Image Preprocessing Method Applied on Facial Emotion Recognition Problem

Mingyang Xu

Xi'an University of Posts & Telecommunications, China
2113203787@qq.com

Abstract. Sometimes people are unable to accurately express their views on a certain product, which may limit in accurately describing emotions. Facial emotion recognition can help solve the problem of people being unable to accurately express their opinions on products, by analyzing facial expressions to obtain emotional feedback. The application value and importance of this technology lies in providing a nonverbal way to understand user emotions, helping manufacturers enhance user satisfaction. Vision Transformer (ViT) model is very powerful on the Computer Vision Problems. In this paper. I used the VIT model to analyze the facial expression dataset and attempted different preprocessing methods. Through the VIT model. I can associate each data sample with the corresponding expression to determine the category of the expression. Then, compared the results obtained using different preprocessing methods and ultimately determine which method performs best in terms of accuracy. After comparing the accuracy obtained through various preprocessing methods, I found that RandomCrop has the highest accuracy and is most suitable for facial expression

Keywords: Facial emotion recognition, Image Preprocessing, Emotional patterns

1 Introduction

Human emotions play a crucial role in daily life [1]. Whether it's social communication, product experience, or emotional expression, our facial expressions play a very important role [2]. Traditionally, people use surveys and other methods to understand users' feelings about products, but sometimes it is difficult to accurately describe their emotions in words. However, many scientists have found that people's facial emotions can clearly reflect their views on products [3]. Therefore, through model training, I can establish a correlation between facial emotions and user perceptions of the product.

Facial emotion recognition is an important research field that involves understanding people's emotional state by analyzing facial expressions [4]. By collecting a large amount of facial expression data and conducting in-depth analysis, I can identify emotional patterns and facial features related to user feedback. This enables us to infer user preferences, satisfaction, and interests, and apply this information to the decision-making process of product improvement. By utilizing

B. H. Ahmad (ed.), *Proceedings of the 2023 International Conference on Data Science, Advanced Algorithm and Intelligent Computing (DAI 2023)*, Advances in Intelligent Systems Research 180,
https://doi.org/10.2991/978-94-6463-370-2_32

facial emotion recognition technology, manufacturers can better understand user needs and optimize product design and functionality.

This method can quickly capture users' emotional reactions, help manufacturers make timely adjustments and improvements, and provide a product experience that better meets user expectations. Through facial emotion recognition, manufacturers can gain more accurate insights into users' emotions and attitudes, thereby effectively improving product design and user experience.

In relevant research, many scholars and researchers have proposed various solutions to solve the problem of facial emotion recognition. They use different technologies and methods, such as computer vision [5], deep learning, and pattern recognition [6], to recognize and analyze facial expressions. These methods have achieved significant results, providing valuable insights for us to understand the relationship between facial emotions and user feedback.

This article aims to explore the application value and importance of facial emotion recognition in product improvement [7]. I established the correlation between facial emotions and user perceptions of products by collecting a large amount of facial expression data and using advanced model training techniques. By analyzing emotional patterns and facial features, I can infer user preferences, satisfaction, and interests, and apply this information to the product decision-making process. Through this research, I can provide manufacturers with deeper user insights, helping them optimize product design, improve functionality, and provide a better product experience.

2 Method

2.1 Pipeline:

A Pipeline is a data processing and machine learning system framework that typically consists of the following steps [8]:

Step1. Data Collection and Preprocessing: Obtain data and clean it, converting it into a format suitable for model usage.

Step2. Model selection and training: Select appropriate models or algorithms based on the characteristics of the data. After selecting the appropriate model, use the training data to train the model, and adjust the model parameters by learning the patterns and correlations of the data.

Step3. Model evaluation and tuning: In this step, the trained model is tested and evaluated using evaluation data.. Based on the evaluation results, the model can be optimized and improved.

2.2 ViT

The structure diagram of the ViT model. I used is shown as Fig, 1, which contains positional embedding, multi-head attention, encoder and decoder blocks. On the top, there is a multiple layer perceptron(MLP) head for classification task [9].

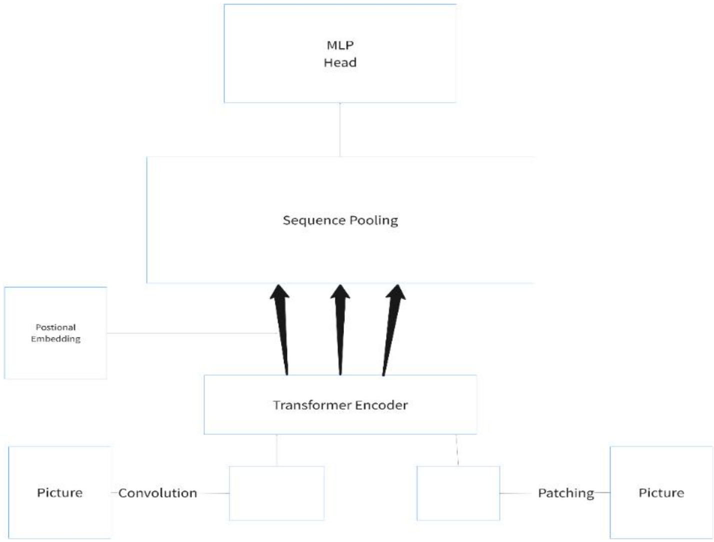

Fig. 1. VIT Structure :Original(Photo/Picture credit :Original)

2.3 Image preprocessing methods

Random cropping: Randomly crop images of different regions from the original image. This method can introduce changes in spatial position and differences in local details, increasing the diversity of data.

Color augmentation: By changing the color attributes of an image such as brightness, contrast, and saturation, the image presents different hues and appearances. This method can simulate different lighting conditions and environments, increasing the diversity of data.

Horizontal or vertical flipping: Flips the image horizontally or vertically to generate a mirrored image. This method can simulate different observation angles and mirror symmetry, increasing data diversity.

Random resizing: Randomly changing the size and scaling of an image. This method can simulate changes in distance, perspective, and scale, increasing data diversity.

Noise addition: Add random noise, such as Gaussian noise or salt and pepper noise, to the image. This method can simulate noise and interference in real scenes, increasing the robustness of the data.

Affine transformation: Changes the shape and posture of an image by performing affine transformation operations such as translation, rotation, scaling, and shearing. This method can simulate different perspectives, postures, and deformations, increasing the diversity of data.

3 Experiment

3.1 Data Set

I use the KMU-FED database, it has 20 kinds of objects for us, every kind of it has 4 types: Front light, Rear light, Left light and Right light， I use 10 kinds of objects for training, and use others for testing, shown as Table1 and Table2

Table 1. The training table

Light\Object	Object 1	Object 2	Object 3	Object 4	Object 5	Object 6	Object 7	Object 8	Object 9	Object 10
1	Front light	Front light	Front light	Front light	Front light	Front light	Front light	Front light	Front light	Front light
2	Rear light	Rear light	Rear light	Rear light	Rear light	Rear light	Rear light	Rear light	Rear light	Rear light
3	Left light	Left light	Left light	Left light	Left light	Left light	Left light	Left light	Left light	Left light
4	Right light	Right light	Right light	Right light	Right light	Right light	Right light	Right light	Right light	Right light

Table 2. The testing table

Light\Object	Object 11	Object 12	Object 13	Object 14	Object 15	Object 16	Object 17	Object 18	Object 19	Object 20
1	Front light	Front light	Front light	Front light	Front light	Front light	Front light	Front light	Front light	Front light
2	Rear light	Rear light	Rear light	Rear light	Rear light	Rear light	Rear light	Rear light	Rear light	Rear light
3	Left light	Left light	Left light	Left light	Left light	Left light	Left light	Left light	Left light	Left light
4	Right light	Right light	Right light	Right light	Right light	Right light	Right light	Right light	Right light	Right light

3.2 Image preprocessing

Image conversion: Use the Image. open() function in the PIL library to load an image file and convert it to RGB format to ensure an image with three channels (red, green, and blue).

Pre processing operations: For each image, apply the following pre processing operations in sequence:

RandomCrop: RandomCrop an area with a size of 224x224 from the image.

ColorJitter: Randomly adjust the brightness, contrast, saturation, and hue of an image.

RandomHorizontalFlip: RandomHorizontalFlip an image horizontally with a certain probability.

RandomVerticalFlip: RandomVerticalFlip an image with a certain probability.

RandomResizedCrop: Randomly crop and adjust the size of the image to 256x256 while maintaining the aspect ratio.

Add noise (Gaussian Blur): Apply Gaussian blur to the image with a blur radius of 3.

RandomAffine: Randomly applies affine transformations, including rotation, translation, scaling, and cropping.

Image conversion and normalization: Resize the preprocessed image to 256x256 pixels and convert it into tensor form. Then, normalize the image using the given mean and standard deviation.

Batch dimension: Finally, add a batch dimension to the preprocessed image and convert it into tensors with shapes of [1, C, H, W] by using the unsqueeze (0) operation on the tensor, where C is the number of channels, H and W are the height and width of the image, respectively.

3.3 Training process

Set training parameters such as batch size, learning rate, and number of training rounds. Load and preprocess training and test set data. Create a data loader for dividing data into batches and randomly shuffling. Define the architecture of the VIT model. Define loss functions and optimizers. Conduct a training cycle, traverse training data batches, and perform forward propagation, loss calculation, backpropagation, and weight updates. After each training round, use the test set for model validation and calculate accuracy. Print the loss and accuracy for each training round. There are some Parameter settings and roles, shown as Table3.

Table 3. Different Roles with Different Settings

Image_ Size	Patch_ Size	Num_ Classes	Dim	Depth	heads	Mlp_ Dim
The size of the input image	Divide the input image into xx sized image blocks	Assuming there are what facial expression categories that need to be classified.	Assuming there are what facial expression categories that need to be classified.	The encoder layer in the Transformer model is what layers.	The number of heads in the attention mechanism in the Transformer model is what	The hidden layer dimension of the Multi Layer Perception Machine (MLP) in the Transformer model is

	what

4 Experiment Result

Different from He, Zhang, Ren, and Sun (2016) introduced the concept of residual networks (ResNet) for image recognition. They demonstrated that deep residual learning could significantly improve the performance of image classification tasks [10].
I just use Pre processing operations, RandomCrop, ColorJitter, RandomHorizontalFlip, RandomVerticalFlip, Add noise and RandomAffine to do pre-processing. The result are shown in Fig2.

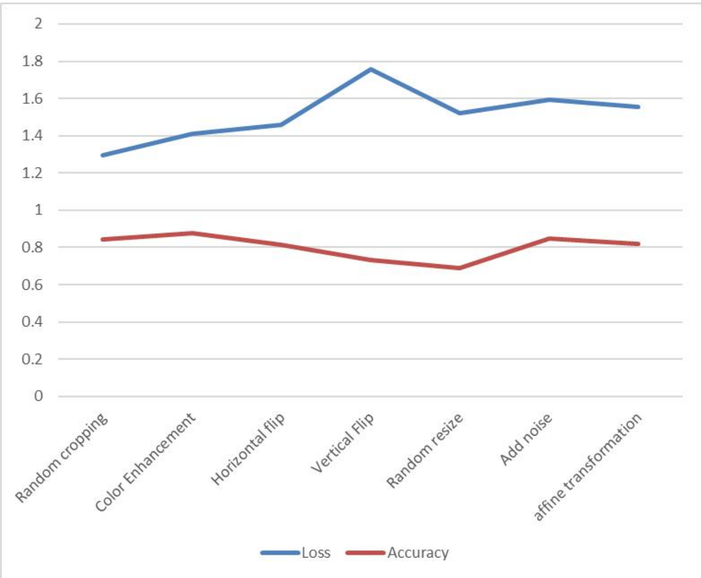

Fig. 2. 1. Line chart for Loss and Accuracy(Photo/Picture credit :Original)

Without any pretreatment ,the loss is about 1.69, and the accuracy is about 0.72. And with the pretreatment， the loss would lower than before and the accuracy would be higher.

5 Conclusion

The utilization of facial emotion recognition technology, particularly through the powerful Vision Transformer (ViT) model, addresses a significant challenge in understanding user emotions when they are unable to accurately express their views on products. This technology offers a nonverbal means of extracting emotional feedback, thereby providing valuable insights for product manufacturers to enhance user satisfaction. Our research in this paper focused on leveraging the VIT model to analyze facial expression datasets and experimenting with various preprocessing methods. Through this approach, I was able to associate each data sample with the corresponding expression. The key takeaway from our study was the comparison of results obtained from different preprocessing methods. Among these methods, RandomCrop emerged as the most effective in terms of accuracy for facial expression recognition. In essence, our work underscores the potential of Vision Transformer models in addressing real-world problems, such as understanding user emotions for product improvement. By selecting the most suitable preprocessing techniques, like RandomCrop in our case, I can significantly enhance the accuracy and reliability of facial expression recognition systems, ultimately benefiting both users and manufacturers alike. This research contributes to the ongoing development of technology that bridges the gap between human emotions and product feedback, thereby facilitating more customer-centric product design and development.

Reference

1. Jacobs M H. Human emotions toward wildlife[J]. Human Dimensions of Wildlife, 2012, 17(1): 1-3.
2. Simanjuntak M B, Lumingkewas M S. REPRESENTATION OF LONGING THROUGH THE LYRICS OF THE SONG" MOTHER HOW ARE YOU TODAY"[C]//Prosiding Seminar Nasional Inovasi Pendidikan. 2022.
3. Xu T, White J, Kalkan S, et al. Investigating bias and fairness in facial expression recognition[C]//Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part VI 16. Springer International Publishing, 2020: 506-523.
4. Pantic M, Patras I. Dynamics of facial expression: recognition of facial actions and their temporal segments from face profile image sequences[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2006, 36(2): 433-449.
5. Ammar L B, Gasmi K, Ltaifa I B. ViT-TB: Ensemble Learning Based ViT Model for Tuberculosis Recognition[J]. Cybernetics and Systems, 2022: 1-20.
6. Li Y. Research and application of deep learning in image recognition[C]//2022 IEEE 2nd International Conference on Power, Electronics and Computer Applications (ICPECA). IEEE, 2022: 994-999.
7. Priesmeyer H R, Mudge S. Using Measures of Emotions to Improve Work Climate, Products, and Decision-Making[J]. Management & Marketing, 2008, 3(1): 17-32.
8. Bertin E, Mellier Y, Radovich M, et al. The TERAPIX pipeline[C]//Astronomical Data Analysis Software and Systems XI. 2002, 281: 228.
9. Pinkus A. Approximation theory of the MLP model in neural networks[J]. Acta numerica, 1999, 8: 143-195.
10. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.