World Journal of Advanced Engineering Technology and Sciences

eISSN: 2582-8266

Cross Ref DOI: 10.30574/wjaets

Journal homepage: https://wjaets.com/

(Research Article)

Facial expression recognition with modified local ternary pattern images using convolutional neural network and extreme learning machine

Kailas Nath V D [1, *] and Jolly B [2]

[1] *Department of Electronics and Communication Engineering, College of Engineering Trivandrum, Trivandrum, Kerala, India.*

[2] *Department of Electronics and Communication Engineering, Government Polytechnic College Ezhukone, Kollam, Kerala, India.*

World Journal of Advanced Engineering Technology and Sciences, 2024, 13(02), 624-631

Publication history: Received on 13 November 2024; revised on 22 December 2024; accepted on 24 December 2024

Article DOI: https://doi.org/10.30574/wjaets.2024.13.2.0630

Abstract

Facial emotion recognition is a crucial aspect of human-computer interaction systems, education, health and various other fields. The common challenges faced in FER are variation in light. Based on researches, Local ternary Pattern (LTP) as a feature vector that can handle the variation of light. Modified Local Ternary Pattern (MLTP) is more discriminative to variation of light and less sensitive to noise. MLTP is a conventional feature vector that requires manual processing. In contrast to MLTP, the Convolution Neural Network (CNN) architecture has an automated feature extractor. This paper proposes MLTP as input to the CNN architecture for handling the variation of light and for automatic feature extractor. Then, for classifying Extreme Learning Machine is used to reduce the training time of the CNN classifier. The Karolinska Directed Emotional Faces (KDEF) dataset is used to assess the proposed method and got an accuracy of 86.28%.

Keywords: Modified Local Ternary Pattern; Convolutional Neural Network, Extreme Learning Machine; KDEF

1. Introduction

Facial Emotion Recognition (FER) is an interesting task to explore in computer vision. It is possible to identify and comprehend human emotions by looking at facial expressions. Automatic Facial Emotion Recognition is used in many real life application like education, mental health and human-computer interaction [1]. FER can be developed using two approaches; convolutional approach and deep learning approach [2]. In recent researches, the convolutional approach does not perform well when handling variation of light. In deep learning approach, it produces high accuracy in case of image classification but need a large data for learning process.

CNN is the most commonly used classifier for object and image classification. But the back propagation algorithm (BP) which is mostly used for training the CNN suffers poor generalization and slow learning. Studies on various tasks verified that by using CNN and ELM combination, the training time can be minimized [3]. In this method, the CNN is used as a feature extractor [4] and ELM is used as the classifier. The impact of this method is verified later in FER task by Jammoussi et al. [5]. However, their work is to test only the effectiveness of CNN-ELM combination in FER without treating the variation of lights. Recent Works shows that LTP can handle variation of light. When compared with LTP, modified LTP is less sensitive and more discriminating to noise [6, 7].

By considering the above facts, we propose a new model by combining the CNN-ELM with MLTP as input. The main idea is combine the two methods so that each may train more quickly and handle fluctuations in light and noise. We use KDEF dataset as it is known for the variations of light and it contains seven emotion class.

[*] Corresponding author: Kailas Nath V D

The remaining of the paper discusses about the following. Section 2 discuss about the previous related works, Section 3 discuss about the methodology that is used in this research. Section 4 shows our results and discussions about them. Finally, section 5 contains the conclusion of this research.

2. Literature Study

FER consists of three main steps: pre-processing, feature extraction and classification [8]. Pre-processing is done to improve the classification performance which contains face detection, normalization or correction. There are many face detection methods like Haar Feature, Seeta Face and Local Assembled Binary(LAB) but Dlib face detection approach is known for its computationally low [9]. The normalization or correction step is used to handle noise, light variation and size.

Feature extraction is an important part in FER to gain good results. It is used for obtaining salient information. Local Binary Pattern (LBP) is one of the effective and local feature descriptor which gains attraction in computer vision and image processing [10, 11]. LBP descriptor is used many real time applications such as medical image analysis, facial image recognition and facial expression recognition. To increase the robustness in noise many LBP variations are designed which include Local Ternary Pattern (LTP) and Modified Local Ternary Pattern (MLTP) [12]. In deep learning, feature extraction is done automatically. CNN has the ability to simultaneously classify input images by using a neural network to extract their various levels.

The final step in FER is to choose an appropriate classifier with the dataset. Variations of CNN are regarded as the robust classifier since they outperform traditional MLP. Backpropagation (BP), which is known to result in slower learning, is used by CNN for training based on gradient descent. While there are researchers that suggest converting 2D convolution into 1D convolution or employing GPUs in parallel, doing so can lead to higher processing costs and perhaps lower accuracy. Using a single hidden layer feedforward neural network (SLFN), the ELM classifier executes generalised inverse on the complete set of data once, preventing BP from learning slowly.

3. Proposed Methodology

This project runs in Python and uses the lab-controlled dataset available in Google Colaboratory. Prior to being transformed into MLTP images, the original images undergo preprocessing. After preprocessing, the images are converted into MLTP and become the input of CNN. After being received by CNN, the attributes are entered into ELM for classification. Figure 1 shows the architecture of the proposed method.

3.1. Dataset

This project makes use of the Karolinska Directed Emotional Faces (KDEF) dataset [13]. Seven class expressions are included in the dataset: afraid, angry, disgusted, happy, neutral, sad, and surprised. This study uses only the frontal images, i.e., 980 images. The original size of KDEF dataset is 562*762*3. Figure 2 shows some sample images in the KDEF dataset.

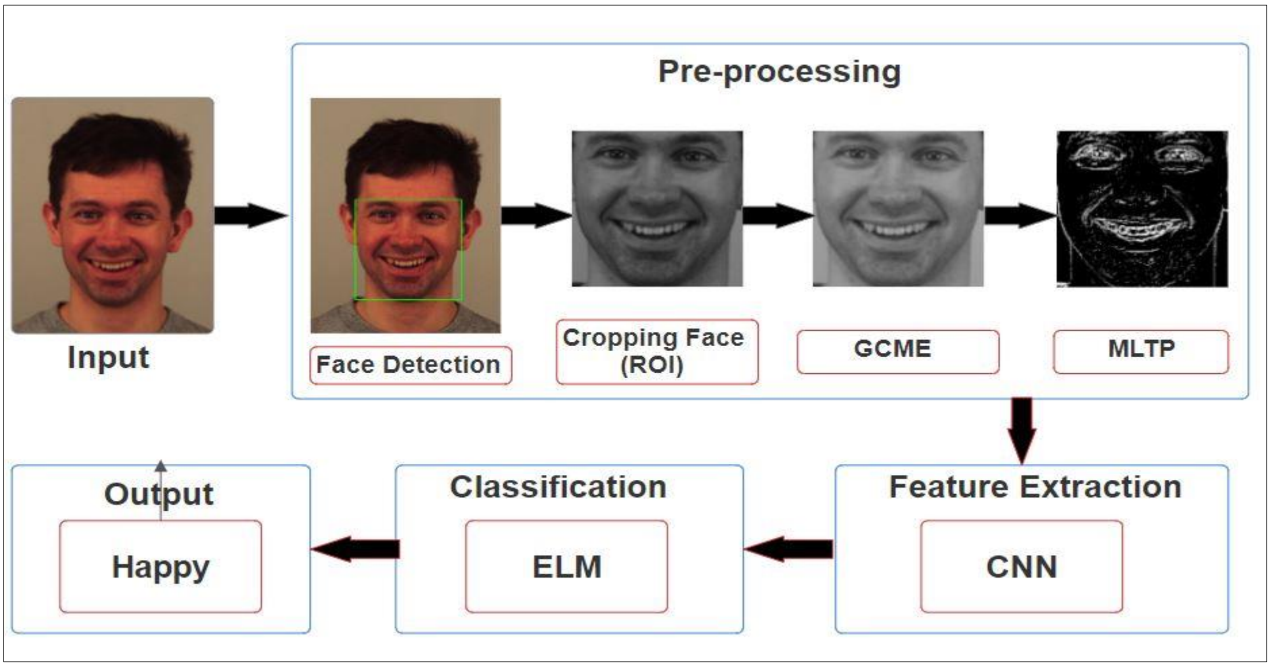

Figure 1 Architecture of the Proposed Methodology

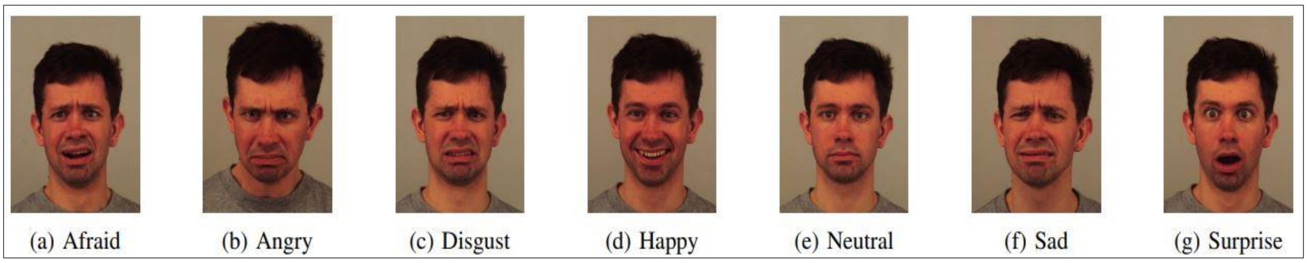

Figure 2 Sample Images of KDEF dataset

3.2. Data Pre-processing

In FER, face detection helps to define the region of interest (ROI). We use Dlib for facial recognition and landmarking. Furthermore, landmarking nodes are used to crop the image. To standardize the images and save processing costs, the images are cropped to 128*128 size and then turned to grayscale. Then, to reduce the impact of light changes, we employ the gamma correction method (GCME), formerly known as the Adaptive Gaussian Transformation Method (AGT-Me) [14]. The algorithm describes the GCME preparation processes as follows:

Input: the grayscale image of KDEF (I)

Output: corrected image G (I)

1. Using min-max normalization, normalize the image.
2. Use (1) to normalize the image.
3. Calculate the optimum gamma γ∗
4. Use γ∗ to adjust for gamma.
5. Obtain the updated image G (I).

$$(I_m + 0,5)/256 \, , \; m = 0,1,2, \dots \dots \dots , 255 \dots \dots (1)$$

3.3. Modified Local Ternary Pattern

LTP is an improvement of the Local Binary Pattern feature extraction methodology. It was created by Tan and Triggs to address issues with altering noise and lighting levels [15]. The basic idea of the modified LTP and the original LTP is mostly the same. However, this experiment did not use MLTP as a feature extractor. After the GCME processing of the image, we compute the upper and lower patterns of the original LTP using (2). To obtain the most significant characteristic, we applied the average calculation to the LTP operator, which is the concatenation of the upper and lower pattern codes. The final result of the modified LTP is an image.

$$f_i(x, gc, t) = \begin{cases} +1, & x \geq gc + t \\ 0, & |x - gc| < t \\ -1, & x \leq gc - t \end{cases} \quad \ldots\ldots\ldots (2)$$

3.4. Feature Extraction

A key step in the categorization process is feature extraction. Techniques for feature extraction that require physical work have been proposed. Consequently, we do not employ the conventional strategy, but rather leverage CNN's capability as an automatic feature extractor. There are five convolution layers, four pooling layers and two dropout layer.

3.4.1. Convolutional Layer

A key component of CNN's architecture that allows it to extract various levels from input is convolution. The convolution process is explained using the following quotations.

$$A|i,j| = \sum_{m=0}^{m-1} \sum_{n=0}^{n-1} F(m,n)\ X\ l(i-3, j-n) \ldots\ldots\ldots (3)$$

The feature map W_c × H_c× D_c that results from the convolution process can be constructed as follows:

$$W_c = \frac{W_{in} - F_w + 2P}{S} + 1$$

$$H_c = \frac{H_{in} - F_h + 2P}{S} + 1 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (4)$$

$$D_c = K$$

Where the input sizes are W_{in}, H_{in} and D_{in}. F stands for filter slides applied to the image, K for the number of filters chosen for convolution, and P for padding.

3.4.2. Rectified Linear Unit (ReLU)

ReLU, the default activation function of CNN, can handle vanishing gradient difficulties, which makes it faster than alternative activation functions. Consequently, it may produce superior results than others. It has the following definition:

$$\begin{cases} \max(0, x), & x \geq 0 \\ 0, & x < 0. \end{cases} \ldots\ldots\ldots\ldots\ldots\ldots (5)$$

3.4.3. Pooling Layer

Pooling was applied combined with a down-sampling procedure to decrease the output size's dimension. To extract the salient feature, use CNN's pooling layer. Due of its superior performance, max pooling is utilized more frequently than sum pooling.

The following is an expression for the pooling operation:

$$W_c = \frac{W_{in} - F_w}{S} + 1$$

$$H_c = \frac{H_{in} - F_h}{S} + 1 \ldots\ldots\ldots\ldots\ldots\ldots (6)$$

$$D_c = D_{in}$$

3.4.4. Dropout Layer

Dropout is a regularization technique that is used for avoiding overfitting by randomly turning a portion of the input units to zero during training. The Dropout layer can be represented as follows:

Let the input to this layer

$$X = [x_1, x_2, \dots \dots \dots, x_n] \text{.......................}(7)$$

And a binary mask

$$M = [m_1, m_2, \dots \dots \dots, m_n] \text{...........}(8)$$

Then, the output Y is given by

$$Y = \frac{X.M}{1-p} \text{................}(9)$$

Where p is the Bernoulli distribution probability.

3.5. Extreme Learning Machine

Guan Bin Huang proposed the categorizing approach known as ELM [16]. Since ELM only uses one hidden layer, it can perform well overall 1,000 times quicker than backpropagation networks. To link the input and hidden layer, the method begins by creating the random weights matrix W and bias b, which are explained as follows:

$$\boldsymbol{W} = \begin{bmatrix} W_{11} & \cdots & W_{1n} \\ \vdots & \ddots & \vdots \\ W_{L1} & \cdots & W_{Ln} \end{bmatrix}_{L\times n} \text{..............................}(10)$$

$$\boldsymbol{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_L \end{bmatrix}_{L\times 1} \text{........................}(11)$$

Where:

L represents hidden nodes.

n is the number of inputs.

The training data X is multiplied by the weighting matrix's transpose to obtain the hidden layer outputs.

$$H_{init} = X.W^T \text{...................}(12)$$

The sigmoid activation function is defined by:

$$H = \frac{1}{1+e^{H_{init}}} \text{.....................}(13)$$

The output of weight matrix β is calculated as:

$$\beta = (\frac{1}{C} + HH^T)^{-1}H^TT \text{...................}(14)$$

Where:

C represents regularization parameter.
T represents desired output matrix.

4. Experiments and Discussion

In this section, the comprehensive evaluation findings of the chosen methodology for evaluating the performance of kdef dataset is shown. It also presents the comparative outcome of the recent FER work.

4.1. Evaluation Criteria

The performance of deep learning and machine learning models may be objectively evaluated with the help of evaluation metrics. The evaluation metrics used for the implementation are as follows:

1. **Accuracy**: Accuracy calculates the proportion of accurately predicted instances to all instances in the dataset.

$$Accuracy = \frac{1}{m}\sum_{i=1}^{m} g(f(x_i) = y_i) \text{……….(15)}$$

2. **Precision**: Precision measures how accurately positive predictions are made. It calculates the proportion of true positives to all positive predictions.

$$Precision = \frac{TP}{TP+FP} \text{…………..(16)}$$

3. **Recall**: Recall determines the ratio of actual positive cases in the dataset to all true positives.

$$Recall = \frac{TP}{TP+FN} \text{………………(17)}$$

4. **F1-Score**: The harmonic mean of recall and precision is the F1 score.

$$F1 - Score = Score = 2 \times Prec \times Rec/Prec + Rec$$

Where:
TP: true positive, denotes elements properly defined as positive.
TN: true negative, denotes elements properly defined as negative.
FP: false positive, denotes positive elements that were classified incorrectly.
FN: false negative, denotes negative elements that were classified incorrectly.

In order to adjust for light changes, we used the GCME approach before converting images to MLTP images. We utilise training and testing data in the same order, with an 80:20 split ratio for each technique, to maintain fairness. The best result for our proposed model is 86.28%. Table 1 shows the detailed architecture that we used. We use CNN as feature extractor and ELM as classifier.

Table 2 shows that comparison of accuracy. Compared with the previous studies, our model yielded 5.13% higher accuracy than achieved by [6]. When compared with the local ternary pattern, our model yielded 0.77% higher accuracy than achieved by [1]. The best accuracy achieved by the authors of [5] is 85.90%. When compared with our model, we got 0.38% higher than achieved by [5].

Table 3 shows the performance of each class. Based on the table shown, the higher recognition rate based on precision, recall and F1 score is happy. This is because happy expression is dissimilar from the other class expression and can be easily identified. The lowest recognition rate is afraid because it is similar to the other class.

Table 1 Detailed Architecture

Layer	Parameter
Input	128, 128, 1
Conv_1	Filter:5×5,32;Stride:1×1;ReLU
Max_Pool_1	Kernel: 2 × 2
Conv_2	Filter:5×5,32;Stride:1×1;ReLU
Max_Pool_2	Kernel: 2 × 2

Conv_3	Filter:5×5,32;Stride:1×1;ReLU
Max_Pool_3	Kernel: 2 × 2
Dropout_1	Probability: 0.2
Conv_4	Filter:5×5,64;Stride:1×1;ReLU
Max_Pool_4	Kernel: 2 × 2
Dropout_2	Probability: 0.1
Conv_5	Filter:4×4,64;Stride:1×1;ReLU
Fully Connected	ELM
Output	Predicted Class

Table 2 Comparison with State

Input Images	Feature Extraction	Classifier	Accuracy (%)	Training Time (sec)
LTP	CNN	CNN	8061	4866.22
	CNN	ELM	85.51	187.68
MLTP	CNN	CNN	81.15	585.642
	CNN	ELM	86.28	21.024

Table 3 Performance of Each Emotion Class on the KDEF Dataset

Class	Precision	Recall	F1-Score
Afraid	0.77	0.76	0.76
Angry	0.89	0.88	0.88
Disgust	0.86	0.87	0.86
Happy	0.97	0.99	0.97
Neutral	0.91	0.89	0.89
Sad	0.79	0.78	0.78
Surprise	0.86	0.96	0.90

5. Conclusion

This paper proposes a novel FER method with modified local ternary pattern using CNN and ELM in KDEF dataset. In this model, CNN is used as the feature extractor. We used ELM for classification instead of BP in CNN to reduce the training speed. The proposed method was tested and got the best accuracy of 86.28%.

Compliance with ethical standards

Disclosure of conflict of interest

The authors declare no conflict of interest.

References

[1] Krisnahati, Ice, Nanik Suciati, and Shintami Chusnul Hidayati. "Face expression recognition with local ternary pattern images using convolutional neural network and extreme learning machine." 2022 6th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE). IEEE, 2022.

[2] Li, Stan Z., et al. "Facial expression analysis." Handbook of face recognition (2005): 247-275.

[3] Yoo, Youngwoo, and Se-Young Oh. "Fast training of convolutional neural network classifiers through extreme learning machines." 2016 International Joint Conference on Neural Networks (IJCNN). IEEE, 2016.

[4] Jogin, Manjunath, et al. "Feature extraction using convolution neural networks (CNN) and deep learning." 2018 3rd IEEE international conference on recent trends in electronics, information \& communication technology (RTEICT). IEEE, 2018.

[5] Jammoussi, Imen, Mounir Ben Nasr, and Mohamed Chtourou. "Facial Expressions Recognition through Convolutional Neural Network and Extreme Learning Machine." 2020 17th International Multi-Conference on Systems, Signals \& Devices (SSD). IEEE, 2020.

[6] Zulkarnain, Syavira Tiara, and Nanik Suciati. "Modified Local Ternary Pattern With Convolutional Neural Network for Face Expression Recognition." Jurnal Ilmiah Teknologi Informasi 19.1 (2021): 10-18.

[7] Ren, Jianfeng, Xudong Jiang, and Junsong Yuan. "Relaxed local ternary pattern for face recognition." 2013 IEEE international conference on image processing. IEEE, 2013.

[8] Huang, Yunxin, et al. "Facial expression recognition: A survey." Symmetry 11.10 (2019): 1189.

[9] D.E. King, "Dlib-ml: A machine learning toolkit", J. Mach. Learn. Res. 10, pp. 1755-1758, 2009.

[10] L. Liu, P. Fieguth, G. Zhao, M. Pietikäinen, and D. Hu,"Extended local binary patterns for face recognition," Inf Sci (N Y), vol. 358–359, pp. 56–72, Sep. 2016, doi: 10.1016/j.ins.2016.04.021.

[11] Chahi, A., Y. Ruichek, and R. Touahni. "Local directional ternary pattern: A new texture descriptor for texture classification." Computer vision and image understanding 169 (2018): 14-27.

[12] Rangsee, Pattarakamon, K. B. Raja, and K. R. Venugopal. "modified local ternary pattern based face recognition using SVM." 2018 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS). Vol. 3. IEEE, 2018.

[13] Lundqvist, D., A. Flykt, and A. Öhman. "The Karolinska directed emotional faces (KDEF), 1998." Department of Neurosciences Karolinska Hospital: Stockholm, Sweden (1998).

[14] Lee, Yong, et al. "Blind inverse gamma correction with maximized differential entropy." Signal Processing 193 (2022): 108427.

[15] Tan, Xiaoyang, and Bill Triggs. "Enhanced local texture feature sets for face recognition under difficult lighting conditions." IEEE transactions on image processing 19.6 (2010): 1635-1650.

[16] Huang, Guang-Bin, Qin-Yu Zhu, and Chee-Kheong Siew. "Extreme learning machine: a new learning scheme of feedforward neural networks." 2004 IEEE international joint conference on neural networks (IEEE Cat. No. 04CH37541). Vol. 2. Ieee, 2004.