV3			
		Precision	Recall	F1-score	Accuracy score	Precision	Recall	F1-score	Accuracy score
Tom	Angry	0.82	0.86	0.84	0.96	0.72	0.75	0.73	0.94
	Happy	0.82	0.94	0.87	0.97	0.76	0.75	0.76	0.94
	Sad	0.87	0.66	0.75	0.95	0.72	0.68	0.7	0.93
	Surprise	0.85	0.85	0.85	0.96	0.76	0.78	0.77	0.94
Jerry	Angry	0.81	0.91	0.86	0.96	0.81	0.8	0.81	0.95
	Happy	0.90	0.87	0.88	0.97	0.71	0.69	0.7	0.91
	Sad	0.88	0.79	0.84	0.96	0.79	0.76	0.78	0.93
	Surprise	0.85	0.87	0.86	0.96	0.72	0.78	0.75	0.93
Micro-average		0.85	0.84	0.84	0.96	0.75	0.75	0.75	0.93
Weighted average		0.85	0.85	0.85	0.96	0.75	0.75	0.75	0.93
Character	Emotion	MobileNetV2				ResNet-50			
		Precision	Recall	F1-score	Accuracy score	Precision	Recall	F1-score	Accuracy score
Tom	Angry	0.71	0.66	0.68	0.93	0.62	0.65	0.64	0.91
	Happy	0.6	0.73	0.66	0.91	0.54	0.77	0.63	0.89
	Sad	0.71	0.52	0.6	0.93	0.46	0.61	0.52	0.88
	Surprise	0.65	0.72	0.68	0.92	0.84	0.37	0.52	0.92
Jerry	Angry	0.66	0.71	0.69	0.92	0.83	0.57	0.67	0.93
	Happy	0.63	0.54	0.58	0.89	0.74	0.64	0.69	0.92
	Sad	0.75	0.69	0.72	0.92	0.64	0.77	0.7	0.90
	Surprise	0.6	0.7	0.65	0.90	0.69	0.7	0.7	0.92
Micro average		0.66	0.66	0.66	0.91	0.64	0.64	0.64	0.90
Weighted average		0.66	0.66	0.66	0.92	0.67	0.64	0.64	0.91

honoured ‘happy jerry’ emotion marker is 198 out of 228 factual ‘happy jerry’ marker. This results into the recall score of (198/228) for the proposed approach as shown in Table 2, whereas the recall score for other three models is comparatively lower for ‘happy jerry’ marker.

For multiclass bracket problem, F1- score is a preferable metric because there may be a large number of factual negatives. It gives the balance between perfection and recall. For case, the F1- score for ‘happy jerry’ is 0.88, for the proposed approach. The other approaches videlicet InceptionV3, MobileNetV2 and ResNet- 50 result in F1-score of 0.75, 0.66, and 0.64, independently, for the same marker. Among these models, VGG16 has outperformed the rest in terms of perfection, recall, and F1- score as shown in Table 2. Also, the table depicts a combined bracket report of the four models on which trial has been conducted. delicacy score for each emotion class then 8) r the two characters taken in this work is also shown in Table 2. The combined delicacy for a particular emotion (say ‘sad’) is calculated by comprising that emotion over both the characters. For illustration, the ‘sad’ emotion delicacy accounts to be 95 which is a normal of two emotion classes (sad for Tom and sad for Jerry). thus, the combined delicacy for each emotion comes out to be Happy [97], Sad [95], Angry [96], and Surprised [96]. Overall scores for each metric are calculated by comprising the scores attained from each class depicted as weighted normal and micro-average. The ultimate is calculated to handle the imbalance in the class distribution if any. VGG16 shows a weighted normal of 0.85 across all criteria as compared to other three models showing the same as 0.75, 0.66, and 0.64, independently, whereas the micro-average comes out to be 0.85 outperforming rest of the models.

Table 3 Classification report for all the models without Mask R-CNN

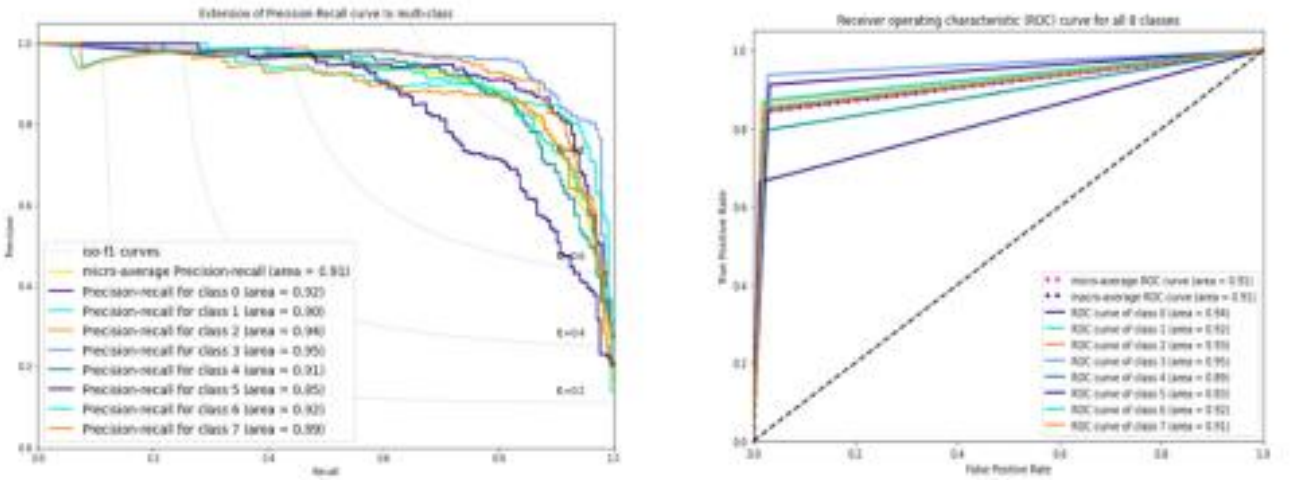
Character	Emotion	VGG16				InceptionV3			
		Precision	Recall	F1-score	Accuracy score	Precision	Recall	F1-score	Accuracy score
Tom	Angry	0.45	0.48	0.46	0.64	0.37	0.35	0.36	0.60
	Happy	0.36	0.37	0.36	0.65	0.35	0.36	0.35	0.59
	Sad	0.38	0.38	0.38	0.65	0.30	0.32	0.30	0.59
	Surprise	0.42	0.46	0.44	0.63	0.35	0.33	0.34	0.61
Jerry	Angry	0.43	0.48	0.45	0.64	0.36	0.35	0.35	0.61
	Happy	0.47	0.47	0.47	0.65	0.36	0.34	0.35	0.60
	Sad	0.45	0.38	0.41	0.63	0.37	0.37	0.37	0.58
	Surprise	0.39	0.45	0.42	0.63	0.32	0.31	0.31	0.61
Micro Average		0.42	0.42	0.42	0.64	0.35	0.34	0.34	0.59
Weighted average		0.42	0.43	0.42	0.64	0.35	0.34	0.34	0.60
Character	Emotion	MobileNetV2				ResNet-50			
		Precision	Recall	F1-score	Accuracy score	Precision	Recall	F1-score	Accuracy score
Tom	Angry	0.28	0.25	0.26	0.58	0.23	0.26	0.24	0.55
	Happy	0.25	0.24	0.24	0.58	0.24	0.25	0.24	0.55
	Sad	0.21	0.20	0.20	0.60	0.23	0.22	0.22	0.56
	Surprise	0.24	0.23	0.23	0.59	0.25	0.25	0.25	0.57
Jerry	Angry	0.26	0.25	0.25	0.59	0.26	0.22	0.24	0.56
	Happy	0.26	0.26	0.26	0.60	0.22	0.21	0.22	0.55
	Sad	0.25	0.27	0.26	0.57	0.24	0.25	0.24	0.58
	Surprise	0.22	0.22	0.22	0.58	0.33	0.27	0.30	0.58
Micro average		0.24	0.24	0.24	0.58	0.25	0.24	0.25	0.56
Weighted average		0.25	0.24	0.25	0.59	0.25	0.24	0.25	0.56

The results of these models without the use of Mask R-CNN are shown in Table 3. Without the use of Mask R-CNN, these models perform inadequately as they're unfit to learn the features of the character's faces which are needed for emotion recognition. This happens because these models learn gratuitous features from the background of the image which aren't needed as the preprocessing stage, i.e., Mask R- CNN isn't used to member the faces.

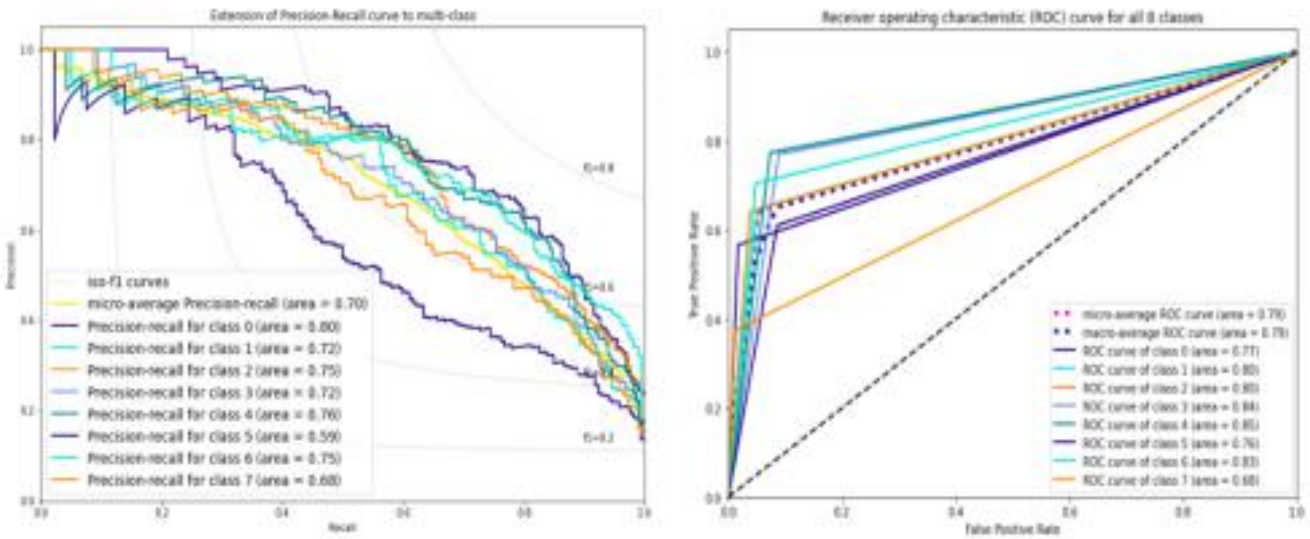
6.2 Extensions

VGG16 outperformed the other three models (ref. Section 6.1) when trained on the created dataset of 8 K labelled images of Tom and Jerry. This section draws out certain fresh results for the best- performing model (VGG16). Figure 12a – d shows the plot for the perfection- recall wind, which depicts the trade- off between perfection and recall for all four DNNs. In an ideal script with high recall and perfection values, a larger area under the wind can be attained. Since the area under the wind is large, indicating high perfection (accurate results) and high recall (maturity positive results). This signifies that both the false-positive rate and false-negative rate are low.

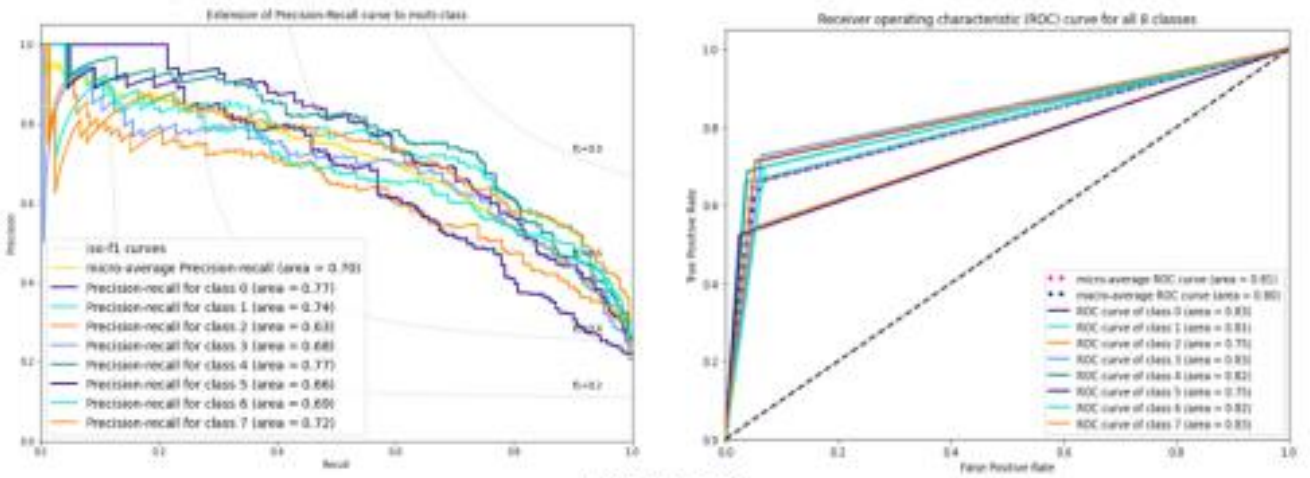
Using VGG16, the micro-average perfection- recall score evaluates to 0.91 for all classes. This score accounts to be the loftiest among all estimated DNNs. The correlation between the false cons and the true cons can be reckoned using the AUC (Area Under wind) in a ROC wind also shown in Fig. 12a – d for all DNNs. As shown, the micro-average ROC score for VGG16 evaluates to 0.91 which exceeds scores attained of the other three DNNs. The figure also shows a macro-average ROC- AUC score for VGG16 that returns the normal without considering the proportion for each marker in the dataset.



(a) VGG-16

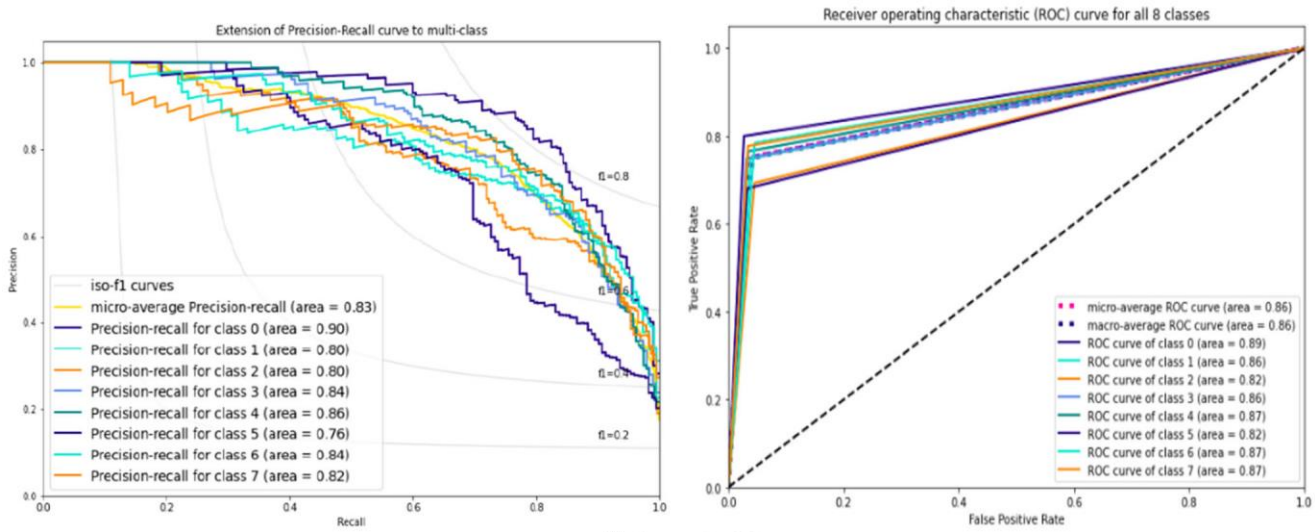


(b) ResNet-50



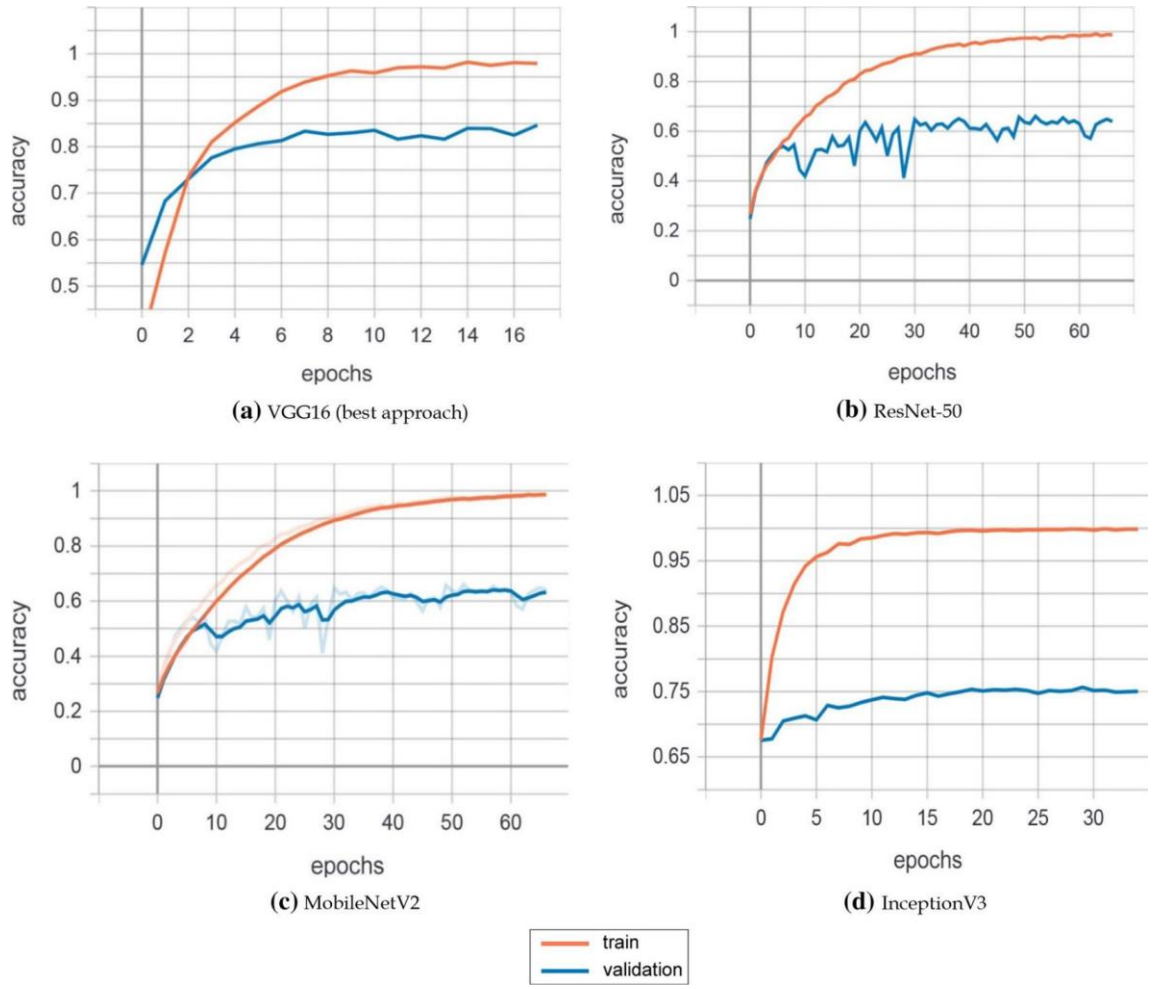
(c) MobileNetV2

Fig. 12 (a–d) Plots for precision-recall curves and ROC curves for all four DNNs



(d) InceptionV3

Fig. 12 continued



(a) VGG16 (best approach)

(b) ResNet-50

(c) MobileNetV2

(d) InceptionV3

Fig. 13 (a–d) Accuracy vs. the number of epochs plots for all the models

Table 4 Comparison between the proposed approach and the existing methods

Comparison Parameters	Proposed Approach	Hill [29]	Li et al. [45]	Ma et al. [46]	Aneja et al. [33]	Aneja et al. [47]
Number of emotions recognized	4 (H, S, A, & Su)	3 (H, S & A)	4 (A, H, S, & N)	6 (H, S, A, F, Su, & D)	7 (S, A, Su, D, F, N, & J)	7 (S, A, Su, D, F, N, & J)
Learning rate	0.0003	0.001	–	–	0.01	0.0001
Accuracy Score for Individual Emotion	0.97 (H), 0.95 (S), 0.96 (A), 0.96 (Su)	–	0.77 (A), 0.65 (H), 0.70 (S)	0.78 (H), 0.49 (S), 0.62 (A), 0.19 (Su)	0.89 (S), 0.85 (A), 0.95 (Su)	0.79 (S), 0.90 (A), 0.94 (Su)
Overall Model Accuracy Score	0.96	0.80	–	–	0.89	–
F1-score	0.85	–	–	–	–	–

*H → Happy, S → Sad, A → Angry, Su → Surprised, F → Fear, D → Disgusted, J → Joy, N → Neutral

A model is said to be a good fit if it's suitable to generalize and learn the features from the training data without overfitting or under- fitting. This means that the model can generalize the features and perform indeed on unseen data. In 13b and c, the models videlicet ResNet- 50 and MobileNetV2 are unfit to generalize the features. Hence, indeed if they perform well on training data with delicacy nearing 1, they're unfit to give analogous results on unseen data as can been seen from the change in the graph for confirmation delicacy wind. The proposed model depicted in 13a is suitable to outperform the below-pronounced models and is suitable to generalize the results giving an average delicacy score of 0.96. As shown in Fig. 13d, the InceptionV3 model is unfit to learn enough features from training data as compared to our proposed model and therefore gives lower scores for all the evaluation criteria. The figure also shows the train and confirmation rigor recorded and colluded using tensor-board from the logs saved while training.

6.3 Comparison

Inspired by the fact that DNN models bear fairly large quantum of data, in our study, we've created a dataset (source: https://github.com/emotionRecog). A fair comparison isn't possible since we've created a new dataset for our proposed tool, which we call, integrated DNN. still, indeed though datasets and number of feelings are varied in the literature, we find it intriguing to report former workshop that were concentrated on emotion bracket. In what follows, let us readdress former workshop (ref. Table 4) and check how fair the comparison can be made.

In In the presented comparison, one of the living works has proposed an emotion recognition model on mortal animated faces [33]. The espoused methodology includes training two different CNNs to fete mortal expressions and stylized characters expressions singly.

A participated embedding point space is also created by mapping mortal faces to character faces using transfer literacy, performing in a delicacy score of 0.89. also, other benefactions [45, 46] have estimated the proposed approaches on amped characters (not specific to a cartoon character) from colourful sources like books, videotape games, etc.

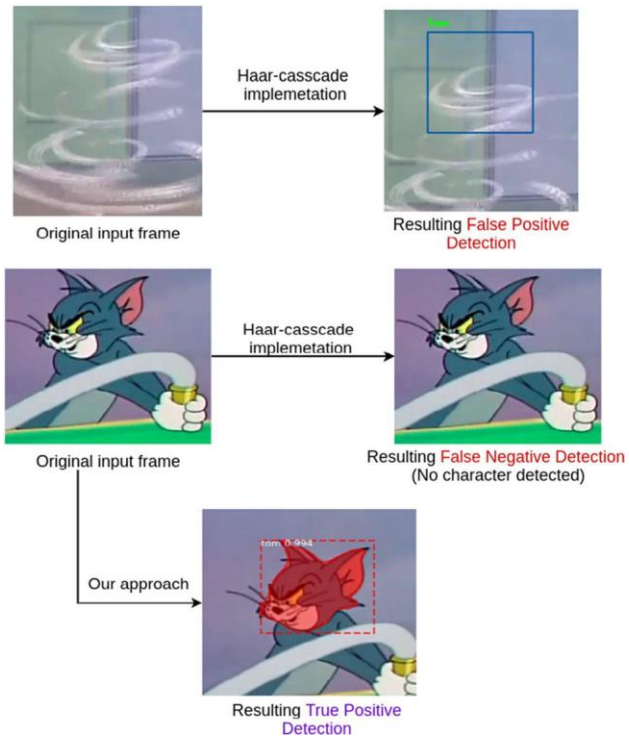


Fig. 14 Comparing our approach in detecting a cartoon character to already existing work

A different kind of approach generates 3D animated faces using mortal facial expressions by transferring the emotion features to the 3D character face using semi-supervised literacy model ‘ExprGen’ [47]. The above- appertained state- of- the- art styles Li et al. [45], Ma et al. [46], and Aneja et al. [47] — do not give their overall end- to- end model delicacy score and rather, present a delicacy score for each emotion marker classified as shown in Table 4. Hill [29] is the only analogous donation that exists in the state- of- the- art styles where the author has proposed an end- to- end emotion recognition model on cartoon vids. This approach gives an overall delicacy score of0.80. Figure 14 depicts the outperformance of the intertwined DNN model (contributed in this paper) using Mask R-CNN over the formerly being methodology [29]. As mentioned before, indeed though datasets are varied from one work to another, we find that our study (with integrated DNN tool) performs better on the dataset of size 8113.

7 Conclusion

Feting feelings from facial expressions of faces other than mortal beings are an intriguing and gruelling problem. Although the being literature has tried to descry and fete objects, still, feting feelings has not been considerably covered. thus, in this paper, we've presented an intertwined Deep Neural Network (DNN) approach that has successfully honoured feelings from cartoon images. We've collected a dataset of size 8 K from two cartoon characters ‘Tom’ & ‘Jerry’ with four different feelings, videlicet happy, sad, angry, and surprise. The proposed integrated DNN approach has been trained on the large dataset and has rightly linked the character, segmented their face masks, and honoured the consequent feelings with a delicacy score of. The approach has employed Mask R- CNN for character discovery and state- of- the- art deep literacy models, videlicet ResNet- 50, MobileNetV2, InceptionV3, and VGG 16, for emotion bracket. The experimental analysis has depicted the outperformance of VGG 16 over others with a delicacy of 96 and F1 score of0.85. The proposed intertwined DNN has also outperformed the state- of- the- art approaches.

The work would be salutary to the animators, illustrators, and cartoonists. It can also be used to make a recommender system that allows druggies to associatively elect emotion and cartoon brace. Studying feelings boxed in cartoons also excerpts other confederated information, which if combined with artificial intelligence can open a plethora of openings, for case, feting feelings from body gestures.

Appendix

See Table 5 .

Table 5 Confusion matrix for all the models

		VGG16								Total Actual Labels	InceptionV3								Total Actual Labels
Character		Jerry Predicted labels				Tom Predicted labels					Jerry Predicted labels				Tom Predicted labels				
	Emotion	**Angry**	**Happy**	**Sad**	**Surprise**	**Angry**	**Happy**	**Sad**	**Surprise**		**Angry**	**Happy**	**Sad**	**Surprise**	**Angry**	**Happy**	**Sad**	**Surprise**	
Jerry Actual labels	**Angry**	187	6	8	3	1	0	0	0	205	164	14	19	8	0	0	0	0	205
	Happy	12	198	3	15	0	0	0	0	228	14	158	16	39	0	1	0	0	228
	Sad	27	5	186	15	0	0	1	0	234	15	22	179	16	1	1	0	0	234
	Surprise	4	12	11	186	0	0	0	0	213	8	26	12	167	0	0	0	0	213
Tom Actual labels	**Angry**	1	0	2	0	160	17	4	3	187	0	0	0	1	140	18	14	14	187
	Happy	0	0	0	0	10	179	2	0	191	1	0	0	0	23	143	12	12	191
	Sad	1	0	0	0	14	16	114	27	172	0	0	0	0	15	18	117	22	172
	Surprise	0	0	1	0	11	7	10	164	193	0	1	0	0	16	6	20	150	193
Total Predicted labels		232	221	211	219	196	219	131	194		202	221	226	231	195	187	163	198	

		MobilenetV2								Total Actual Labels	Resnet_50								Total Actual Labels
Character		Jerry Predicted labels				Tom Predicted labels					Jerry Predicted labels				Tom Predicted labels				
	Emotion	**Angry**	**Happy**	**Sad**	**Surprise**	**Angry**	**Happy**	**Sad**	**Surprise**		**Angry**	**Happy**	**Sad**	**Surprise**	**Angry**	**Happy**	**Sad**	**Surprise**	
Jerry Actual labels	**Angry**	146	22	22	14	1	0	0	0	205	116	12	52	11	9	2	3	0	205
	Happy	30	124	14	60	0	0	0	0	228	6	146	32	32	2	5	5	0	228
	Sad	31	18	161	24	0	0	0	0	234	8	13	181	21	4	2	5	0	234
	Surprise	14	32	18	149	0	0	0	0	213	6	26	18	150	4	5	3	1	213
Tom Actual labels	**Angry**	0	1	0	0	123	42	8	13	187	1	0	0	1	121	31	29	4	187
	Happy	0	0	0	0	18	139	10	24	191	1	0	0	0	22	147	18	3	191
	Sad	0	0	0	0	14	30	90	38	172	1	0	0	1	18	41	105	6	172
	Surprise	0	0	0	0	17	20	18	138	193	1	1	0	2	14	41	62	72	193
Total Predicted labels		221	197	215	247	173	231	126	213		140	198	283	218	194	274	230	86	

Authors' contributions All authors have equally contributed toward the formation of this paper.

Funding Not Applicable.

Declarations

Conflicts of interest the authors declare that they have no competing interests.

Availability of data and material
https://github.com/emotionRecog.

References

1. Ekman P, Friesen WV (1976) Measuring facial movement. Environ psychol nonverbal behav 1(1):56–75
2. Shivhare S. N., Khethawat S. (2012). Emotion detection from text. arXiv preprint. arXiv:1205.4944
3. Gupta V, Singh VK, Mukhija P, Ghose U (2019) Aspect-based sentiment analysis of mobile reviews. J Intell Fuzzy Syst 36(5):4721–4730
4. Piryani R, Gupta V, Singh VK (2017) Movie Prism: A novel system for aspect level sentiment profiling of movies. J Intell Fuzzy Syst 32(5):3297–3311
5. Rao Y, Xie H, Li J, Jin F, Wang FL, Li Q (2016) Social emotion classification of short text via topic-level maximum entropy model. Inf Manag 53(8):978–986
6. Venkataramanan K., Rajamohan H. R. (2019). Emotion Recog-nition from Speech. arXiv preprint, arXiv:1912.10458
7. Gupta V, Juyal S, Singh GP, Killa C, Gupta N (2020) Emotion recognition of audio/speech data using deep learning approaches. J Inf Optim Sci 41(6):1309–1317
8. Casale S., Russo A., Scebba G., Serrano S. (2008). Speech emotion classification using machine learning algorithms. In: 2008 IEEE international conference on semantic computing, pp. 158–165
9. Jiang D. N., Cai L. H. (2004). Speech emotion classification with the combination of statistic features and temporal features. In: IEEE International Conference on Multimedia and Expo, Vol. 3, pp 1967-1970
10. Kim, M. H., Joo, Y. H., Park, J. B. (2005). Emotion detection algorithm using frontal face image. International Conference on Control and Robotics Systems, 2373–2378.
11. Bargal S. A., Barsoum E., Ferrer C. C., Zhang C. (2016). Emotion recognition in the wild from videos using images. In: Proceedings of the 18th ACM International Conference on Multimodal Interaction. pp 433–436.
12. Elngar, A. A., Jain, N., Sharma, D., Negi, H., Trehan, A., & Srivastava, A. (2020). A deep learning based analysis of the big five personality traits from handwriting samples using image processing. Journal of Information Technology Management, 12(Special Issue: Deep Learning for Visual Information Analyt-ics and Management.), 3–35.
13. Guo Y., Gao H. (2006). Emotion recognition system in images based on fuzzy neural network and HMM. In: 5th IEEE Inter-national Conference on Cognitive Informatics, Vol. 1, pp 73–78.
14. Lisetti C, Nasoz F, LeRouge C, Ozyer O, Alvarez K (2003) Developing multimodal intelligent affective interfaces for tele-home health care. Int J Hum Comput Stud 59(1–2):245–255
15. Gupta V, Jain N, Katariya P, Kumar A, Mohan S, Ahmadian A, Ferrara M (2021) An emotion care model using multimodal
16. Derntl B, Seidel EM, Kryspin-Exner I, Hasmann A, Dobmeier M (2009) Facial emotion recognition in patients with bipolar I and bipolar II disorder. Br J Clin Psychol 48(4):363–375
17. Jain R, Jain N, Aggarwal A, Hemanth DJ (2019) Convolutional neural network based Alzheimer's disease classification from magnetic resonance brain images. Cogn Syst Res 57:147–159
18. Jain, N., Chauhan, A., Tripathi, P., Moosa, S. B., Aggarwal, P., & Oznacar, B. (2020). Cell image analysis for malaria detection using deep convolutional network. Intelligent Decision Tech-nologies, (Preprint), 1–11.
19. Bahreini K, Nadolski R, Westera W (2016) Data fusion for real-time multimodal emotion recognition through webcams and microphones in e-learning. Int J Hum-Comput Interact 32(5):415–430
20. Ray A., & Chakrabarti A. (2012). Design and implementation of affective e-learning strategy based on facial emotion recognition. In: Proceedings of the International Conference on Information Systems Design and Intelligent Applications, pp 613–622.
21. Chu HC, Tsai WWJ, Liao MJ, Chen YM (2018) Facial emotion recognition with transition detection for students with high-functioning autism in adaptive e-learning. Soft Comput 22(9):2973–2999
22. Shen L, Wang M, Shen R (2009) Affective e-learning: Using ''emotional'' data to improve learning in pervasive learning environment. J Educ Technol Soc 12(2):176–189
23. Piryani R, Gupta V, Singh VK, Ghose U (2017) A linguistic rule-based approach for aspect-level sentiment analysis of movie reviews. In: Bhatia SK, Mishra KK, Tiwari S, Singh VK (eds) Advances in computer and computational sciences. Springer, Singapore, pp 201–209
24. Ren F, Quan C (2012) Linguistic-based emotion analysis and recognition for measuring consumer satisfaction: an application of affective computing. Inf Technol Manage 13(4):321–332
25. Piryani R, Gupta V, Singh VK (2018) Generating aspect-based extractive opinion summary: drawing inferences from social media texts. Computacio´n y Sistemas 22(1):83–91
26. Garbas J. U., Ruf T., Unfried M., Dieckmann A. (2013). Towards robust real-time valence recognition from facial expressions for market research applications. In: 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, (pp 570–575).
27. Robinson L, Spencer MD, Thomson LD, Sprengelmeyer R, Owens DG, Stanfield AC, Johnstone EC (2012) Facial emotion recognition in Scottish prisoners. Int J Law Psychiatry 35(1):57–61
28. Peleshko, Dmytro, Kateryna Soroka. (2013). Research of usage of Haar-like features and AdaBoost algorithm in Viola-Jones method of object detection. International Conference on the Experience of Designing and Application of CAD Systems in Microelectronics.

29. Hill JW (2017) Deep Learning for Emotion Recognition in Cartoons(Unpublished master's dissertation). The University of Lincoln, Lincoln School of Computer Science, UK
30. Ekman P, Oster H (1979) Facial expressions of emotion. Annu Rev Psychol 30(1):527–554
31. Gajarla V, Gupta A (2015) Emotion detection and sentiment analysis of images. Georgia Institute of Technology, Atlanta
32. Minaee, S., & Abdolrashidi, A. (2019). Deep-emotion: Facial expression recognition using attentional convolutional network. arXiv preprint, arXiv:1902.01019
33. Aneja D., Colburn A., Faigin G., Shapiro L., Mones B. (2016). Modeling stylized character expressions via deep learning. In: Asian conference on computer vision, pp 136–153
34. Zhao J, Meng Q, An L, Wang Y (2019) An event-related potential comparison of facial expression processing between cartoon and real faces. PLoS ONE 14(1):e0198868
35. Kendall LN, Raffaelli Q, Kingstone A, Todd RM (2016) Iconic faces are not real faces: enhanced emotion detection and altered neural processing as faces become more iconic. Cognitive Res: Princ Implic 1(1):19
36. Li, S., Zheng, Y., Lu, X., & Peng, B. (2019). iCartoonFace: A Benchmark of Cartoon Person Recognition. arXiv preprint, arXiv:1907.13394
37. Zhou Y., Jin Y., Luo A., Chan S., Xiao X., Yang X. (2018). ToonNet: a cartoon image dataset and a DNN-based semantic classification system. In: Proceedings of the ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry, pp. 1–8.
38. Cowie R, Douglas-Cowie E, Tsapatsoulis N, Votsis G, Kollias S, Fellenz W, Taylor JG (2001) Emotion recognition in human-computer interaction. IEEE Signal Process Mag 18(1):32–80
39. Xu C, Cui Y, Zhang Y, Gao P, Xu J (2020) Person-independent facial expression recognition method based on improved Wasserstein generative adversarial networks in combination with identity aware. Multimedia Syst 26(1):53–61
40. Siddiqi MH, Ali R, Khan AM, Kim ES, Kim GJ, Lee S (2015) Facial expression recognition using active contour-based face detection, facial movement-based feature extraction, and non-linear feature selection. Multimedia Syst 21(6):541–555
41. Rolling L (1981) Indexing consistency, quality and efficiency. Inf Process Manage 17(2):69–76
42. Byrt T (1996) How good is that agreement? Epidemiol 7(5):561
43. Pantic M, Rothkrantz LJM (2000) Automatic analysis of facial expressions: the state of the art. IEEE Trans Pattern Anal Mach Intell 22(12):1424–1445
44. Lin K, Zhao H, Lv J, Li C, Liu X, Chen R, Zhao R (2020) Face Detection and Segmentation Based on Improved Mask R-CNN. Discrete Dyn Nat Soc 2020:1–11
45. Li Y., Yu F., Xu Y. Q., Chang E., Shum H. Y. (2001). Speech-driven cartoon animation with emotions. In: Proceedings of the ninth ACM international conference on Multimedia, pp 365–371.
46. Ma X., Forlizzi J., Dow S. (2012). Guidelines for depicting emotions in storyboard scenarios. In: International design and emotion conference.
47. Aneja D., Chaudhuri B., Colburn A., Faigin G., Shapiro L., Mones B. (2018). Learning to generate 3D stylized character expressions from humans. In: IEEE Winter Conference on Applications of Computer Vision, pp 160–169.