International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 42, No. 3, June, 2025, pp. 1367-1381
Journal homepage: http://iieta.org/journals/ts

Expressive Cartoon Card (ECC) Generation Using Deep Learning for Privacy Preserving Social Media Posts

Vasantha Sandhya Venu[1], Shirina Samreen[2*]

[1] Department of Computer Science and Engineering, Vardhaman College of Engineering, Hyderabad 501218, India
[2] Department of Computer Science, College of Computer and Information Sciences, Majmaah University, Al Majmaah 15341, Saudi Arabia

Corresponding Author Email: s.samreen@mu.edu.sa

https://doi.org/10.18280/ts.420313

Received: 22 August 2024
Revised: 7 January 2025
Accepted: 19 April 2025
Available online: 30 June 2025

Keywords:
anime, cartoon images, deep learning, privacy preserving, social media posts

ABSTRACT

Nowadays, it is common to express one's feelings through a quote or an image conveying emotion. Due to ever-increasing cybercrimes, people use cartoonized or anime images to preserve privacy before posting or displaying their profile pictures on social media. We present a novel expressing approach which facilitates more effective communication on social media while preserving privacy through the generation of anime-based personal images expressing emotion with tagged captions. The proposed system is designed to generate personalized Expressive Cartoon Card (ECC) for individuals employing customized AnimeGANv2, DNN, and smallGPT-2 models. A test dataset of colour images and another dataset for quotes are constructed from other large datasets. The effectiveness of visual quality and similarity of anime and real photos is evaluated using PSNR, SSIM and FID metrics. For the three anime styles, the average FID scores are 24.74, 34.39 and 49.08, and overall accuracies of anime emotion detection are 0.97, 0.94, and 0.94. GPT2 performance is analyzed using training and validation loss curves, which resulted in 1.0 and 1.9 at 30 epochs demonstrating its apt for quote generation for the detected emotion.

1. INTRODUCTION

Preserving privacy is crucial when it comes to an individual's photo on social media. Social media platforms have become a common way for people to share their lives with friends and family; nevertheless, it's crucial to consider any potential risks that might be connected to sharing personal photos online. Posting a photo online can make you vulnerable to cyberbullying, harassment, and stalking.

Anime is a style of Japanese animated entertainment that typically features colorful and highly stylized animation, distinct character designs, and complex storylines. It has gained widespread popularity around the world, especially among younger audiences, due to its unique style, rich storytelling, and diverse themes. Anime is more often used for entertainment. To protect your privacy, it's essential to be careful about what you post on social media. Many people choose to have anime profile pictures on social media as they can maintain their privacy and remain anonymous. This can be particularly appealing for those who do not want to reveal their real identity or appearance online. Many people simply enjoy the art style of anime and find the colorful and expressive characters to be visually appealing, and also using an anime profile picture can signal to others their interests. This can help to create a sense of community and connection with like-minded individuals. Thus, the use of an anime profile picture can be a way to express oneself, have fun online and retain privacy.

Profile photos can be a powerful tool for expressing emotions and communicating with others online. People often choose profile photos that reflect their current mood or emotions, or that convey a particular message to the audience. For example, someone who is feeling happy and upbeat may choose a profile photo that shows them smiling or laughing. This can communicate to others that they are approachable and friendly, and may help them to connect with others who share their positive outlook. Similarly, someone who is feeling sad or upset may choose a profile photo that reflects their current mood, such as a photo of them looking down or with a sober expression. This can communicate to others that they may be going through a difficult time and may need support or understanding.

Quotes can be a powerful tool for self-expression, as they serve as a way of conveying a particular mood or emotion. For example, someone who is feeling down or going through a difficult time may use a quote that reflects their feelings of sadness or despair. Alternatively, someone who is feeling optimistic or inspired may use a quote that reflects their sense of hope and positivity. So, the use of quotes in profiles can be a meaningful way for people to express themselves and connect with others who share similar feelings or interests.

Other ways to express oneself include emojis, animal or avatar stickers, comedian photos with expressions, and memes. However, none feel as natural as physical expressions, where a person's facial expressions and message convey emotions directly. Hence, there is a need for a more effective way of communication on social media while preserving privacy.

Existing solutions have focused on various approaches. Some authors have developed GAN-based models for cartoonizing photos, primarily creating cartoon or anime-style images without emphasizing facial expressions. Others have explored emotion recognition using CNN models and vision transformers to assess a person's mental state based on facial expressions in different contexts. Additionally, researchers have employed Recurrent Neural Networks (RNNs) and Transformers, like GPT, to generate quotes tailored to a user's mood or preferences by learning patterns in text data. This paper introduces a novel technique, which integrates three different algorithms for more personalized way and privacy preserving which none of the authors have worked on. First, Enhanced Cartoonization algorithm to retain expressions of the original photo while cartoonization. Second, emotion recognition and validation algorithm to validate generated cartoons comparing them with original photo's facial expression while other authors compared cartoon images with original to check for similarity of facial features. Third, Quote generation algorithm gives a suitable quote as caption to the cartoon card based on the emotion identified from the cartoon image.

2. LITERATURE SURVEY

A cartoonized selfie Generative Adversarial Network (scGAN), which emphasize certain facial regions while ignoring minor details, primarily employs an attentive-based adversarial network. It creates a cycle-like architecture, to be more precise. In cartoon portraits, a total variation loss is utilized to draw attention to key edges and contents. In order to place more emphasis on delicate facial features like the eyes, a cycle loss is added, and to increase the method's resilience while removing artifacts, a perceptual loss is added [1]. There are two ways to transform images of actual people into anime. The first technique uses a completely new dataset to train photo-to-cartoon algorithms. The second technique breaks down cartoonization into three steps: line-drawing the original image, adding color clues to the line-drawing, and finally coloring. A lighter approach called LWAnimeGAN was created in order to train more quickly, and that takes up less space [2]. Dong et al. [3] have designed the architecture for the cartoon loss function. It can mimic both the drawing and colouring processes so that users can learn how to create cartoon images with smooth surfaces and vibrant colours. In addition, it applies an initialization method to facilitate and stabilize the model training in case of reusing the discriminator [3]. A pixelation technique based on art style is implemented to retain the integrity and essential elements of the image in artistic perspective. This technique also makes it possible to generate high-quality pixels of any size without the need for special datasets. The AOP algorithm produces pixel images that are more aesthetically pleasing than those produced by other algorithms and instruments [4]. In both grayscale and colorful photos, various techniques for creating a cartoonized painterly impression were presented. The painterly look on photos is achieved using the vector quantization concept. LBG, KPE, and KMCG algorithms are utilized to produce cartoonized, artistic outcomes. Based on the amount of time required and the effects produced, the outcomes of applying each algorithm to photographs are contrasted. The outcomes of the aforementioned techniques can be employed in a variety of digital art tools and movie to comic conversion applications [5]. MS-CartoonGAN is a new shift in GAN architecture, which produces cartoons in different styles. In order to accomplish this, three different types of losses such as threshold adversarial, style loss and level oriented semantic loss [6]. A useful and complex method for computer vision as well as computer graphics is one that turns pictures of real-world settings into cartoon images. This approach is built on traditional image processing techniques, which have lately neglected to stylize images in creative forms like paintings as deep learning has gained popularity. Yet, instructional approaches result in cartoonization by requiring a lot of time, energy, and data [7]. The significance of pre and post processing tasks for better cartooning of videos is highlighted by the findings of comprehensive tests followed by qualitative user evaluation [8]. Chen et al. [9] used cartoon pictures and mismatched photographs. They used two new losses related to stylistic variation, semantic content and edge adversarial. The processing time for GAN model's initial implementation on a single threaded task is almost two times as long for producing cartoon copies. Transfer learning reduced model implementation from 32-bit float to 8-bit int data [10]. To distinguish between three different white-box representations of images in GAN, texture, surface, and structure representations are employed for cartooning [11]. Cartoon-like abstracted pictures were produced by combining techniques for edge recognition, smoothing/segmentation of region, color harmonization, and brightness/saturation correction. A subjective assessment was done using the results of relevant work. Additionally, a trial research with human judges was carried out in which the suggested method was contrasted with five already available image cartooning tools [12]. An artificial two-sub system structure for the recognition of complicated emotions was developed. The design mirrors how the human brain responds to dilemmas and makes decisions [13]. Mou et al. [14] presented a fusion of hybrid attention and ConvLSTM for multimodal inputs to model the recognition of driver emotion. Comprehensive trials were conducted on driving simulator generated data, to confirm the efficacy of the suggested strategy. A model designed for detecting emotions is employed in suspect interrogation, as a tool to aid persons with nerve injury to recognize emotions, and Covid patient timelines. Comparisons were made between the outputs of CNN variants and Vision Transformers, which worked on AffectNet, Expanded CK+, Tsinghua, RAF, and KDEF datasets [15]. The Approach developed by Sakai et al. [16], suggests a two-phase generative framework to replicate the expression of a patient on some public face. In this technique, face swapping and the expression of the patient by using an upgraded Cycle GAN. Then, a realistic virtual patient face was created. Understanding epistemic emotions is crucial for enhancing the effectiveness of teaching and learning since they are linked to knowledge creation and academic success. Epistemic emotions are difficult to identify since they are often not overtly exhibited on the outside. Algorithms were trained to assess six epistemic emotions using heart rate variability (HRV) and video data (Bore, Confuse, Curious, Frustration, Interest and Surprise) [17].

A database of multimodal face expressions that include both thermal infrared and naturally occurring visible videos was created. It not only updates additional heat infrared data but also improves the data on emotional intensity. It classifies each feeling into one of three class levels. Thirty individuals' spontaneous seven emotions were recorded in the database. During the experiment, emotions were elicited using both audio and visual cues. Additionally, the constructed database is examined utilizing cutting-edge models including YOLO, ResNet50, and other hybrid models [18]. A toolkit is presented to identify common emotions and two new states on data representing various racial and ethnic groupings [19]. The research paper [20] offers the Relation-aware Network (RANet) for classifying facial expressions. To build associations between spatial positions and channels, RANet includes two relational attention modules. Global relationships assist RANet to focus on discriminative face regions. To efficiently compute spatial attention, the separable convolution has been used. A fresh label distribution learning technique that takes uncertainty into account to increase the robustness of deep models to handle ambiguity and uncertainty. In order to create emotion distributions for training samples in an adaptable manner, it makes use of neighborhood information in the valence-arousal space. By incorporating provided labels into the label distributions, it also takes their ambiguity into account. To get more training supervision and boost recognition accuracy, it is simply integrated into a deep network [21].

3. PROPOSED METHOD

Cartoonized images have been widely used for profile pictures over social media to preserve privacy and on the other side emotion-based quotes are being displayed in profile pictures to express one's feelings to others. The proposed system blends these two features to come up with a new appealing technique that facilitates an individual to express their feelings in a more attractive and realistic way. It comprises three main modules: Photo cartoonization, Emotion Recognition and Validation, and Quoted Cartoon Card Creation.

Initially, the user captures his/her photo, this photo is inputted to the photo cartoonization module to generate an anime picture and also given to an emotion recognizer to find out the emotion of the person in the captured photo. For anime picture generation, a pre-trained animeGANv2 model is employed as it works well on real-world images producing a photo with better visual quality. This pre-trained model supports three different styles of anime: Miyazaki Hayao, Kon Satoshi (Paprika) and Makoto Shinkai. The user selects one of these three styles for anime picture generation. The Emotion Recognition and Validation module includes the model designed using Deep Neural Network (DNN) for Emotion Recognition and a comparator for emotion validation. Generated anime is passed as input to an emotion recognizer which identifies emotion. As a part of the validation, the emotion of the anime picture and the original picture are both compared. If the emotion identified in both the pictures is the same then it is passed to the next module, i.e., Quote Generation otherwise the user needs to recapture his/her picture in a better lighting conditions and start the process again. Randomly, five quotes related to detected emotion are generated using a small GPT-2 model. Next five cartoon cards are constructed for the given input image using an anime photo, each with a different quote as a caption. The proposed ECC generation system is described in Figure 1.

3.1 Description of datasets

Datasets involved in implementing the proposed system fall into three categories: a) Anime style datasets, b) Test photo dataset, and c) Quote dataset. Three different anime style datasets, which are mentioned in Table 1 are employed by the AnimeGANv2 pretrained model [22]. For testing this system, a colorful test photo dataset is created covering five emotions such as neutral, sad, angry, happy, and fear from various other datasets whose details are mentioned in Table 2. To construct a test photo dataset diverse enough to represent various facial expressions across genders and age groups, photos were extracted from three different datasets to ensure balance.

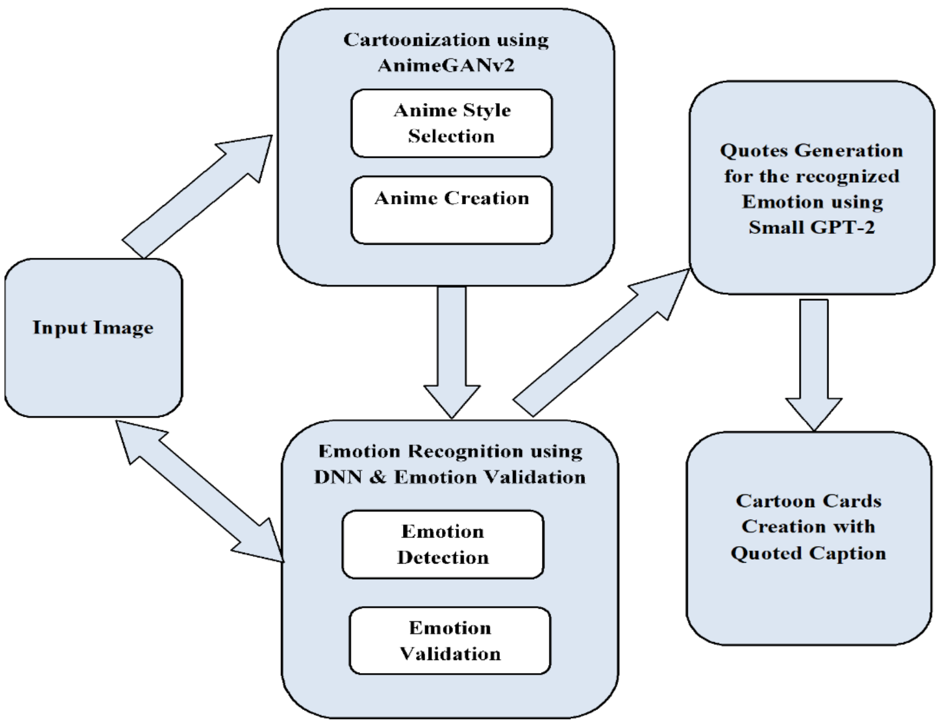

Figure 1. Basic architecture of the proposed ECC generation system

Table 1. Anime style datasets details

Style of Anime	Count of Pictures
Miyazaki Hayao from “The Wind Rises” film	1752
Makoto Shinkai from “Your Name & Weathering with you” film	1445
Kon Satoshi from “Paprika” film	1284

Table 2. Test photo datasets details

Dataset Details	Total Faces Count	Extracted Test Images Count
V7Lab AI Generated Person Faces[23]	100000	99
Human Faces from Kaggle [24]	7219	74
Flickr Faces (FF-HQ) [25]	52000	20
Facial age dataset [26]	9978	15

The majority of these were sourced from V7Labs due to their superior quality. Photos of individuals aged above 5 years and below 50 years were considered, assuming they are more active on social media. A Quotes dataset is constructed that includes only those quote categories that are relevant for recognized emotions [27-30]. Its details are mentioned in Table 3.

Table 3. Quotes dat27aset details

Keywords/tags	Emotion	Extracted Count of Quotes	Source
Inspirational, life, motivational, dreams, hope, peace	Neutral	15560	Goodreads quotes [28] Quotes-500k [29]
Happy, smile, humor, happiness, friendship, joyful, cheerful, jovial	Happy	13065	Goodreads quotes [28] Quotes-500k [29] Azquotes [30]
Fear, afraid, fearful, mistakes, out-of-control, scare, death, horror, panic	Fear	12730	Goodreads quotes [28] Quotes-500k [29]
Anger, ego, misinterpretation, misunderstanding, overthinking, ignorance, bother	Angry	9850	Goodreads quotes [28] Quotes-500k [29]
Sad, unhappy, sorrow, sadness, suffering, pain, alone, suicide	Sad	12560	Goodreads quotes [28] Quotes-500k [29]

To create a rich set of quotes for each emotion, three different datasets were considered. AZQuotes was included as it provides a wide range of quotes related to the "Happy" emotion with unique keywords not found in GoodReads and the 500k datasets. Meanwhile, GoodReads and the 500k datasets cover essential keywords related to all emotions, ensuring comprehensive representation.

3.2 Photo cartoonization

This module deals with cartoonization of a user's photo using an enhanced pre-trained AnimeGANv2 model, which is explained in Algorithm 1. AnimeGANv2 is modified to include a quantization process to obtain a more animated look. User inputs captured photos and needs to select any one of the three anime styles available: Hayao, Shinkai, and Paprika. These photos are preprocessed to make them compatible for the model input. Next, k-means-based quantization is applied. The selected model converts quantized output into anime photos, which are then verified for similarity threshold with real photos. In case of poor similarity falling under the threshold value, users need to recapture and input the photo into this module. The anime photos above the set threshold are passed for the next module, i.e., the Emotion Recognition and Validation module.

3.2.1 Enhanced AnimeGANv23.3

Color quantization using the k-means process, detailed in Algorithm 1 is applied to the AnimeGANv2 model to produce a more cartoonized effect in the resulting anime images. The color quantization technique is applied to reduce the number of distinct colors in an image while preserving its overall appearance.

Algorithm 1: K-means Color Quantization
Given:
X: Set of RGB color vectors representing the pixels in the image.
$X=\{x_1,x_2,\ldots,x_n\}$, where $x_i \in \mathbb{R}^3$.
C: Set of centroids. $C=\{c_1,c_2,\ldots,c_k\}$, where $c_j \in \mathbb{R}^3$.
Sj: Set of color vectors assigned to cluster j.
Objective: Minimize the sum of squared distances between color vectors and their respective cluster centroids:

$$J(C)=\sum_{j=1 \text{to } k}\sum_{xi\in Sj}\|xi-cj\|^2$$

Initialization: Randomly initialize k centroids $C=\{c_1,c_2,\ldots,c_k\}$.
Assignment: Assign each color vector x_i to the cluster j with the closest centroid:

$$S_j=\{x_i:\|x_i-c_j\|^2\leq\|x_i-c_l\|^2,\forall l,1\leq l\leq k\}$$

Update: Recalculate centroids based on the mean of color vectors in each cluster:

$$c_j=1/|S_j|\sum_{xi\in Sj}x_i$$

Repeat: Until convergence, repeat the assignment and update steps.

The final centroids C represent the reduced color palette for the quantized image, and each pixel is assigned the color of its corresponding centroid. Convergence can be determined by observing when centroids no longer change significantly or after a fixed number of iterations.

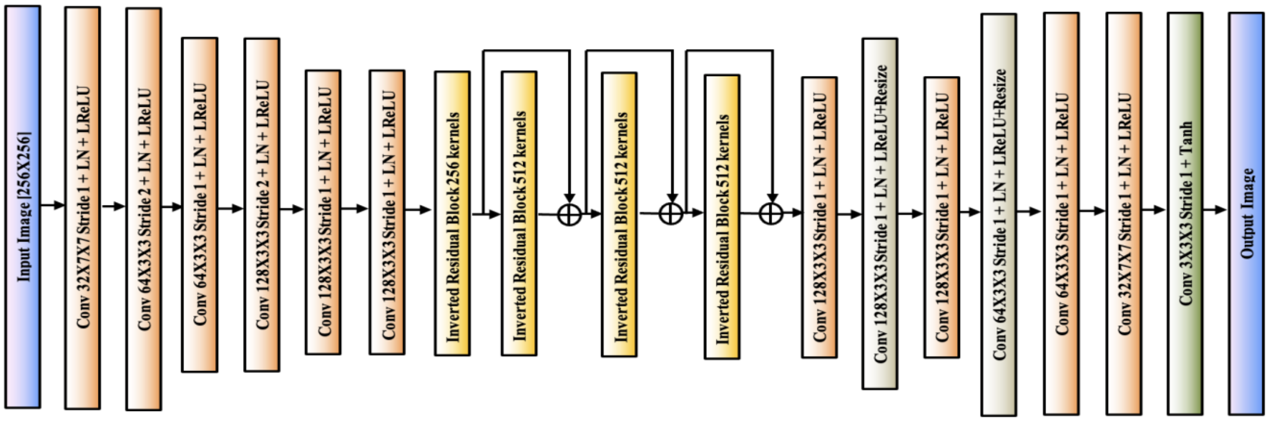

Figure 2. Architecture diagram of enhanced AnimeGANv2 [31]

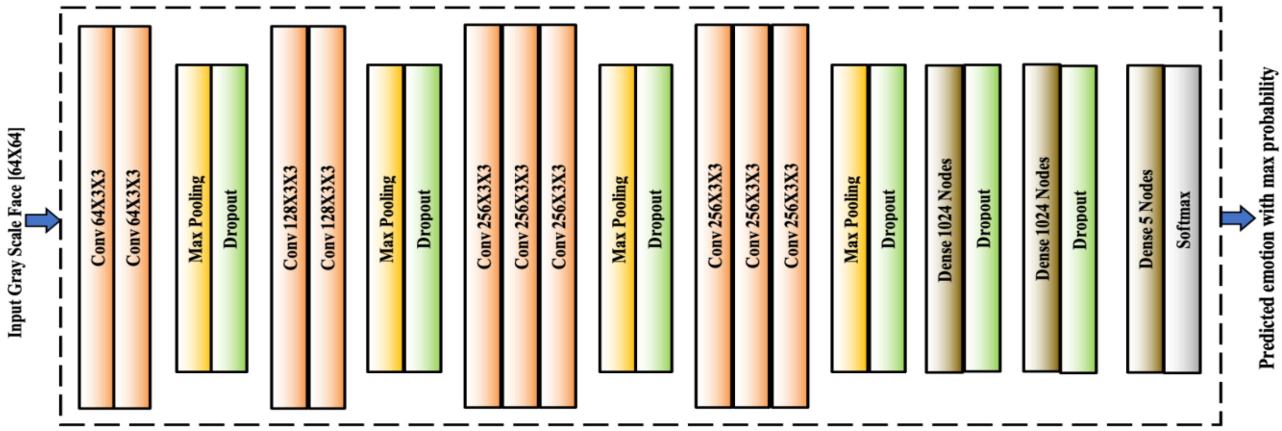

Figure 3. Architecturediagram of DNN [32]

AnimeGANv2 [31] is an image-to-image translation framework involving a combination of convolution and inverted residual block layers, which is headed by K-Color Quantization as depicted in Figure 2. It is specifically designed for anime-style images. It is based on a GAN architecture and utilizes a variety of techniques, including multi-scale generators, perceptual loss, and style loss. The goal of AnimeGANv2 is to convert input images into anime-style images that look natural and appealing. It includes four different pre-trained models that are specialized for different anime styles: Hayao, Shinkai, and Paprika. Multi-scale generators produce multiple images at different resolutions to produce a final combined output image having high-quality images with fine details and smooth edges. Perceptual loss involves comparing the generated features to the original at multiple levels of abstraction. By using perceptual loss, AnimeGANv2 can generate images that look more natural and realistic. Style loss to preserve the style of the original image. By using style loss, AnimeGANv2 can produce anime-style images that are faithful to the style of the original image. It uses spectral normalization to normalize the weights of the neural network, which helps prevent the network from becoming unstable during training. Additionally, it uses batch normalization and residual connections to improve the performance of the network.

3.3 Emotion recognition and validation

The key feature of the proposed system is to retain emotion from real photos to anime. To identify an expression over the face, facial features such as eyebrows, eyes, nose, mouth, and their movements are analyzed. There are various methods for facial expression recognition, including traditional computer vision techniques and deep learning methods. This module employs SSD to first extract the face region, which is then passed to a pre-trained DNN model to classify facial expression from a custom defined emotion list. This process is applied to both real and anime pictures; subsequently, their emotions are compared. If emotions are matching, then the anime photo is passed further to the next module for quoted cartoon card generation.

3.3.1 DNN model

The model DNN [32] is trained on the FERPlus large-scale dataset for facial expressions. It supports seven basic emotions and one additional emotion, contempt. It is customized to the proposed system requirements to have five distinct emotions: Neutral, Happy, Fear, Sad, and Anger. It takes an input 64 × 64 grayscale image, which is passed to a set of convolutional layers for extracting essential features. The later part of the architecture includes dense layers employed for prediction, which is depicted in Figure 3. The output of the model is a vector of scores for each emotion class. The scores are then passed through the application of softmax function, which gives out a probability distribution over the emotion classes. The one with maximum probability is picked as an emotion label.

The process of detecting emotions from images involves several steps, including image preprocessing, face detection, feature extraction, and emotion classification. A detailed explanation of the entire process is given below.

3.3.2 Input image processing

The first step is to load the input image X, which could be either an anime or a real photo, into the system. This image X is represented as a matrix, where each element corresponds to a pixel's intensity value. For further processing, a copy of this image X_{copy} is created to preserve the original image while allowing annotations such as bounding boxes and emotion labels later in the process:

$$X_{copy}= X$$

3.3.3 Face detection using SSD

Once the image is loaded, the next step is to detect faces within the image using the Single Shot Detector (SSD) algorithm. SSD is a deep learning-based approach that efficiently detects objects, in this case, faces, within an image. The output of the SSD is the bounding box coordinates F(X) of the detected face:

$$F(X)=SSD(X) \tag{1}$$

The detected face region $R_F$is then extracted from the original image based on these coordinates:

$$R_F=X[F(X)] \tag{2}$$

3.3.4 Face region preprocessing

After detecting the face, it is important to preprocess the face region to prepare it for emotion recognition by a DNN. First, padding is applied to the face region R_F to ensure the model has sufficient context:

$$P_F=pad(R_F) \tag{3}$$

Next, the padded face region P_F is converted to grayscale, which reduces the complexity of the image by removing color information while retaining the essential features required for emotion detection:

$$G_P=convert\_to\_grayscale(P_F) \tag{4}$$

The grayscale face image G_P is then resized to a fixed dimension of 64×64pixels to match the input size expected by the DNN model:

$$R_G=cv2.resize(G_P,64\times 64) \tag{5}$$

Finally, the resized grayscale image R_G is reshaped to fit the required input format of the DNN:

$$R_{input}=reshape(R_G) \tag{6}$$

3.3.5 Emotion recognition using DNN

With the face region prepared, the next step involves using a pre-trained DNN to recognize the emotion expressed by the detected face. The reshaped image R_{input} is fed into the DNN, where it undergoes forward propagation to produce raw emotion scores S:

$$S=DNN(R_{input}) \tag{7}$$

These raw scores are then passed through a softmax function, which converts them into a probability distribution PPP over the predefined emotion classes (e.g., SAD, ANGER, NEUTRAL, FEAR, HAPPY):

$$P=Softmax(S) \tag{8}$$

The emotion with the highest probability is selected as the predicted emotion E:

$$E=argmax(P) \tag{9}$$

3.3.6 Annotate and display the image

To visualize the detected emotion, the predicted emotion Eis is annotated on a copy of the original image X_{copy} using text labels, and a bounding box is drawn around the detected face:

$$X_{annotated}=cv2.putText(cv2.rectangle(Xcopy,F(X)),E) \tag{10}$$

Finally, the annotated image $X_{annotated}$ is displayed using a plotting library like matplotlib.pyplot, allowing you to see the detected face along with its predicted emotion:matplotlib.pyplot.imshow($X_{annotated}$).

3.3.7 Repeat for real photo

The same process is repeated for a real photo to compare the emotions predicted by the model for both images.

3.4 Quoted cartoon card creation

Quotes in profile pictures can serve as a form of personal expression, allowing individuals to convey their thoughts, values, or emotions to their online audience. The chosen quote can reflect their personality, beliefs, or current state of mind. Some people use inspirational quotes to motivate themselves and others. Humorous quotes can be a way to share a laugh or simply lighten the mood among your followers. Posting a quote with a profile picture shares your thoughts and feelings with your friends and followers, allowing them to understand the context of your happiness or sorrow better. An anime version of a real photo looks more appealing and at the same time, preserves privacy. This module takes an anime-flavoured photo and its emotion as inputs. GPT-2 simple is trained and fine-tuned on our 3.4. quote’s dataset, which generates three quotes relevant to the input emotion. The anime photo is patched with a quote, generating the Expressive Cartoon Card ECC). Three ECCs are generated, one for each quote, so that the user has a choice to select the most appropriate one. The ECC generation process is detailed in Algorithm 2.

Algorithm 2: Emotion Quote Generator

1. Load the required Python libraries:
   - gpt_2_simple for quote generation
   - numpy for numerical operations
   - matplotlib for display images
   - csv, pandas for reading .csv file
2. Map emotions to keywords
3. Set gpt-2 model_name= 124M // gpt-2 small
4. Load model using gpt2.download_gpt2
5. Load Quotes dataset using pd.read_csv("q.csv")
6. Fine tune gpt-2 small with steps=30 // Train gpt-2
7. Save the model

// Inference model to generate quotes for input keyword

8. Call generate function with parameters settings:
prefix = keyword,
nsamples =3,
length =100,
temperature=0.8,
batch_size=3

9. Prepare Expressive Cartoon Card(ECC) for each quote
10. Display top 3 ECCs for user selection

3.4.1 Generative pre-trained transformer-2 (GPT-2) simple

GPT-2 [33] uses the decoder-only transformer architecture depicted in Figure 4, in which a self-attention mechanism is employed to capture long-term dependencies in the text. The GPT-2 generates text that is contextually relevant and coherent. This means that it can consider the context in which it is generating and produce text that is consistent with that context. In this paper, it is tailored for generating quotes based on prefix keywords corresponding to emotions. Stacking multiple identical layers allows the model to capture both low-level and high-level patterns in the text. It uses positional encodings to inject information about the position of each word/token in the sequence. After the self-attention mechanism, each layer includes feedforward neural networks. These networks transform the output of the attention mechanism into a form that is suitable for further processing. Layer normalization is applied after each sub-layer (self-attention and feedforward) to stabilize training. Additionally, residual connections are used to facilitate the flow of gradients during training.

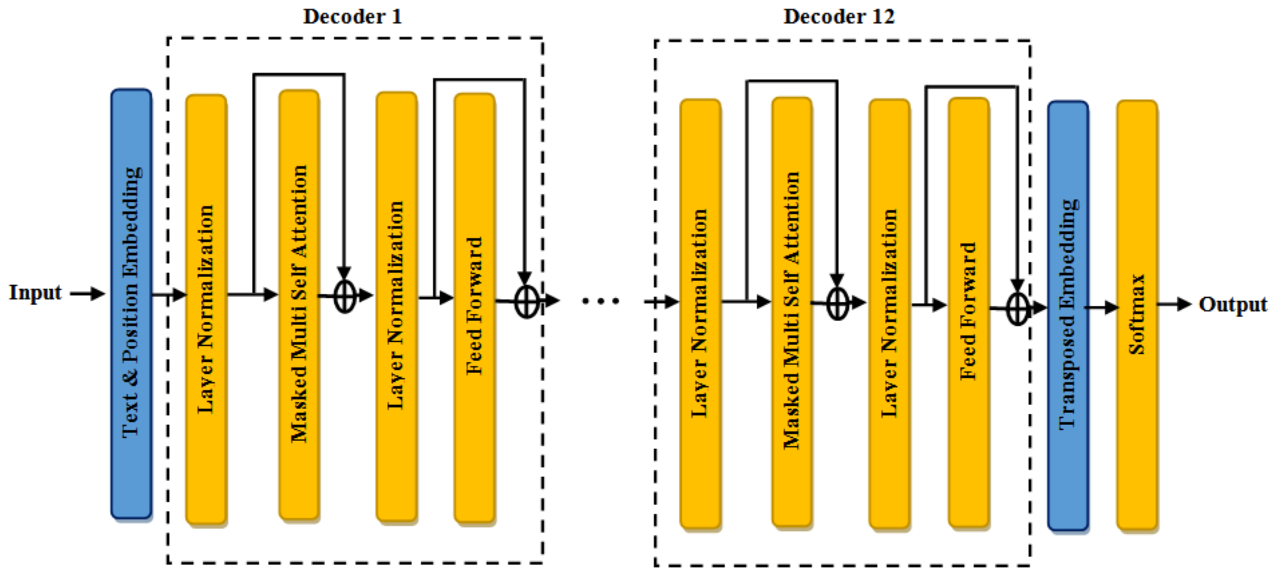

Figure 4. Architecture diagram of GPT-2 simple [33]

4. RESULTS, DISCUSSION AND ANALYSIS

The proposed system experimentation requires the installation of following packages: onnxruntime-gpu, numpy, opencv-python, bleedfacedetector, opencv-contrib-python, dlib and gpt-2-simple. Metrics used for evaluating the image quality of generated anime are described below:

Peak Signal-to-Noise Ratio (PSNR): PSNR given in Eq. (11), measures image quality by comparing the peak signal power to the noise power, often in decibels.

$$PSNR=10 \cdot log_{10}(MAX^2/MSE) \tag{11}$$

where, MAX is the maximum pixel value, and MSE is the Mean Squared Error between the original and the compressed image.

Structural Similarity-Index (SSIM): SSIM given in Eq. (12), gauges structural similarity between two images, considering luminance, contrast, and structure.

$$\mathrm{SSIM}(x,y)= (2\mu_x\mu_y+c_1)(2\sigma_{xy}+c_2) / (\mu_x^2+\mu_y^2+c_1)(\sigma_x^2+\sigma_y^2+c_2) \tag{12}$$

where μ_x,μ_y are average luminance, $\sigma_x,\sigma y$ are standard deviations, σ_{xy} is covariance, and c_1,c_2 are stability constants.

Fréchet Inception Distance (FID): FID given in Eq. (13), measures similarity between two datasets using Fréchet distance between Gaussian distributions of features from the InceptionV3 network.

$$\mathrm{FID}=\|\mu_1-\mu_2\|^2+\mathrm{Tr}(\Sigma_1+\Sigma_2-2(\Sigma_1\Sigma_2)^{0.5}) \tag{13}$$

where, μ_1,μ_2 are means, Σ_1,Σ_2 are covariance matrices of feature vectors.

Module 1 takes the user's natural photo as input, applies k-means quantization process, and then converts it into an anime style photo. It comes in three different styles: Hayao (style1), Shinkai (style2), and Satoshi (style3). Figure 5 shows three different anime styles generated for given four sample photos.The three anime styles of a user input photo are passed to module

In module 2, real and anime style photos emotions are identified. Emotion recognition and comparison results for the three anime styles are depicted in Figures 6-8.

Next module of the proposed ECC generation system deals with generation of quoted cartoon cards. Initially, the user selects one of the three anime styles, for which three quotes are generated. Subsequently, these quotes are tagged to the anime photo producing ECCs. The final results of the ECC generation process for five different emotions are shown in Figure 9.

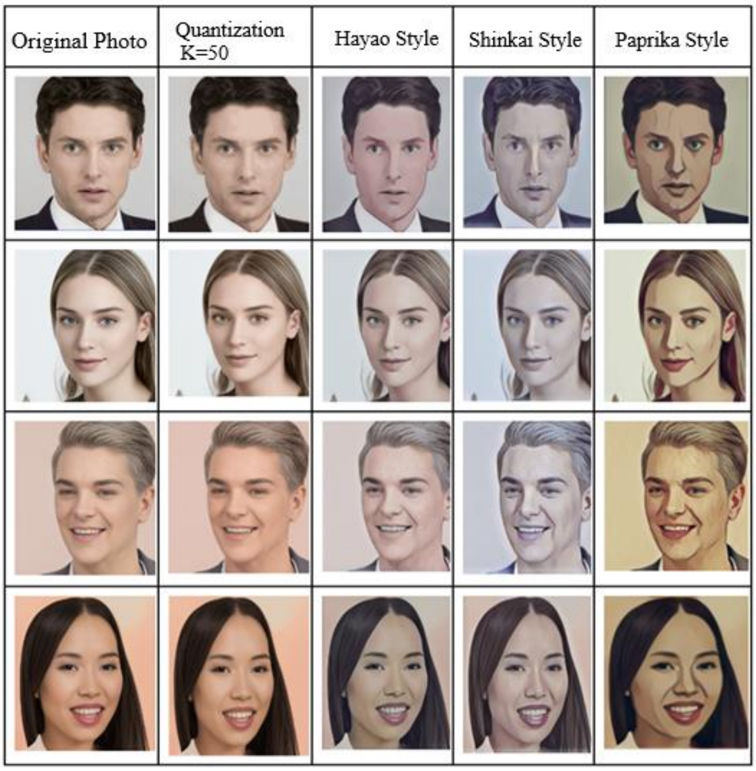

Figure 5. Anime photos of four sample images in Hayao, Shinkai and Paprika styles [23]

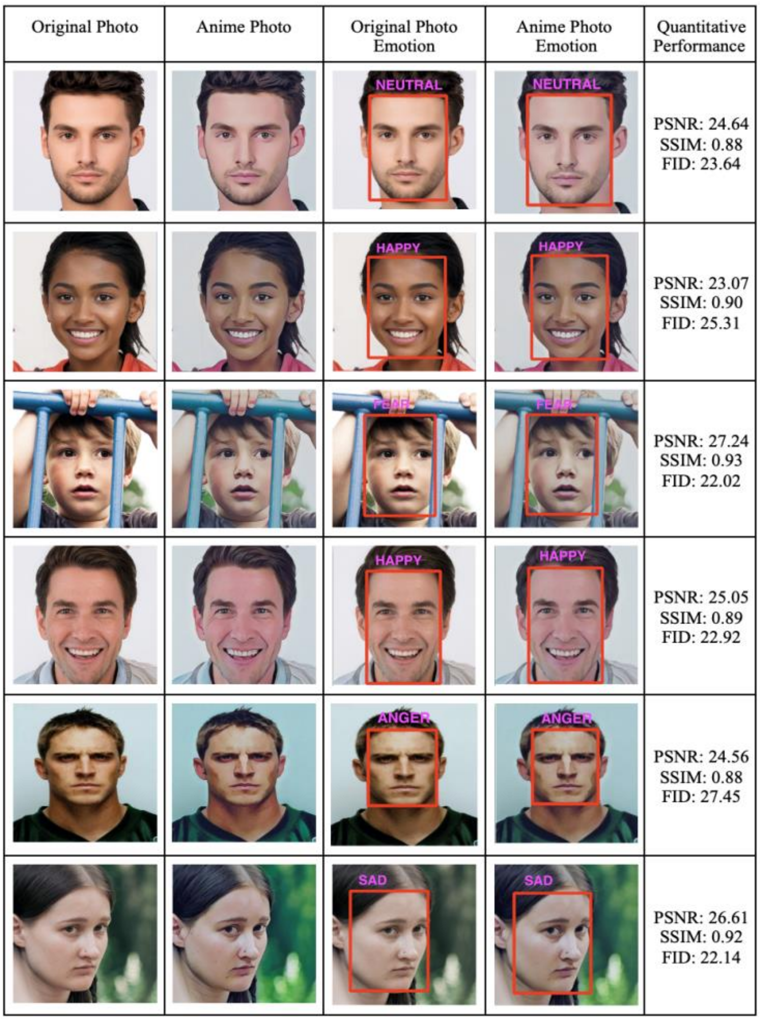

Figure 6. Emotion recognition results for Hayao style [24]

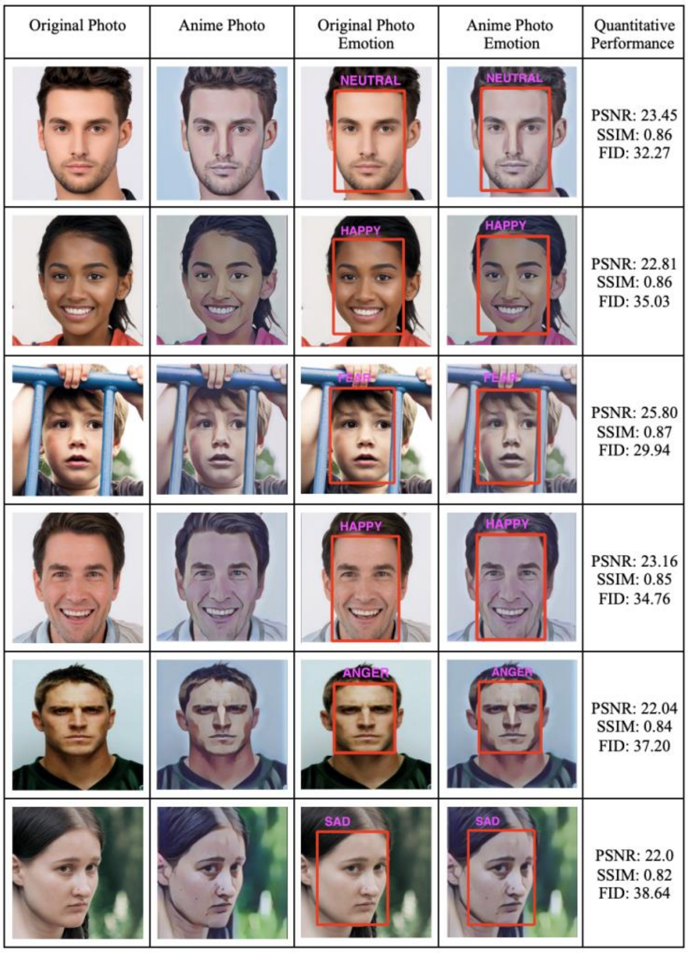

Figure 7. Emotion recognition results for Shinkai style [25]

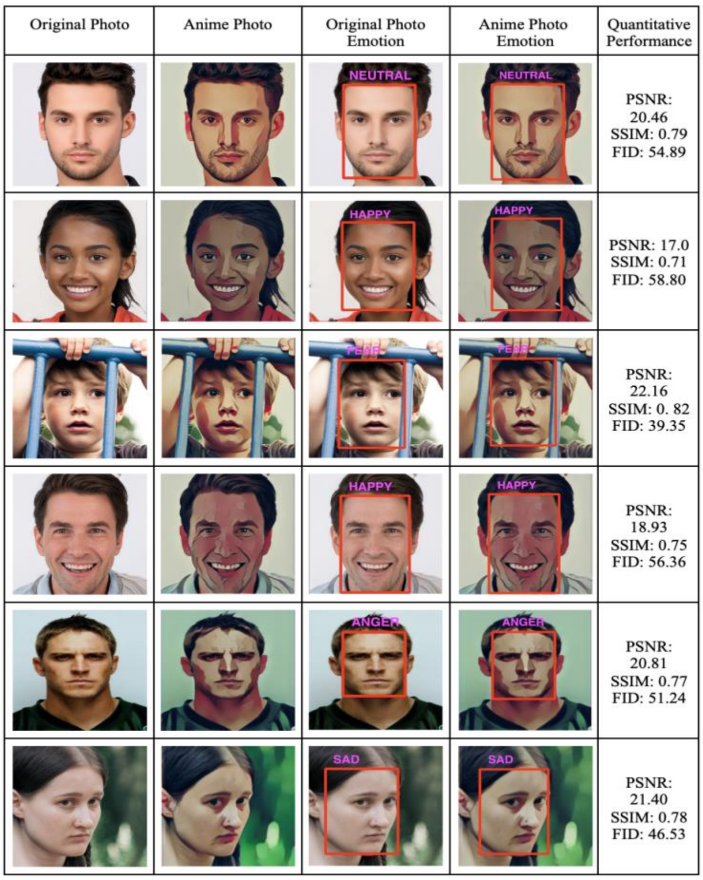

Figure 8. Emotion recognition results for Satoshi style [26]

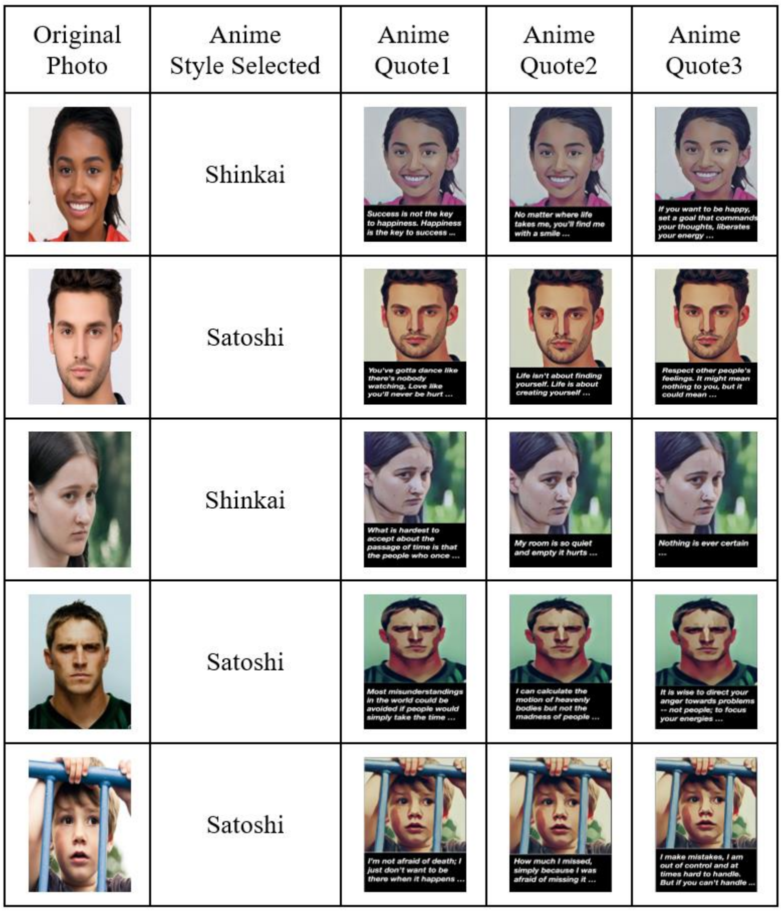

Figure 9. Results of cartoon cards tagged with quotes

5. COMPARATIVE ANALYSIS

The quantitative analysis results of the proposed three anime styles and their comparison with other techniques are included in Table 4 and Table 5 respectively. The Hayao Style performed best when compared with other two styles, which is depicted in Figure 10. Hayao and Shinkai styles have shown superior performance over contemporary methods, which is shown in Figure 11.

The confusion matrix of anime emotion prediction is shown in Table 6 and performance of three styles in recognition of the five emotions in terms of precision (pr), recall (re) and F1-Score (F1) are detailed in Table 7. Hayao and Shinkai styles performed better for emotion "Happy" comparatively as shown in Figures 12 and 13 on the other hand Satoshi style recognized "Happy" and "Fear emotions comparatively good over other emotions as shown in Figure 14.

Table 4. Quantitative analysis of proposed anime styles

	Style 1			Style 2			Style 3		
	BS	LS	AS	BS	LS	AS	BS	LS	AS
PSNR↑	27.24	23.07	25.16	25.80	22.0	23.90	22.16	17.0	19.58
SSIM↑	0.93	0.88	0.91	0.87	0.82	0.85	0.82	0.71	0.77
FID↓	22.02	27.45	24.74	29.94	38.64	34.29	39.35	58.80	49.08

Note: BS: Best Score, LS: Least Score, AS: Average Score

Table 5. Proposed anime styles in comparison to other methods in terms of FID

Method	FID↓
Arefeen toon2real [34]	39.25
Hao 3D Cartoon with Expressions [35]	146.8
Dahua LineGAN [36]	38.74
Yifang Gated Cycle Mapping [37]	79.74
Proposed Anime Style 1	24.74
Proposed Anime Style 2	34.39
Proposed Anime Style 3	49.08

Table 6. Confusion matrix of emotion predicted (row) Vs Actual emotion(column) of 3 styles

	Happy		Sad		Angry		Fear		Neutral	
S1	R	A	R	A	R	A	R	A	R	A
H	45	45	0	0	0	0	0	0	0	0
S	0	0	34	34	0	0	1	1	1	1
An	0	0	0	0	39	39	0	0	1	1
F	0	0	0	0	0	0	36	36	0	0
N	0	0	2	2	0	0	1	1	48	48
S2										
H	45	44	0	0	0	0	0	0	0	0
S	0	0	34	32	0	0	1	1	1	2
An	0	1	0	1	39	37	0	0	1	1
F	0	0	0	1	0	1	36	36	0	0
N	0	0	2	2	0	1	1	1	48	47
S3										
H	45	44	0	0	0	2	0	0	0	0
S	0	0	34	33	0	0	1	1	1	1
An	0	1	0	0	39	37	0	0	1	2
F	0	0	0	0	0	0	36	36	0	1
N	0	0	2	3	0	0	1	1	48	46

Note: R: Real Photo, A: Anime Photo, S1: Style 1, S2: Style 2, S3: Style 3, H: Happy, S: Sad, An: Angry, F: Fear, N: Neutral

Table 7. Performance analysis of proposed anime styles

Emotion	Precision				Recall				F1-Score			
	R	S1	S2	S3	R	S1	S2	S3	R	S1	S2	S3
Happy	1.0	1.0	1.0	0.96	1.0	1.0	0.98	0.98	1.0	1.0	0.99	0.96
Sad	0.94	0.94	0.91	0.94	0.94	0.94	0.89	0.92	0.94	0.94	0.90	0.93
Angry	0.97	0.97	0.92	0.92	1.0	1.0	0.95	0.95	0.99	0.99	0.94	0.93
Fear	1.0	1.0	0.95	0.97	0.95	0.95	0.95	0.95	0.97	0.97	0.95	0.96
Neutral	0.94	0.94	0.92	0.92	0.96	0.96	0.94	0.92	0.95	0.95	0.94	0.92
M_{Avg}	0.97	0.97	0.94	0.94	0.97	0.97	0.94	0.94	0.97	0.97	0.94	0.94
W_{Avg}	0.97	0.97	0.94	0.94	0.97	0.97	0.94	0.94	0.97	0.97	0.94	0.94

Note: R: Real Photo, S1: Style 1, S2: Style 2, S3: Style 3, M_{Avg}: Macro Average, W_{Avg}: Weighted Average

Table 8. Overall accuracy of the proposed DNN for three styles

Real Photo	Anime Style1	Anime Style2	Anime Style3
0.971	0.971	0.942	0.942

Table 9. Comparison of proposed styles' overall accuracy with contemporary emotion recognition methods

Model	Overall Accuracy
CNN+Adam [38]	60.6%
VGG 16 on Cartoon Images [39]	96.0%
CUSTOM_CNN ON FER-2013, JAFFE, CK+ [40]	91.6%
VGG 16 [41]	73.0%
FERNET_DEEP CNN[42]	69.6%
EMOLA_FABA[43]	64.5%
Proposed DNN on Style1	**97.1%**
Proposed DNN on Style2	94.2%
Proposed DNN on Style3	94.2%

The overall emotion prediction accuracy of Hayao style is better than the other two styles comparatively, and also it outperforms other contemporary models whereas the other two styles are equally good, which are detailed in Tables 8 and 9, respectively. Hayao Style performance is near to real photos from Figure 15 but the other two styles are also equally good when compared to contemporary models while producing more cartoon-look from Figure 9 and Figure 16. Realistic visuals (e.g., Real Photo) consistently outperform cartoonized representations in emotion detection, as they preserve critical expression details necessary for accurate classification. In contrast, highly cartoonized visuals as in Style3 significantly compromise the precision, recall, and F1-Score for nuanced emotions like Sad, Neutral, and Fear, where subtle facial features are crucial. Emotions with distinct and exaggerated features, such as Happy and Angry, demonstrate robustness across styles, being less impacted by the loss of detail in stylized photos. If the emotion recoginized is incorrect when compared to the original visual, then user needs to retake in better lighting conditions and angle, which forms constraint for the proposed utility.

GPT-2 for quote generation has performed well in emotion related quotes generation, which is analyzed using training and validation losses as depicted in Figure 17. GPT-2 is a cost-effective option for training and deployment, ideal for budget-limited projects. It can be fine-tuned locally, ensuring data privacy without relying on external APIs like GPT-3. Its efficiency on smaller hardware setups makes it suitable for environments without high-end GPUs or cloud resources.

Table 10 shows the proposed GPT2's comparison with other models. The Cosine Similarity score is used as a metric to measure similarity between emotion tags to keywords extracted from the generated quote [44]. ROUGE and BLEU scores [45] are computed as the highest score from the set to evaluate the quality of the quote retrieved from the quotes dataset when compared against each emotion tag from the set for contextual relevance. The proposed model's performance when compared with other models or methods demonstrates significant improvement. The cosine similarity score of 96.8% indicates an exceptionally high semantic overlap, showing that the keyword is strongly aligned with the quote sentence's meaning. The ROUGE score of 0.84 reflects a high recall of n-grams, highlighting significant content overlap between the keyword and the sentence. Meanwhile, the BLEU score of 0.78 demonstrates good precision, suggesting that most n-grams match well, with minor variations in phrasing or structure. Together, these metrics confirm a strong contextual and textual relevance between the keyword and the quote.

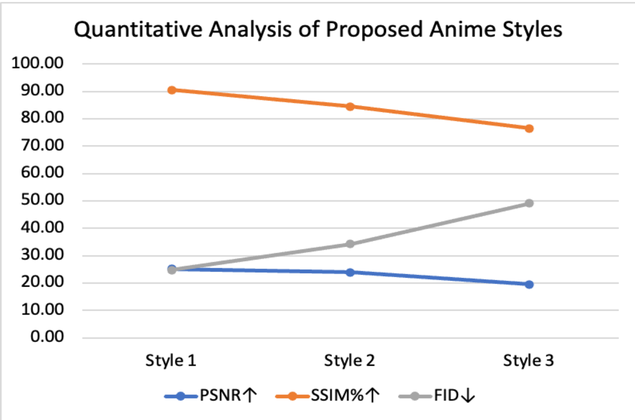

Figure 10. Quantitative analysis of proposed three anime styles

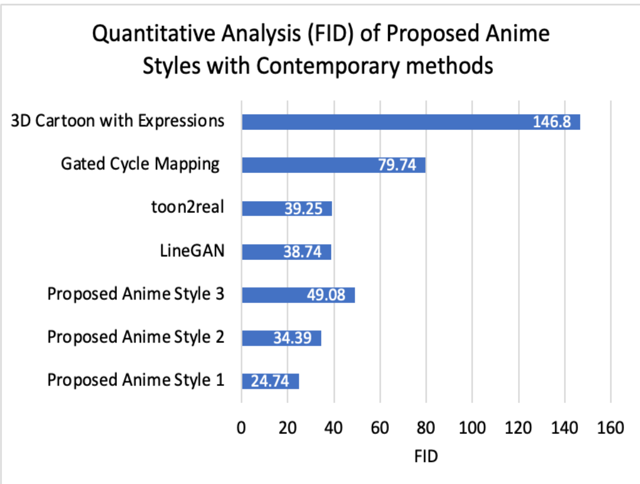

Figure 11. Comparison of proposed anime styles with other methods

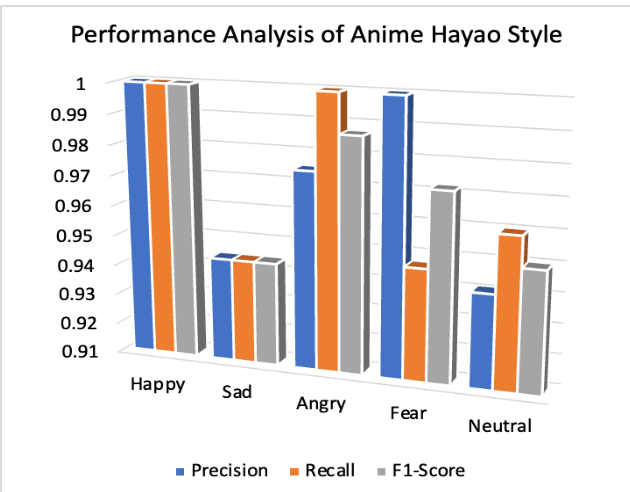

Figure 12. Hayao style's performance analysis

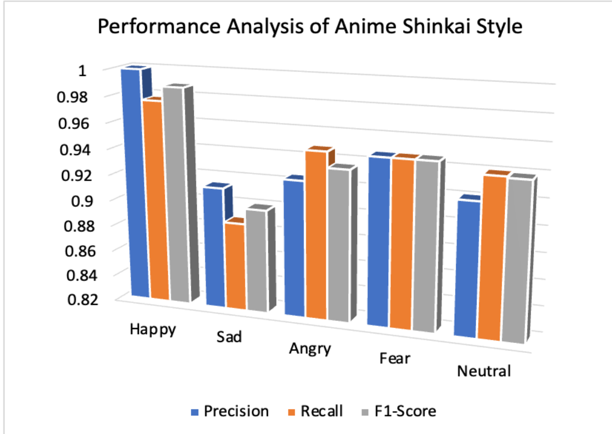

Figure 13. Shinkai style's performance analysis

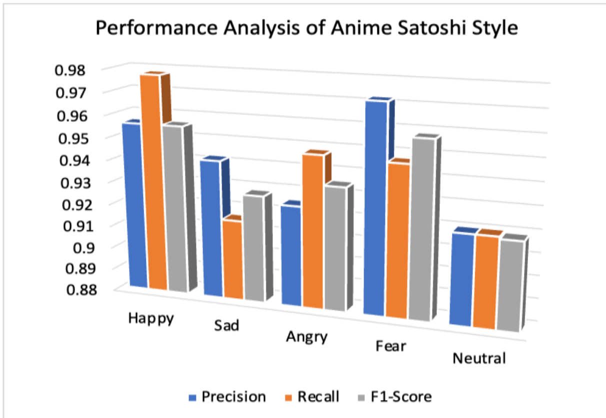

Figure 14. Satoshi style's performance analysis

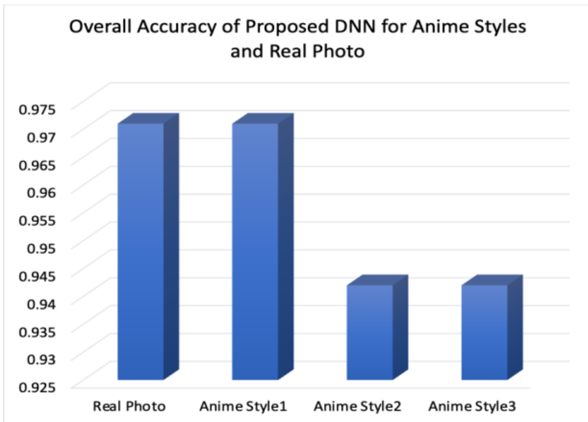

Figure 15. Comparison of three styles with respect to overall accuracy

Table 10. Comparative analysis of proposed GPT2s

Method/Model	Cos_Similarity%	ROUGE	BLEU
Quote generator for Image Tag [44]	83.2%	-	-
ChipperAggregati [45]	-	0.57	0.45
TextRank [46]	71.3%	-	-
Instruction tuning based on Personality Traits [47]	32.31%	21.21	0.93
GPT2S	96.8%	0.84	0.78

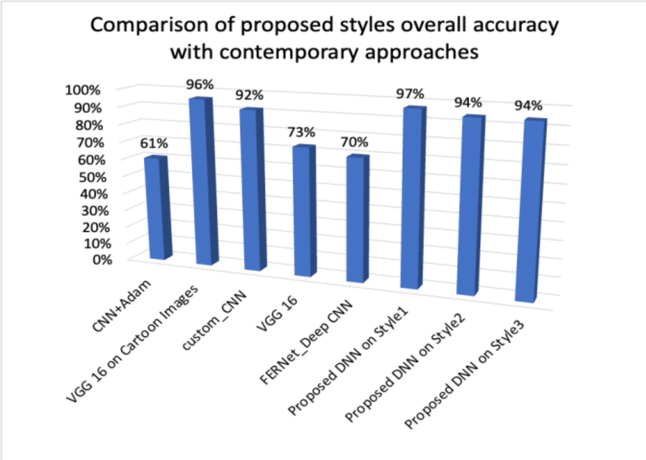

Figure 16. Comparison of three styles with other models in terms of overall accuracy

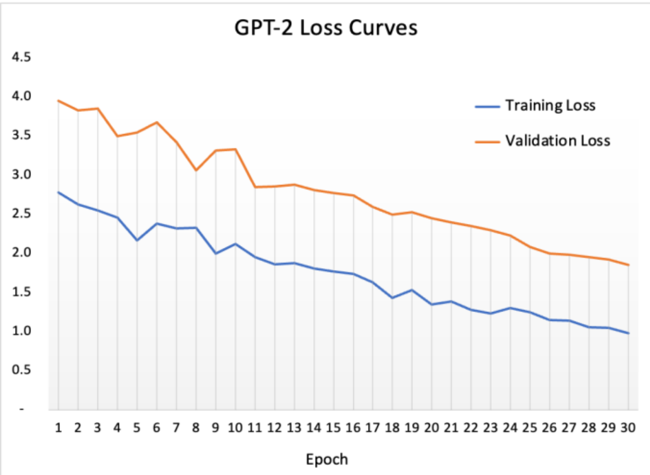

Figure 17. GPT-2 simple loss curves

6. CONCLUSION AND FUTURE WORK

People update profile pictures in social media accounts to express their current state of living and feelings. The AnimeGANv2, which is pipelined with k-means quantization technique produced cartoon-look privacy preserving pictures holding essential facial expressions. The average score of PSNR, SSIM, and FID of three styles are 24.74, 34.39, and 49.08, shows that more realistic cartoon-look photos are generated. The SSD algorithm extracted the face region from the anime photo. Next, the DNN model design, which is tailored to the most general five emotions recognition, is implemented to validate anime emotion with real one. It has exhibited an overall accuracy of 0.97 for Hayao style and 0.94 for other two styles. Test results of quotes generation for given emotion is implemented using GPT-2 simple. It has shown reduced training and validation loss of 1.0 and 1.9 at 30 epochs. Hence, the proposed Expressive Cartoon Card (ECC) system marks a significant advancement in social media content creation by providing a realistic, privacy-preserving, and visually appealing approach, outperforming contemporary GANs and quote generators in terms of computational efficiency. Integrating preprocessing algorithms to enhance low-quality images can reduce emotion validation errors and improve prediction accuracy. Expanding beyond the limited styles supported by AnimeGANv2, advanced algorithms can enable a wider range of transformations, such as Face-to-Kpop, Portrait-to-Pixar, and Nordic Myth-inspired designs. Moreover, developing context-based quote generation datasets tailored to specific conversations or situations, as well as exploring diverse emotion datasets with greater variation, can enhance the system's adaptability and precision. Finally, the introduction of specialized social media applications for sharing Expressive Cartoon Cards in personal groups can open new avenues for creative and personalized communication.

ACKNOWLEDGMENT

The authors would like to thank Deanship of Scientific Research at Majmaah University for supporting this work under Project Number R-2025-1735.

REFERENCES

[1] Li, X., Zhang, W., Shen, T., Mei, T. (2019). Everyone is a cartoonist: Selfie cartoonization with attentive adversarial networks. In 2019 IEEE International Conference on Multimedia and Expo (ICME), Shanghai, China, pp. 652-657. https://doi.org/10.1109/ICME.2019.00118

[2] Zhao, X., Zhou, Y., Wu, J., Xu, Q., Zhang, Y. (2021). Turn real people into anime cartoonization. In 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE), Guangzhou, China, pp. 387-393. https://doi.org/10.1109/ICCECE51280.2021.9342433

[3] Dong, Y., Tan, W., Tao, D., Zheng, L., Li, X. (2021). CartoonLossGAN: Learning surface and coloring of images for cartoonization. IEEE Transactions on Image Processing, 31: 485-498. https://doi.org/10.1109/TIP.2021.3130539

[4] Lei, P., Xu, S., Zhang, S. (2024). An art-oriented pixelation method for cartoon images. The Visual Computer, 40(1): 27-39. https://doi.org/10.1007/s00371-022-02763-0

[5] Patankar, A.B., Kubde, P.A., Karia, A. (2016). Image cartoonization methods. In 2016 International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, pp. 1-7. https://doi.org/10.1109/ICCUBEA.2016.7860045

[6] Shu, Y., Yi, R., Xia, M., Ye, Z., Zhao, W., Chen, Y., Lai, Y.K., Liu, Y.J. (2021). Gan-based multi-style photo cartoonization. IEEE Transactions on Visualization and Computer Graphics, 28(10): 3376-3390. https://doi.org/10.1109/TVCG.2021.3067201

[7] Wang, S., Qi, J. (2022h). A novel image cartoonization algorithm without deep learning. In CIBDA 2022; 3rd International Conference on Computer Information and Big Data Applications, Wuhan, China, pp. 1-4.

[8] Shahcheraghi, Z., See, J. (2013). On the effects of pre- and post-processing in video cartoonization with bilateral filters. In 2013 IEEE International Conference on Signal and Image Processing Applications, Melaka, Malaysia, pp. 37-42. https://doi.org/10.1109/ICSIPA.2013.6707974

[9] Chen, Y., Lai, Y.K., Liu, Y.J. (2018). Cartoongan: Generative adversarial networks for photo cartoonization. In Proceedings of the IEEE Conference on Computer

Vision and Pattern Recognition, pp. 9465-9474. https://doi.org/10.1109/CVPR.2018.00986
[10] Acharjee, J., Deb, S. (2021). Cartoonize images using TinyML strategies with transfer learning. In 2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), Zallaq, Bahrain, pp. 411-417. https://doi.org/10.1109/3ICT53449.2021.9581835
[11] Wang, X., Yu, J. (2020). Learning to cartoonize using white-box cartoon representations. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8090-8099. https://doi.org/10.1109/CVPR42600.2020.00811
[12] Júnior, J.D.O.A., de Queiroz, J.E.R., Gomes, H.M. (2013). An approach for non-photorealistic rendering that is appealing to human viewers. In 2013 XXVI Conference on Graphics, Patterns and Images, Arequipa, Peru, pp. 242-249. https://doi.org/10.1109/SIBGRAPI.2013.41
[13] Wu, Y.C., Chiu, L.W., Lai, C.C., Wu, B.F., Lin, S.S. (2023). Recognizing, fast and slow: Complex emotion recognition with facial expression detection and remote physiological measurement. IEEE Transactions on Affective Computing, 14(4): 3177-3190. https://doi.org/10.1109/TAFFC.2023.3253859
[14] Mou, L., Zhao, Y., Zhou, C., Nakisa, B., Rastgoo, M.N., Ma, L., Huang, T., Yin, B., Jain, R., Gao, W. (2023). Driver emotion recognition with a hybrid attentional multimodal fusion framework. IEEE Transactions on Affective Computing, 14(4): 2970-2981. https://doi.org/10.1109/TAFFC.2023.3250460
[15] Panlima, A., Sukvichai, K. (2023). Investigation on MLP, CNNs and Vision Transformer models performance for Extracting a Human Emotions via Facial Expressions. In 2023 Third International Symposium on Instrumentation, Control, Artificial Intelligence, and Robotics (ICA-SYMP), Bangkok, Thailand, pp. 127-130. https://doi.org/10.1109/ICA-SYMP56348.2023.10044742
[16] Sakai, T., Seo, M., Matsushiro, N., Chen, Y.W. (2023). Simulation of facial palsy using an improved cycle GAN and face restoration network. In 2023 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, pp. 1-4. https://doi.org/10.1109/ICCE56470.2023.10043574
[17] Bounyong, S., Yoshida, R., Yoshioka, M. (2023). Epistemic emotion detection by video-based and heart rate variability features for online learning. In 2023 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, USA, pp. 1-2. https://doi.org/10.1109/ICCE56470.2023.10043446
[18] Nguyen, H., Tran, N., Nguyen, H.D., Nguyen, L., Kotani, K. (2023). KTFEv2: Multimodal facial emotion database and its analysis. IEEE Access, 11: 17811-17822. https://doi.org/10.1109/ACCESS.2023.3246047
[19] Bishay, M., Preston, K., Strafuss, M., Page, G., Turcot, J., Mavadati, M. (2023). Affdex 2.0: A real-time facial expression analysis toolkit. In 2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG), Waikoloa Beach, USA, pp. 1-8. https://doi.org/10.1109/FG57933.2023.10042673
[20] Ma, X., Ma, Y. (2023). Relation-aware network for facial expression recognition. In 2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG), Waikoloa Beach, USA, pp. 1-7. https://doi.org/10.1109/FG57933.2023.10042525
[21] Le, N., Nguyen, K., Tran, Q., Tjiputra, E., Le, B., Nguyen, A. (2023). Uncertainty-aware label distribution learning for facial expression recognition. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 6088-6097. https://doi.org/10.1109/WACV56688.2023.00603
[22] Release Style dataset & Style video. TachibanaYoshino/AnimeGANv2, 2020. https://github.com/TachibanaYoshino/AnimeGANv2/releases/tag/1.0, accessed on March 2023.
[23] V7 Darwin, AI Data Labeling & ML Training Data Platform, 2018. https://www.v7labs.com/open-datasets/generated-faces, accessed on April 2023.
[24] Human Faces, Kaggle, 2020. https://www.kaggle.com/datasets/ashwingupta3012/human-faces, accessed on May 2023.
[25] Flickr-Faces-HQ Dataset (FFHQ), Kaggle, 2020. https://www.kaggle.com/datasets/arnaud58/flickrfacesh q-dataset-ffhq, accessed on May 2023.
[26] Facial age, Kaggle, 2019. https://www.kaggle.com/datasets/frabbisw/facial-age, accessed on June 2023.
[27] Murthy, A.R., Kumar, K.A. (2021). A review of different approaches for detecting emotion from text. In IOP Conference Series: Materials Science and Engineering, 1110(1): 012009. https://doi.org/10.1088/1757-899X/1110/1/012009
[28] Popular Quote, 2002. https://www.goodreads.com /quotes, accessed on June 2023.
[29] Quotes- 500k, Kaggle, 2020. https://www.kaggle.com/datasets/manann/quotes-500k, accessed on July 2023.
[30] 100 Happiness Quotes You Need to See Before You Die, AZ Quotes, 2012. https://www.azquotes.com/quotes/topics/happiness.html, accessed on Nov 2023.
[31] Chen, J., Liu, G., Chen, X. (2020). AnimeGAN: A novel lightweight GAN for photo animation. In Artificial Intelligence Algorithms and Applications: 11th International Symposium, ISICA 2019, Guangzhou, China, pp. 242-256. Springer Singapore. https://doi.org/10.1007/978-981-15-5577-0_18
[32] Barsoum, E., Zhang, C., Ferrer, C.C., Zhang, Z. (2016). Training deep networks for facial expression recognition with crowd-sourced label distribution. In Proceedings of the 18th ACM International Conference on Multimodal Interaction, New York, United States, pp. 279-283. https://doi.org/10.1145/2993148.2993165
[33] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I. (2019). Language models are unsupervised multitask learners. OpenAI blog, 1(8): 9.
[34] Arefeen Sultan, K.M., Imrul Jubair, M., Nahidul Islam, M.D., Hossain Khan, S. (2021). toon2real: Translating cartoon images to realistic images. *arXiv preprint* arXiv:2102.01143v1. https://doi.org/10.48550/arXiv.2102.01143
[35] Wang, H., Lin, G., Hoi, S.C., Miao, C. (2022). 3D cartoon face generation with controllable expressions from a single GAN image. arXiv preprint arXiv:2207.14425. https://doi.org/10.48550/arXiv.2207.14425
[36] Lv, D., Pu, Y., Nie, R. (2022). LineGAN: An image

colourisation method combined with a line art network. IET Computer Vision, 16(5): 403-417. https://doi.org/10.1049/cvi2.12096
[37] Men, Y., Yao, Y., Cui, M., Lian, Z., Xie, X., Hua, X.S. (2022). Unpaired cartoon image synthesis via gated cycle mapping. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, USA, pp. 3501-3510. https://doi.org/10.1109/CVPR52688.2022.00349
[38] Sarvakar, K., Senkamalavalli, R., Raghavendra, S., Kumar, J.S., Manjunath, R., Jaiswal, S. (2023). Facial emotion recognition using convolutional neural networks. Materials Today: Proceedings, 80: 3560-3564. https://doi.org/10.1016/j.matpr.2021.07.297
[39] Jain, N., Gupta, V., Shubham, S., Madan, A., Chaudhary, A., Santosh, K.C. (2022). Understanding cartoon emotion using integrated deep neural network on large dataset. Neural Computing and Applications, 34: 21481-21501. https://doi.org/10.1007/s00521-021-06003-9
[40] Khattak, A., Asghar, M.Z., Ali, M., Batool, U. (2022). An efficient deep learning technique for facial emotion recognition. Multimedia Tools and Applications, 81(2): 1649-1683. https://doi.org/10.1007/s11042-021-11298-w
[41] Dwijayanti, S., Iqbal, M., Suprapto, B.Y. (2022). Real-time implementation of face recognition and emotion recognition in a humanoid robot using a convolutional neural network. IEEE Access, 10: 89876-89886. https://doi.org/10.1109/ACCESS.2022.3200762
[42] Bodapati, J.D., Srilakshmi, U., Veeranjaneyulu, N. (2022). FERNet: A deep CNN architecture for facial expression recognition in the wild. Journal of the Institution of Engineers (India): Series B, 103(2): 439-448. https://doi.org/10.1007/s40031-021-00681-8
[43] Li, Y., Dao, A., Bao, W., Tan, Z., Chen, T., Liu, H., Kong, Y. (2024). Facial affective behavior analysis with instruction tuning. In European Conference on Computer Vision (pp. 165-186). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-72649-1_10
[44] Goel, S., Madhok, R., Garg, S. (2018). Proposing Contextually Relevant Quotes for Images. In Advances in Information Retrieval. ECIR 2018. Lecture Notes in Computer Science, 10772: 591-597. https://doi.org/10.1007/978-3-319-76941-7_49
[45] Yepes, A.J., You, Y., Milczek, J., Laverde, S., Li, R. (2024). Financial report chunking for effective retrieval augmented generation. arXiv preprint arXiv:2402.05131. https://doi.org/10.48550/arXiv.2402.05131
[46] Fakhrezi, M.F., Bijaksana, M.A., Huda, A.F. (2021). Implementation of automatic text summarization with TextRank method in the development of Al-qur'an vocabulary encyclopedia. Procedia Computer Science, 179: 391-398. https://doi.org/10.1016/j.procs.2021.01.021
[47] Huang, Y. (2024). Orca: Enhancing role-playing abilities of large language models by integrating personality traits. arXiv preprint arXiv:2411.10006. https://doi.org/10.48550/arXiv.2411.10006