Genetic Programming-Based Feature Learning for Facial Expression Classification

Ying Bi, Bing Xue and Mengjie Zhang
School of Engineering and Computer Science,
Victoria University of Wellington, PO Box 600, Wellington 6140, New Zealand
Email:{Ying.Bi, Bing.Xue, Mengjie.Zhang}@ecs.vuw.ac.nz

***Abstract*—Facial expression classification is an important but challenging task in artificial intelligence and computer vision. To effectively solve facial expression classification, it is necessary to detect/locate the face and extract features from the face. However, these two tasks are often conducted separately and manually in a traditional facial expression classification system. Genetic programming (GP) can automatically evolve solutions for a task without rich human intervention. However, very few GP-based methods have been specifically developed for facial expression classification. Therefore, this paper proposes a GP-based feature learning approach to facial expression classification. The proposed approach can automatically select small regions of a face and extract appearance features from the small regions. The experimental results on four different facial expression classification data sets show that the proposed approach achieves significantly better results in almost all the comparisons. To further show the effectiveness of the proposed approach, different numbers of training images are used in the experiments. The results indicate that the proposed approach achieves significantly better performance than any of the baseline methods using a small number of training images. Further analysis shows that the proposed approach not only selects informative regions of the face but also finds a good combination of various features to obtain a high classification accuracy.**

I. Introduction

Understanding facial expressions of motion is important for human communication [1]. Ekman and Friesen [2] classified facial expressions into six groups: surprise, fear, disgust, anger, happiness, and sadness. By including the neutral category, there are totally seven well-known facial expressions. The classification of facial expressions is often based on images, which can be easily obtained or sampled by a sensor or camera. Facial expression classification based on images refers to classify face images with different expressions into one of the predefined groups. Facial expression classification is a challenging task, as it aims to analyse abstract and high-level content of face images. It has a wide range of applications in computer vision and pattern recognition, such as emotion analysis, law enforcement, and interactive video [3].

Typically, algorithms of facial expression classification have two main steps: feature extraction and expression classification [4]. The first step is to extract informative features such as geometric and appearance features from the images. The geometric features usually select a large number of facial fiducial points from the face images and extract features based on these points [5]. Therefore, the precise location of such points in a face image is important for obtaining effective geometric features. The appearance features aim to extract the appearance of the face, such as shape and texture [5]. These features include the local binary patterns (LBP), histogram of oriented gradients (HOG), and Gabor features [1]. Then, facial expression classification builds a stable classification system to classify different expressions based on the extracted features. Commonly used methods are support vector machines (SVMs) [6], k-nearest neighbour (*k*NN) [7], and sparse representation-based classification (SRC) [8]. However, in these above processes, feature detection and extraction are manually performed, which require rich domain knowledge and are time-consuming.

Instead of manually extracting features, many methods have been developed to automatically extract features from images for classification [5]. Typical methods are convolutional neural networks (CNNs) and genetic programming (GP) [9, 10]. In these methods, features are automatically learned/extracted and then classification is performed using these features on a training set. Based on the classification performance of the training set, these methods can search for the best solutions/features via the learning process. Thus, these methods are more effective for different image classification tasks than the methods using manually extracted features [11]. However, most of the existing methods are neural network (NN)-based methods, which often require a large amount of training data [9, 10]. Therefore, this paper aims to develop a non-NN-based method for facial expression classification, even using a small number of training images.

Genetic programming (GP) is an evolutionary algorithm with powerful search/learning ability. GP can automatically evolve computer programs to solve a problem without predefined solution structure [12]. Compared with other evolutionary algorithms, GP has a flexible representation, e.g., tree-based representation, enabling it to find solutions with variable depths. GP has been applied to many tasks and achieved promising results, such as symbolic regression, classification, clustering, scheduling, and image analysis [10, 13].

In recent years, GP has been widely applied to image classification [10]. Existing works have shown that GP can extract effective and domain-specific features for object classification, texture image classification, and scene classification [9, 11, 14, 15]. However, very few GP-based methods have been developed for facial expression classification. The methods such as [9, 11, 14] could be used for facial expression

978-1-7281-6929-3/20/$31.00 ©2020 IEEE

classification, but they may not be effective and efficient because they were developed for different tasks. The features effective for texture classification may not be effective for facial expression classification. Therefore, it is necessary to develop new GP-based feature learning methods based on the task type, i.e., facial expression classification.

The goal of this paper is to develop a new GP-based feature learning approach for facial expression classification. The new approach can automatically select small regions of images and extract informative appearance features from these regions. To achieve this, a program structure, a function set and a terminal set are developed in the new approach. For simplification, the new approach is termed as facialGP. The performance of the facialGP approach will be examined on four different facial expression classification data sets using various numbers of training images. A number of algorithms will be used as baseline methods to show the effectiveness of the facialGP approach. In addition, further analysis of the example program evolved by facialGP will be conducted to provide more insights into it.

II. Related Work

A. Facial Expression Classification

A general procedure of facial expression classification is shown in Fig. 1. The procedure consists of image preprocessing, feature extraction, dimensionality reduction, and classification. In this procedure, face detection aims to find the location of the face in the image, which can be conducted by finding some landmarks or points in the face [16]. From the detected face or points, different features can be extracted for solving facial expression classification. Chao et al. [4] improved LBP to extract specific facial features by emphasising local information of face images. Then a dimensionality reduction approach is employed to reduce the dimension of the extracted LBP features. The use of LBP features can also be seen in [17], where a pairwise feature selection method is employed to reduce the dimension of features. Scale-invariant feature transform (SIFT) has also been applied to extract features from the detected keypoints for facial expression classification [18].

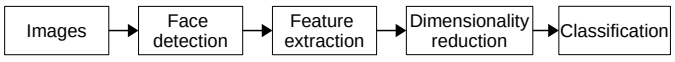

Fig. 1. A general procedure of facial expression classification.

In general, high-dimensional features are extracted from a large number of landmarks/points [17]. To effectively and efficiently solve facial expression classification, dimensionality reduction techniques are applied to reduce the dimension of features before classification. Commonly used methods are principal component analysis (PCA) [19] and linear discriminant analysis (LDA) [16], which are unsupervised and supervised methods, respectively.

The aforementioned methods have achieved good performance on facial expression classification data set, such as the Japanese Female Facial Expression (JAFFE) database [20]. However, careful designs of face detection and feature extraction are needed to achieve successful facial expression classification. In contrast, representation learning algorithms, i.e., deep learning algorithms, have been proposed to automatically learn features for classification without domain knowledge. Zhao et al. [5] combined deep belief networks (DBNs) and multilayer perceptron (MLP) for facial expression classification. Lopes et al. [21] applied convolutional neural networks (CNNs) for facial expression classification, where a preprocessing step was employed to extract facial-specific features. However, deep learning algorithms often require a large number of training images to train the models. To this end, this study explores a non-NN-based algorithm for feature learning to solve facial expression classification.

B. GP for Feature Learning and Image Classification

The solutions of GP are often represented by trees, where internal nodes are functions and leaf nodes are features/variables. An example tree is shown in Fig. 2. This tree can be formulated as $(x_1 * x_2 - x_3) + x_1 + 0.9$.

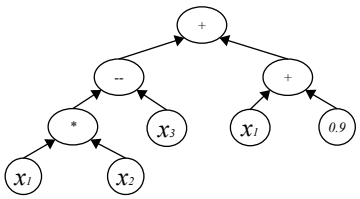

Fig. 2. An example GP tree.

Based on the flexible tree-based representation, many GP-based algorithms have been developed for image classification [13]. Atkins et al. [22] proposed a multi-tier GP for simultaneous image filtering, region detection, feature extraction, feature construction, and classification. The multiple processes of image classification have been integrated into a single GP tree using a strongly typed version of GP [23]. Al-Sahaf et al. [24] developed a two-tier GP method by removing the image filtering tier of the multi-tier GP method and improving the region detection functions. The two-tier GP method is generally faster than the multi-tier GP for image classification. To extract informative features, Lensen et al. [25] employed the HOG descriptor as a function node of GP trees and developed a GP-HOG method for image classification. However, these methods have only been examined on binary image classification. Shao et al. [11] developed a GP-based feature learning algorithm for image classification with simultaneously maximizing the classification accuracy and minimising the tree size. This method employed a number of image filters and max-pooling operators as functions so that domain-specific features can be extracted from images. Bi et al. [9] developed a GP-based method with convolution operators for feature learning. The filters and the size of filters can be automatically selected by this GP method, which is more flexible than that in CNNs. The performance of this method has been examined on six different data sets. Although a number of GP-based

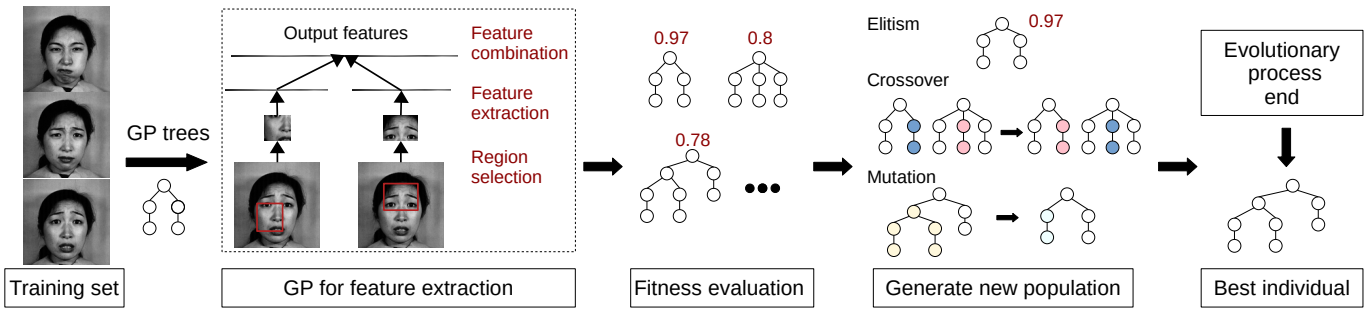

Fig. 3. Outline of the proposed facialGP approach.

methods have been developed for image classification, none of them is, particularly, for facial expression classification. Effectively solving facial expression classification needs local region selection and appearance feature extraction. These can be integrated into a single GP tree to achieve automatic local region selection and feature extraction. Therefore, this paper develops a GP-based feature learning algorithm for facial expression classification.

III. The Proposed Approach

This section describes the details of the proposed facialGP approach. First, the overall algorithm is outlined. Second, it introduces the program structure, the function set and the terminal set of facialGP. Finally, the fitness function and the test process are presented.

A. Overall Algorithm

The overall algorithm of facialGP is shown in Fig. 3. The facialGP approach has a population of individuals/solutions and searches for the best solution(s) through an evolutionary process. Each individual of facialGP can perform region selection, feature extraction and feature combination. Thus, given an image to an individual, the output of the individual is a set of features. The features are extracted using some feature extraction methods from the locally detected regions. These features can be used to feed into a classification algorithm for expression classification.

At the initialisation step, facialGP randomly initialises a population of individuals/trees using a tree generation method. At each generation, the population is evaluated using a fitness function, which will be introduced in the following subsection. The elitism operator directly copies the best individuals into the next generation. The crossover and mutation operators are employed to generate new offspring for the next generation. Tournament selection is often used in GP to select the individuals to form the parents for crossover and mutation. If the termination criterion is satisfied, the evolutionary process will be terminated and the best individual will be returned.

B. Program Structure

The facialGP approach uses a simple program structure to perform region selection, feature extraction, and feature combination. An example program is shown in Fig. 4. This

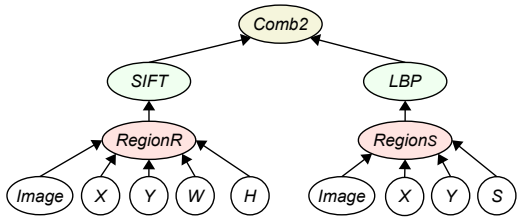

Fig. 4. An example program to show the program structure of facialGP.

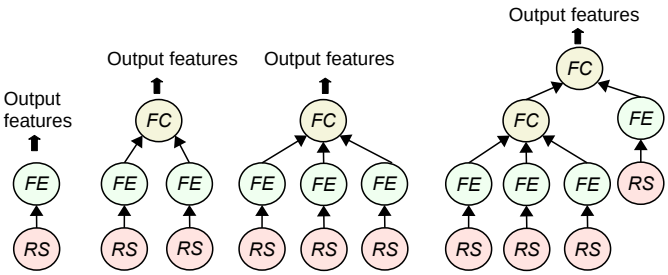

Fig. 5. Possible program structures of facialGP.

example program can be used to extract features as shown in Fig. 3, i.e., selecting two local regions, extracting features from these two regions, respectively, and combining these features to form the output features.

With this design, facialGP can find local regions of the face using region selection functions during the evolutionary process. The regions could contain specific expression features, which may help achieve a good expression classification performance. Besides region selection, another important step is feature extraction. In a traditional facial expression classification system, many appearance features have been extracted, such as LBP and SIFT features [4, 17]. Therefore, facialGP uses these methods to extract features from the automatically selected regions. The features extracted by different functions are combined to form the final output features.

Fig. 5 shows possible program structures of facialGP. In this figure, RS represents region detection function, FE represents feature extraction, and FC represents feature combination. From this figure, it can be found that the complexity and depth of the program/individual are flexible. A simple program (e.g., the left one in Fig. 5) only selects one region of the face and uses a feature extraction function to extract features from the region. A complex program (e.g., the right one in Fig. 5) may

have a large number of nodes to select multiple regions and extract high-dimensional features. This design allows facialGP to evolve solutions/programs with variable depths for feature extraction in facial expression classification.

C. Function Set and Terminal Set

Based on the program structure, there are three types of functions in the proposed facialGP approach. The functions are two region selection functions, three feature extraction functions and two feature combination functions.

Region Selection Functions: The two region selection functions are termed as $RegionS$ and $RegionR$, where S indicates square and R indicates rectangle. These two functions can select square and rectangle regions from the image. The $RegionS$ function has four arguments: $Image$, X, Y, and S. The $RegionR$ function has five arguments: $Image$, X, Y, W, and H. The $Image$ represents the input image, where the features will be extracted from. X and Y represents the top left coordination of the selected region in the image. The $S \times S$ and $W \times H$ represent the size of the selected region in the two functions. Given these arguments, the $RegionS$ function returns a small square region, i.e., $Image[X : min(width,\ X+S),\ Y : min(height,\ Y+S)]$. Given five arguments, the $RegionR$ function returns a small rectangle region, i.e., $Image[X : min(width,\ X+W),\ Y : min(height,\ Y+H)]$. Note that $width$ and $height$ represent the width and height of the image.

Feature Extraction Functions: To narrow the search space and to allow facialGP to search for more specific expression features, three different feature extraction functions are employed. These functions are $SIFT$ [26], LBP [27] and $Concatanation$. The $SIFT$ function [26] produces 128 SIFT features of gradient magnitude and orientation from a selected region. The SIFT features aim to extract shape features of the face. The LBP function [27] produces 59 uniform LBP features from a selected region. The LBP features aim to extract texture features of the face. The $Concatanation$ function concatenates each array of the selected region and returns a feature vector. The $Concatanation$ function does not produce new features by returning the raw pixel values. The raw pixel values may be informative for classification. It is noted that the $SIFT$ and LBP operators have been used in [28], but the facialGP method uses less operators, leading to a smaller search space.

Feature Combination Functions: Feature combination functions concatenate the features from different child nodes into a feature vector to form the output features. Two feature combination functions are $Comb2$ and $Comb3$, which take two and three arguments, respectively. The feature combination functions allow facialGP to produce a combination of different features from various regions.

Terminals: Six terminals, $Image$, X, Y, W, H, and S, are employed in the proposed facialGP approach. The $Image$ represents the input image, which is a 2D array with values in the range of $[0,\ 1]$. Note that the pixel values of the image are scaled from $[0,\ 255]$ into $[0,\ 1]$. The other terminals are the parameters for the $RegionS$ and $RegionR$ functions, which have been introduced in the previous subsection. The ranges of X and Y are set to $[0,\ width-10]$ and $[0,\ height-10]$. The range of the W, H and S terminals are set to $[10,\ 30]$, which allow the size of the selected region to be between 10×10 and 30×30.

D. Fitness Function

To evaluate the features produced by facialGP, a classification algorithm, linear SVM, is employed for classifying expressions. Linear SVM is more popular for image classification than the other classification algorithms [11]. A training set, consisting of images and labels, are employed for evaluation during the evolutionary process. To improve the generalisation performance of the features learned by facialGP, five-fold cross-validation on the training set is employed in the fitness evaluation of facialGP. In the fitness evaluation process, the produced features by facialGP are normalised using the min-max normalisation method. The normalised features and the class labels are split into five folds. Each time four folds are used to train the SVM and the remaining one fold is used to test the classifier. The mean classification accuracy of the five folds is set as the fitness value of each individual.

E. Test Process

The test process evaluates the performance of the best individual found by the proposed facialGP approach on the unseen test set. The test set has not been used during the evolutionary process. In the test process, the training set, which is the same as that used in the evolutionary process, and the test set is fed into the best individual to obtain the features. Then the features of training and test sets are normalised. The normalised features and class labels of the training are used to train the linear SVM and the trained classifier is used to classify the test set. The classification accuracy of the test set is obtained and reported.

IV. Experiments Design

This section designs the experiments that are conducted to examine the performance of the proposed facialGP approach on different facial expression classification data sets.

A. Baseline Methods

To show the effectiveness of the facialGP approach, nine different baseline methods are employed for comparisons. They are linear SVM, *k*NN, LDA, SRC [8], random forest (RF), SIFT, LBP, LeNet [29], and CNN [30]. The linear SVM, *k*NN, LDA, SRC, and RF methods directly use the raw pixel values to train the classifier. The SIFT and LBP methods extract SIFT and LBP features from the face images and use linear SVM to perform expression classification. The LeNet and CNN methods are deep learning methods, which automatically learn features for classification. The CNN method has two convolutional layers with 32 3×3 filters and each layer connects with a 2×2 max-pooling layer. The final two layers are a dense layer of 128 hidden units and an output

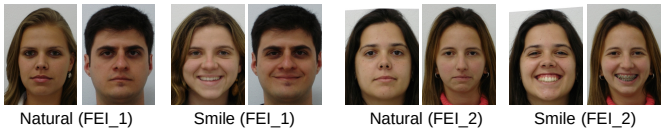

Fig. 6. Example images of **FEI_1** and **FEI_2**. These two data sets have facial images with natural and smile expressions.

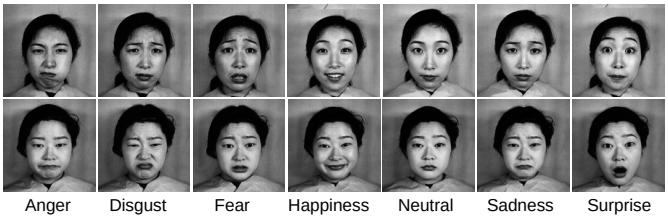

Fig. 7. Example images of **JAFFE**. This data set has seven different expressions: anger, disgust, fear, happiness, neutral, sadness, and surprise.

layer with the number of units equals the number of classes. For *k*NN, the number of neighbours is set to 1. In RF, the number of trees is 500 and the maximum tree depth is 100 [30].

B. Data sets

In the experiments, four different facial expression classification data sets are used. They are FEI_1 [31], FEI_2 [31], JAFFE [20], and GENKI [32]. Based on these four data sets, various numbers of training images are employed in the experiments, which aim to investigate whether the training set scale affects the performance of facialGP.

The FEI_1 and FEI_2 data sets have face images of different people into two facial expressions, i.e., natural and smile. Example images are shown in Fig. 6. These two data sets have 200 images, i.e., 100 images per class, respectively. The original images are colour images and are with a size of 260×360. We downsample the images using the ratio of 1/4 (the image size is changed to 65×90) and convert the colour images into greyscale to reduce computational cost. In the experiments, we use 50 images (25 images per class), 100 images (50 images per class), and 150 images (75 images per class) as the training sets, respectively. The remaining images except for training images are used as the test set.

The JAFFE data set is a well-known facial expression classification data set. It has seven common facial expressions, i.e., anger, disgust, fear, happiness, neutral, sadness, and surprise. The original 256×256 images are resized to 64×64. This data set has a total number of 217 images, i.e., about 30 images per class. Example images from each class of JAFFE are shown in Fig. 7. In the experiments, we use 35 images (5 images per classes), 70 images (10 images per class), and 140 images (20 images per class) as the training sets, respectively. The remaining images are used for testing.

Unlike the previous three data sets, the GENKI data set contains real-world face images with different expressions. The GENKI data set has two classes of facial expressions and it is a challenging task due to the high image variations.

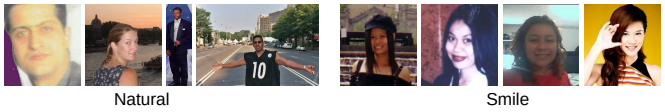

Fig. 8. Example images of **GENKI**. This data set has real-world facial images into two expressions: natural and smile.

As shown in Fig. 8, the faces have occlusion and locate at various positions with different orientations in the images. The GENKI data set has 4000 colour images with various sizes (most images have the size of 170×192 to 179×192). In the experiments, the images are converted to greyscale and resized to 85×95. The same as that on the other data sets, we use 400 images (200 images per class), 1000 images (500 images per class), and 2000 images (1000 images per class) for training and the remaining images for testing, respectively.

C. Parameter Settings

In the facialGP approach, the population size is 100 and the maximum number of generations is 50. The crossover rate is 0.8, the mutation rate is 0.19, and the elitism rate is 0.01. The parameter settings for facialGP refer to that in [33]. The *ramped-half-and-half* method is used for tree generation at the population initialisation step and in the mutation operation. The minimum tree depth is 2 and the maximum tree depth is 6. The selection method is tournament selection with size 7. The parameter settings for CNN and LeNet are the same as that in [30]. The implementation of facialGP is based on the *DEAP* (*Distributed Evolutionary Algorithm in Python*) [34] package and the implementations of the classification algorithms are based on the *scikit-learn* package [35]. We conduct experiments 30 independent runs of each method (including facialGP) on each data set.

V. Results and Discussions

This section discusses the experimental results obtained by the proposed facialGP approach and the nine baseline methods on four data sets using various numbers of training images. The test results, i.e., maximum accuracy, mean accuracy and standard deviation, are listed in Tables I - IV. To show the significant difference of performance improvement, Wilxocon rank-sum test is employed to compare the results from the 30 runs obtained by facialGP and a baseline method. The "+" and "–" symbols in Tables I - IV indicate that facialGP achieves significantly better and worse results than the compared method. The "=" symbol indicates that facialGP achieves similar results to the compared method. The final row of Tables I - IV presents the summary of the significance tests.

Test Accuracy on FEI_1: The test accuracy (maximum, mean and standard deviation) of the FEI_1 data set is listed in Table I. It can be found from the table that the proposed facialGP approach achieves significantly better results in almost all the comparisons. Using a small number of training images, i.e., 25 images per class, facialGP achieves significantly better results than any of the compared methods. Increasing the number of training images from 50 to 100, an increase of

TABLE I
COMPARISON OF TEST ACCURACY (%) ON FEI_1 USING VARIOUS NUMBERS OF TRAINING IMAGES

	50 images		100 images		150 images	
	Max	Mean±Std	Max	Mean±Std	Max	Mean±Std
SVM	79.33	78.71±0.17+	86.00	86.00±0.00+	88.00	87.53±0.86+
*k*NN	56.67	56.67±0.00+	52.00	52.00±0.00+	46.00	46.00±0.00+
LDA	80.67	80.67±0.00+	**97.00**	**97.00±0.00**–	96.00	**96.00±0.00**=
SRC	77.33	77.33±0.00+	80.00	80.00±0.00+	86.00	86.00±0.00+
RF	87.33	84.69±0.92+	90.00	89.20±0.41+	86.00	86.00±0.00+
SIFT	80.00	80.00±0.00+	86.00	86.00±0.00+	84.00	84.00±0.00+
LBP	63.33	59.49±2.41+	63.00	57.13±3.46+	76.00	64.40±6.73+
LeNet	90.00	85.58±3.41+	92.00	89.57±1.61+	94.00	91.13±2.56+
CNN	87.33	85.00±1.55+	93.00	89.37±1.40+	94.00	90.47±2.39+
facialGP	**94.67**	**91.62±1.87**	95.00	92.27±2.07	**100.0**	95.73±2.15
Overall	9+		8+, 1–		8+, 1=	

TABLE II
COMPARISON OF TEST ACCURACY (%) ON FEI_2 USING VARIOUS NUMBERS OF TRAINING IMAGES

	50 images		100 images		150 images	
	Max	Mean±Std	Max	Mean±Std	Max	Mean±Std
SVM	78.67	78.67±0.00+	87.00	86.97±0.18+	94.00	94.00±0.00+
*k*NN	48.00	48.00±0.00+	50.00	50.00±0.00+	52.00	52.00±0.00+
LDA	77.33	77.33±0.00+	92.00	**92.00±0.00**=	94.00	94.00±0.00+
SRC	76.67	76.67±0.00+	84.00	84.00±0.00+	92.00	92.00±0.00+
RF	89.33	86.16±1.13=	90.00	88.33±0.88+	94.00	92.27±1.26+
SIFT	56.67	56.67±0.00+	74.00	74.00±0.00+	82.00	82.00±0.00+
LBP	54.67	53.07±1.12+	59.00	55.60±1.73+	60.00	55.20±3.31+
LeNet	81.33	73.04±3.84+	92.00	87.57±1.87+	98.00	**96.47±1.46**–
CNN	80.00	76.71±1.92+	92.00	90.93±1.14=	98.00	95.40±2.30=
facialGP	**92.00**	**86.67±4.13**	**96.00**	90.87±2.91	**100.0**	95.00±2.21
Overall	8+, 1=		7+, 2=		7+, 1=, 1–	

TABLE III
COMPARISON OF TEST ACCURACY (%) ON JAFFE USING VARIOUS NUMBERS OF TRAINING IMAGES

	35 images		70 images		140 images	
	Max	Mean±Std	Max	Mean±Std	Max	Mean±Std
SVM	42.70	41.33±0.58+	59.44	57.71±0.68+	89.04	87.72±0.76+
*k*NN	15.17	15.17±0.00+	11.19	11.19±0.00+	34.25	34.25±0.00+
LDA	28.09	28.09±0.00+	51.05	51.05±0.00+	80.82	80.82±0.00+
SRC	36.52	36.52±0.00+	56.64	56.64±0.00+	86.30	86.30±0.00+
RF	37.08	33.78±1.71+	55.24	51.73±1.79+	82.19	76.53±2.21+
SIFT	24.72	24.72±0.00+	23.78	23.78±0.00+	34.25	34.25±0.00+
LBP	22.47	21.39±0.53+	25.87	22.24±1.39+	41.10	34.25±4.26+
LeNet	44.38	36.54±4.31+	64.34	58.42±3.83+	93.15	88.54±2.58+
CNN	41.01	35.22±2.04+	67.83	60.47±4.59+	94.52	90.36±2.29+
facialGP	**57.87**	**46.01±5.89**	**79.02**	**69.00±4.61+**	**95.89**	**92.15±2.33**
Overall	9+		9+		9+	

the test accuracy can be found in most of the algorithms, especially LDA. LDA achieves better results than facialGP using 100 training images. But the performance of LDA is significantly affected by the number of training images. LDA only achieves a maximum accuracy of 80.67% using 50 training images, which is 14% less than that achieved by facialGP. Compared with these baseline methods, the proposed facialGP approach achieves better and stable results when decreasing or increasing the number of training images on the FEI_1 data set.

Test Accuracy on FEI_2: The results of FIE_2 are listed in Table II. In the three cases, facialGP obtains 22 "+", 4 "=", and 1 "–" out of the total 27 comparisons. Similar to the pattern observed on FEI_1, facialGP achieves better results than any of the compared methods when using 50 training images. In the second case, i.e., using 100 training images, LDA obtains a mean accuracy of 92.00%, which is slightly higher than that by facialGP of 90.87%. But facialGP achieves a maximum accuracy of 100% using 150 training images, which is 4% higher than that by LDA (96%). From the results, it can be found that the performance of LeNet is increased when using more training images. For example, LeNet obtains a mean accuracy of 96.47% using 150 training images, which is 23.43% higher than that using 50 training images. This confirms that deep learning methods require a large number of training images to train the models. In contrast, the proposed facialGP approach can achieve better classification performance in the three cases, i.e., a small numbers of training images, a medium number of training images and a large number of training images.

Test Accuracy on JAFFE: Compared with FEI_1 and FEI_2, JAFFE is more challenging and it has seven classes of different expressions. The test results of JAFFE are listed in Table III. It is noticeable that facialGP achieves significantly better results than any of the baseline methods in the three cases of using various numbers of training images. The results show that facialGP is more effective for solving difficult facial expression classification tasks than the baseline methods. One possible reason is that facialGP learns effective appearance features from the automatically selected regions. The features learned by the proposed facialGP approach are more effective than the LBP features, the SIFT features, the raw pixel values, and the features learned by CNNs. The comparisons also indicate that facialGP is more effective for solving difficult facial expression tasks than the baseline methods.

Test Accuracy on GENKI: GENKI is a real-world data set with high image variations. The classification results of GENKI are listed in Table IV. It can be found that the proposed facialGP approach achieves significantly better classification accuracy than any of the baseline methods on GENKI in the three cases. In the first case, facialGP obtains a maximum accuracy of 69.67%, which is over 8% higher than the best accuracy of all the baseline methods. Increasing the number of training images, the best classification accuracy of facialGP increases to 72.90%, which is 6.1% higher than that of the baseline methods. This confirms that the proposed facialGP approach is more effective for difficult facial expression classification, especially using a small number of training images, than the baseline methods.

To sum up, the experimental results show that the proposed facialGP approach achieves significantly better results in almost all the comparisons in the three data split cases of the four facial expression classification data sets. This indicates that the proposed facialGP approach is an effective approach for facial expression classification by automatically performing region selection and feature extraction. The comparisons of the proposed facialGP approach and the baseline methods using various numbers of training images show that the proposed facialGP approach is more effective using a small number of training images. It is also noticeable that the performance of simple deep learning methods (LeNet and CNN) is affected

TABLE IV
COMPARISON OF TEST ACCURACY (%) ON GENKI USING VARIOUS NUMBERS OF TRAINING IMAGES

	400 images		1000 images		2000 images	
	Max	Mean±Std	Max	Mean±Std	Max	Mean±Std
SVM	56.61	56.52±0.06+	57.03	56.56±0.29+	57.30	56.73±0.27+
*k*NN	54.22	54.22±0.00+	53.37	53.37±0.00+	55.85	55.85±0.00+
LDA	54.86	54.86±0.00+	53.63	53.63±0.00+	55.30	55.30±0.00+
SRC	54.33	54.33±0.00+	55.80	55.80±0.00+	55.25	55.25±0.00+
RF	60.47	59.99±0.28+	60.80	59.97±0.42+	63.25	61.96±0.58+
SIFT	60.17	60.17±0.00+	61.60	61.60±0.00+	62.90	62.90±0.00+
LBP	57.58	55.17±3.25+	59.27	52.91±4.58+	59.30	54.16±3.74+
LeNet	61.64	59.15±1.00+	66.77	63.05±2.05+	74.30	71.09±1.86+
CNN	61.64	59.13±1.06+	66.80	63.70±1.72+	72.75	69.67±1.68+
facialGP	**69.67**	**66.65±1.66+**	**72.90**	**70.59±1.76+**	**77.40**	**74.92±1.78**
Overall		9+		9+		9+

by the number of training images. Because LeNet and CNN both have many trainable parameters and the current training set is not sufficient to obtain optimal parameters. Therefore, the performance of LeNet and CNN is worse than facialGP. In addition, the proposed facialGP approach is less affected by decreasing the number of training images. It is noticeable that facialGP achieves significantly better performance than any of the baseline methods on the four data sets using a small number of training images.

VI. FURTHER ANALYSIS

This section further analyses the solutions evolved by the proposed facialGP approach to provide insight into it. An example program/solution of facialGP is shown in Fig. 9. This example program achieves 94.52% accuracy on JAFFE using 140 training images (20 images per class). This example program selects two regions of the face as circled in the example images in Fig. 9. The first (left) region is a 14×29 region, which contains the eye and nose areas of the face. The raw pixel values of this region are extracted by the $Concatenate$ function to form a 406D feature vector. The second (right) region is a 29×26 region, which contains the mouth area of the face. Note that $41 + 29$ is larger than 64 so that the size of the second region is 23×26. From this region, the $SIFT$ function is employed to extract 128 gradient and orientation features. In total, this example program produces 534 features from an input 64×64 image.

From Fig. 9, it can be found that the two selected regions contain salient information of the face. The first region contains the eye and nose areas of the face, which have different appearances in the seven expressions. Therefore, the raw pixel values of this region are extracted. We test these extracted 406 features for classification and obtain an accuracy of 91.78% on the same test set of JAFFE. The second region contains the mouth area of the face, which contains informative shape features so that the $SIFT$ function is employed for feature extraction. The extracted 128 features by $SIFT$ obtain an accuracy of 80.82% on the same test set of JAFFE. In contrast, the combination of the 406 features and the 128 features achieves an accuracy of 94.52%, which is much higher than that achieved by using 406 or 128 features individually. It is suggested that using a combination of these features achieves better expression classification results than using the features

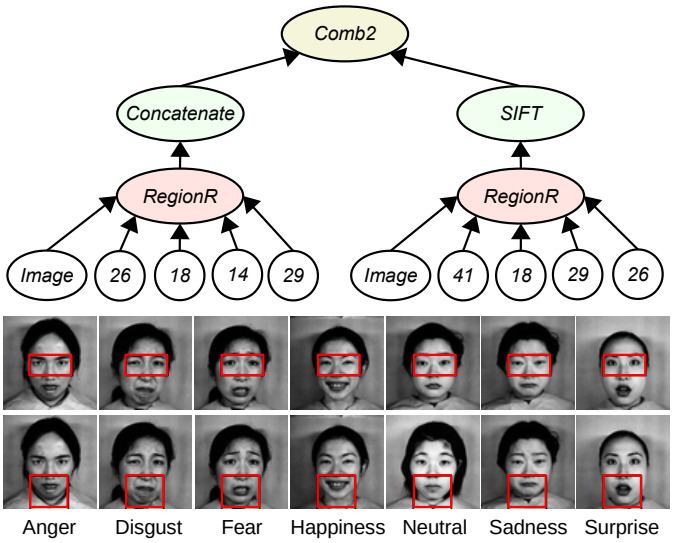

Fig. 9. An example program evolved by facialGP on JAFFE (140 training images) and the images with selected regions by the example program.

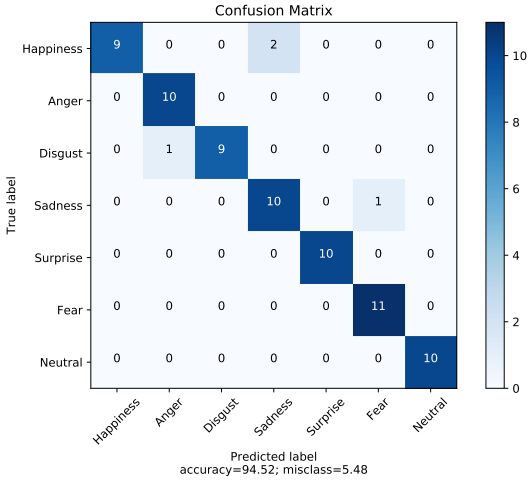

Fig. 10. Confusion matrix obtained by the example program in Fig. 9.

extracted from a single region. This indicates that the proposed facialGP approach not only selects informative regions of the face but also finds a good combination of various features to achieve a high classification accuracy.

Fig. 10 shows the confusion matrix obtained by the example program in Fig. 9. It can be found that only four images are classified wrongly. Two images in the $Happiness$ class are classified into the $Sadness$ class, one image in the $Disgust$ class is classified into the $Anger$ class, and one image in the $Sadness$ class is classified into the $Fear$ class. It can also be found that all the images in the $Anger$, $Surprise$, $Fear$, and $Neutral$ classes are correctly classified. This indicates that some images/classes of the JAFFE data sets are difficult to be classified. In the future, more attention should be paid to the classes that have wrongly classified images.

VII. CONCLUSIONS

The goal of this paper was to develop a new GP-based feature learning approach for facial expression classification.

This goal has been successfully achieved by developing the facialGP approach to feature learning for facial expression classification. The proposed facialGP approach can automatically select regions of the face image and extract features from the selected regions. The performance of the proposed facialGP approach has been examined on four facial expression data sets of varying difficulty. The experimental results show that the proposed facialGP approach achieves significantly better classification performance than nine baseline methods in almost all the comparisons. In addition, the performance of the proposed facialGP approach is examined under the scenarios of using different numbers of training images. The experimental results show that facialGP achieves significantly better results than any of the baseline methods using a small number of training images. Further analysis of the example program evolved by facialGP shows that it not only selects informative regions of the face but also finds a good combination of different features to achieve a high classification accuracy.

The proposed facialGP approach provides an example of a GP-based feature learning approach for facial expression classification. In the future, the performance of facialGP will be further improved on a small number of training instances. In addition, new region selection operators will be developed to select regions that are invariant to the position.

REFERENCES

[1] P. Kumar, S. Happy, and A. Routray, "A real-time robust facial expression recognition system using hog features," in *Proceedings of International Conference on Computing, Analytics and Security Trends (CAST)*. IEEE, 2016, pp. 289–293.

[2] P. Ekman and W. V. Friesen, *Unmasking the face: A guide to recognizing emotions from facial clues*. Ishk, 2003.

[3] G. Ali, M. A. Iqbal, and T.-S. Choi, "Boosted nne collections for multicultural facial expression recognition," *Pattern Recognition*, vol. 55, pp. 14–27, 2016.

[4] W.-L. Chao, J.-J. Ding, and J.-Z. Liu, "Facial expression recognition based on improved local binary pattern and class-regularized locality preserving projection," *Signal Processing*, vol. 117, pp. 1–10, 2015.

[5] X. Zhao, X. Shi, and S. Zhang, "Facial expression recognition via deep learning," *IETE Technical Review*, vol. 32, no. 5, pp. 347–355, 2015.

[6] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[7] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.

[8] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2008.

[9] Y. Bi, B. Xue, and M. Zhang, "An evolutionary deep learning approach using genetic programming with convolution operators for image classification," in *Proceedings of IEEE Congress on Evolutionary Computation (CEC)*. IEEE, 2019, pp. 3197–3204.

[10] H. Al-Sahaf, Y. Bi, Q. Chen, A. Lensen, Y. Mei, Y. Sun, B. Tran, B. Xue, and M. Zhang, "A survey on evolutionary machine learning," *Journal of the Royal Society of New Zealand*, vol. 49, no. 2, pp. 205–228, 2019.

[11] L. Shao, L. Liu, and X. Li, "Feature learning for image classification via multiobjective genetic programming," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 7, pp. 1359–1371, Jul 2014.

[12] J. R. Koza, *Genetic programming: on the programming of computers by means of natural selection*. MIT press, Cambridge, 1992.

[13] Y. Bi, B. Xue, and M. Zhang, "A survey on genetic programming to image analysis," *Journal of Zhengzhou University (Engineering Science)*, vol. 39, no. 06, pp. 3–13, 2018.

[14] H. Al-Sahaf, A. Al-Sahaf, B. Xue, M. Johnston, and M. Zhang, "Automatically evolving rotation-invariant texture image descriptors by genetic programming," *IEEE Transactions on Evolutionary Computation*, vol. 21, no. 1, pp. 83–101, Feb 2017.

[15] Y. Bi, B. Xue, and M. Zhang, "An automatic feature extraction approach to image classification using genetic programming," in *Proceedings of International Conference on the Applications of Evolutionary Computation*, 2018, pp. 421–438.

[16] J. H. Shah, M. Sharif, M. Yasmin, and S. L. Fernandes, "Facial expressions classification and false label reduction using lda and threefold svm," *Pattern Recognition Letters*, 2017.

[17] M. J. Cossetin, J. C. Nievola, and A. L. Koerich, "Facial expression recognition using a pairwise feature selection and classification approach," in *Proceedings of International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2016, pp. 5149–5155.

[18] Z. Huang and F. Ren, "Facial expression recognition based on active appearance model & scale-invariant feature transform," in *Proceedings of IEEE/SICE International Symposium on System Integration*. IEEE, 2013, pp. 94–99.

[19] M. Hussain, S. A. Khan, N. Ullah, N. Riaz, and M. Nazir, "Computationally efficient invariant facial expression recognition," *Research Journal of Recent Sciences*, vol. 2277, p. 2502, 2014.

[20] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with gabor wavelets," in *Proceedings of the Third IEEE international conference on automatic face and gesture recognition*, Apr 1998, pp. 200–205.

[21] A. T. Lopes, E. de Aguiar, A. F. De Souza, and T. Oliveira-Santos, "Facial expression recognition with convolutional neural networks: coping with few data and the training sample order," *Pattern Recognition*, vol. 61, pp. 610–628, 2017.

[22] D. Atkins, K. Neshatian, and M. Zhang, "A domain independent genetic programming approach to automatic feature extraction for image classification," in *Proceedings of IEEE Congress on Evolutionary Computation (CEC)*, 2011, pp. 238–245.

[23] D. J. Montana, "Strongly typed genetic programming," *Evolutionary Computation*, vol. 3, no. 2, pp. 199–230, 1995.

[24] H. Al-Sahaf, A. Song, K. Neshatian, and M. Zhang, "Two-tier genetic programming: Towards raw pixel-based image classification," *Expert System with Application*, vol. 39, no. 16, pp. 12 291–12 301, Nov 2012.

[25] A. Lensen, H. Al-Sahaf, M. Zhang, and B. Xue, "Genetic programming for region detection, feature extraction, feature construction and classification in image data," in *Proceedings of European Conference on Genetic Programming*. Springer, 2016, pp. 51–67.

[26] A. Vedaldi and B. Fulkerson, "Vlfeat: An open and portable library of computer vision algorithms," in *Proceedings of the 18th ACM international conference on Multimedia*, 2010, pp. 1469–1472.

[27] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, Jul 2002.

[28] Y. Bi, B. Xue, and M. Zhang, "Genetic programming with a new representation to automatically learn features and evolve ensembles for image classification," *IEEE Transactions on Cybernetics*, 2020. DOI: 10.1109/TCYB.2020.2964566.

[29] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov 1998.

[30] Z.-H. Zhou and J. Feng, "Deep forest," *National Science Review*, vol. 6, no. 1, pp. 74–86, Oct 2018.

[31] C. E. Thomaz, "Fei face database," *online: http://fei.edu.br/˜cet/facedatabase.html*, Mar 2012.

[32] "The mplab genki database, genki-4k subset." *online: http://mplab.ucsd.edu*.

[33] Y. Bi, B. Xue, and M. Zhang, "An automated ensemble learning framework using genetic programming for image classification," in *Proceedings of the Genetic and Evolutionary Computation Conference*. ACM, 2019, pp. 365–373.

[34] F.-A. Fortin, F.-M. De Rainville, M.-A. Gardner, M. Parizeau, and C. Gagné, "DEAP: Evolutionary algorithms made easy," *Journal of Machine Learning Research*, vol. 13, no. Jul, pp. 2171–2175, Jul 2012.

[35] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, Oct 2011.