A Region Adaptive Image Classification Approach Using Genetic Programming

Qinglan Fan, Bing Xue and Mengjie Zhang
Evolutionary Computation Research Group,
Victoria University of Wellington, PO Box 600
Wellington 6140, New Zealand
Email: {qinglan.fan, bing.xue, mengjie.zhang}@ecs.vuw.ac.nz

***Abstract*—Feature extraction, as one essential step of image classification, can potentially reduce image data dimensionality and capture effective information for improving performance. However, most existing image descriptors are designed to conduct specific tasks and might not be sufficient for different types of images. Genetic programming (GP) can automatically extract multiple important and discriminative features by incorporating diverse image descriptors into a GP program. Furthermore, different regions in an image have different structural characteristics. In this paper, we propose a region adaptive image classification approach based on GP, which can automatically extract informative image features by automatically applying different image descriptors in different regions of an image. A new flexible GP program structure with a new function set and a new terminal set is developed in this approach. The performance of the proposed method is evaluated on four various data sets and compared with other state-of-the-art classification methods. Experimental results illustrate that the proposed approach is capable of achieving better or competitive performance than these baseline methods. Further analysis of some good programs shows the high interpretability of the proposed method.**

***Index Terms*—feature extraction, image classification, genetic programming, region adaptive**

I. Introduction

Image classification is an important image analysis technique that can classify an image according to its visual content, with a wide range of real-world applications such as face recognition [1], pedestrian detection [2], and medical diagnosis [3]. The main challenges to improve classification accuracy are the high dimensionality and diversity of the image data.

Feature extraction is a crucial component of the image classification process [4], [5], which could effectively reduce image dimensionality and capture distinctive image features. A number of image descriptors have been constructed to extract useful image features such as histogram (Hist) [6], Local Binary Pattern (LBP) [7], Scale Invariant Feature Transform (SIFT) [8], Speeded Up Robust Features (SURF) [9], and Histogram of Orientated Gradients (HOG) [10]. However, these descriptors are designed to conduct specific tasks, which might not achieve satisfying performance in different types of image classification tasks. In addition, different regions in a single image tend to have different structural characteristics. It is difficult to achieve promising results on different types of images classification tasks by using a single image descriptor.

Based on the extracted image features, selecting a suitable classifier is also a key step for image classification. Generally, domain knowledge and human intervention are required to determine how to generate the best combination of the image descriptors and classifiers when dealing with different tasks. But domain experts are not always available and often costly to emply [11].

Genetic programming (GP) [12] is an evolutionary computation (EC) technique, which has been widely applied to image feature extraction [5], [13], [14] and classification problems [4], [15]–[23] without domain knowledge and human intervention. GP is a population-based search approach that mimics Darwin's principles. It begins by randomly generating a set of solutions, and further updates these solutions by using selection and genetic operators, i.e. crossover and mutation until finding the best solution. The representation of each solution is typically a tree structure, where diverse types of data can be used as nodes of the tree. Therefore, a variety of image descriptors can be fed into a single tree to extract useful features, and then the image can be categorized into different groups according to the output of the tree. That is to say, GP can achieve automatic feature extraction and image classification simultaneously.

A number of research studies show that GP-based image classification methods achieve good performance. The majority of existing image classification approaches based on GP try to capture discriminative features among images from different classes. For example, Lensen et.al. [17] presented a GP-based image classification method by directly incorporating HOG as function nodes of a GP tree. The images can be classified into different classes based on HOG features. However, this may not be sufficient for complex image classification problems due to limited image features. In general, a single image contains smooth, texture, and edge regions that have different structural characteristics [24]. By using differences between regions in a single image, extracting more rich distinctive image features for different types of classification tasks is the main motivation of this paper.

A. Goals

The goal of this paper is to develop a new GP-based method for different types of image classification, which can

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

automatically extract adaptive and informative image features by utilizing different image descriptors in different regions of an image. To achieve this, we propose a region adaptive image classification approach based on GP, which can capture two types of image features (texture and edge) and perform classification using these features simultaneously. The fractal dimension (FD) can quantitatively measure the roughness of an image and extract the texture features [25]. Another important texture descriptor, LBP, has been employed in texture classification [26]. For edge regions, Sobel and Prewitt are efficient methods to detect the edge structure. By designing a new GP program representation, the best combinations of texture and edge descriptors can be selected automatically due to its flexible tree-structure. The performance of the proposed method will be examined on four different types of image data sets of varying difficulty and compared with five GP-based methods and 29 non-GP approaches. Specifically, the goal can be diveded into four following objectives:

- Develop a new GP program to incorporate different functions for image classification tasks.
- Design a new function set and a new terminal set for extracting high-level distinctive image features including texture and edge.
- Evaluate the effectiveness of the presented method and compare it with existing state-of-the-art image classification methods.
- Analyze the program trees of good individuals to understand how they capture useful features for image classification.

II. Related work

This section reviews typical related work which utilizes GP for image feature extraction and classification. The limitations of these methods are discussed, showing the motivations of our proposed region adaptive image classification approach.

Atkins et al. [15] proposed a domain-independent image classification system that has three tiers (3T-GP) with different functions and performs image filtering, feature extraction, and classification, respectively. This method conducts feature extraction and classification simultaneously with the framework of GP, which achieves comparative performance to the domain-specific features. Motivated by the promising results of the 3T-GP approach, AI-Sahaf et al. [16] developed a two-tier GP (2T-GP) method by removing the filtering tier and detecting regions of different shapes and sizes such as rectangle, circle, line, and column. It is much faster than 3T-GP and achieves better performance by extracting higher-level features. However, both 3T-GP and 2T-GP extract features based on pixel statistics, they may not be effective for complex image classification.

To extract a set of important features from images, Lensen et al. developed HOG-GP [17] and SURF-GP [19] methods for image classification where HOG or SURF was fed into a designed GP program. Experimental results show that they work better than 2T-GP. Almeida et al. [22] developed a GP program to discover an effective combination of time-series similarity functions for the remote sensing image classification. Experimental results show that it achieves competitive classification performance when comparing with traditional methods utilizing a single similarity function. Shao et al. [4] presented a multi-objective GP approach for image classification in which the accuracy rate and the tree complexity are considered as the fitness objectives. The high-level image features can be extracted, which exhibit better performance than hand-crafted feature representation. By introducing a new mutation method and a new crossover technique, Price et al. [23] improved the GP evolutionary process to generate advanced image features for classification. However, only using one type of image descriptor or extracting simple features through the mathematics statistical method is not sufficient for different image classification tasks.

Bi et al. [21] proposed a multi-layer GP (MLGP) method for image classification in which each layer targets a subtask, e.g., region detection and feature extraction. The high-level feature is extracted by feeding three types of operators including histogram equalization operator, image filters, and image descriptors to the GP program. Experimental results on different datasets such as object detection show that the method achieves better results than baseline methods. Moreover, considering that different regions in a single image have different structural characteristics, extracting more rich image features for different types of image classification needs further investigation.

III. The Proposed Approach

In this paper, we develop a new GP program for different types of image classification. A new function set and a new terminal set are proposed to enable GP to automatically select the best combination of texture and edge descriptors, and generate distinctive features.

A. GP program structure design

For different types of image classification tasks, i.e., facial expression, scene detection, and objective detection, it is crucial to extract adaptive and discriminative features according to the content of images. As we mentioned before, texture and edge are the two main types of features in an image. The proposed GP program could automatically extract these features from an image based on its content and perform classification simultaneously.

As shown in Fig. 1, an example program tree takes an MIT street image [27] as input, captures the texture and edge features in different regions using **FD** and **SobelY** respectively. The value returned from the root is then used as the parameter decision. In our study, the image will be labeled as one class if the output is positive, or as the other class if the return value is zero or negative. The designed GP program could detect important regions from an image, extract their features and perform classification simultaneously. To be able to combine various texture and edge descriptors into a single tree, strongly-typed GP [28] is used to restrict the input and

output types of nodes. The details will be discussed in the next section.

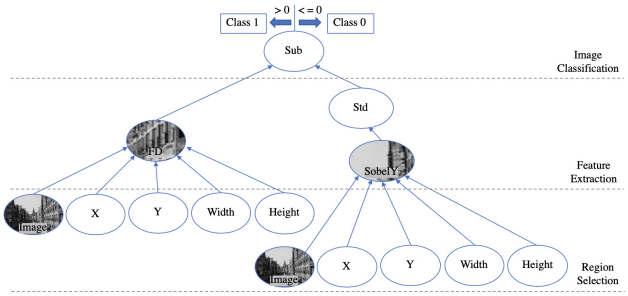


Fig. 1. An example of the program structure.

B. Function Set

According to the designed GP program representation, two types of operators are used to construct the function set. Table I describes the different types of function operators and outlines the inputs and outputs of each type.

TABLE I
THE FUNCTION SET

Operator	Input	Output	Description
LBP	A region	A region	Uniform LBP descriptor
FD	A region	Double	The fractal dimension of a region
SobelX	A region	A region	Sobel filter along X axis
SobelY	A region	A region	Sobel filter along Y axis
PrewittX	A region	A region	Prewitt filter along X axis
PrewittY	A region	A region	Prewitt filter along Y axis
Std	A region	Double	The standard deviation of a region
Add	(Double, Double)	Double	Arithmetic oprators
Sub	(Double, Double)	Double	Arithmetic oprators
Mul	(Double, Double)	Double	Arithmetic oprators
Div	(Double, Double)	Double	Arithmetic oprators

The first type of operators contains **LBP**, **FD**, **SobelX**, **SobelY**, **PrewittX**, **PrewittY**, and **Std** which are employed for feature extraction. **FD** and **LBP** are utilized to capture texture features. The majority of natural images can be represented as a fractal set [29]. **FD** is one of the most important fractal features that could precisely describe the local variations of an image. It has been successfully applied to feature extraction of texture images, which effectively improves the classification accuracy [30]. **LBP** is another effective texture descriptor that has gray and rotation invariance. It achieves impressive performance on face recognition. **SobelX**, **SobelY**, **PrewittX**, and **PrewittY** are effective edge detectors, which could capture the edge structure information from images along different directions. **Std** means the standard deviation and shows the variation within a region. In this study, we use **Std** to quantify the variation of texture and edge features.

The second type of operators is **Add** $(+)$, **Sub** $(-)$, **Mul** $(\times)$, and **Div** $(/)$, which represent arithmetic functions and allow GP to use multiple extracted features for classification. The **Div** is protected by returning 0 if the denominator is 0.

C. Terminal Set

As shown in Table II, there are three types of terminals in the terminal set that are children nodes of texture or edge descriptors, namely, **Image**, **X**, **Y**, **Width**, and **Height**. The terminal **Image** represents the input image, which is a 2D matrix corresponding to the pixel values in the range $[0, 1]$ (The raw image is normalized by dividing 255 to overcome the problems brought by uneven illumination). The **X** and **Y** terminals are responsible to generate random integers as coordinates for the top-left point of a selected region. They are between $[0, ImageWidth - 3]$ or $[0, ImageHeight - 3]$. **Width** and **Height** mean the width and height of a selected region, with the range of $[3, ImageWidth]$ and $[3, ImageHeight]$ respectively. Therefore, the minimum size of the selected region is 3×3 to capture informative regions from the image [16].

TABLE II
THE TERMINAL SET

Terminal	Value Range	Description
Image	$[0, 1]$	Input image
X	$[0, ImageWidth - 3]$	The Coordinate of the top-left point of a selected region
Y	$[0, ImageHeight - 3]$	The Coordinate of the top-left point of a selected region
Width	$[3, ImageWidth]$	The width of the selected region
Height	$[3, ImageHeight]$	The height of the selected region

D. Fitness Measure

A fitness function (f) is needed to measure a program's goodness in GP. In this paper, f is the classification accuracy, which is commonly used for binary image classification. It is defined as follows,

$$f = \frac{TP + TN}{TP + FP + TN + FN},$$

where TP and FP are the numbers of instances being correctly and incorrectly classified as positive instances, TN and FN are the numbers of instances being correctly and incorrectly classified as negative instances.

IV. Experiment Design

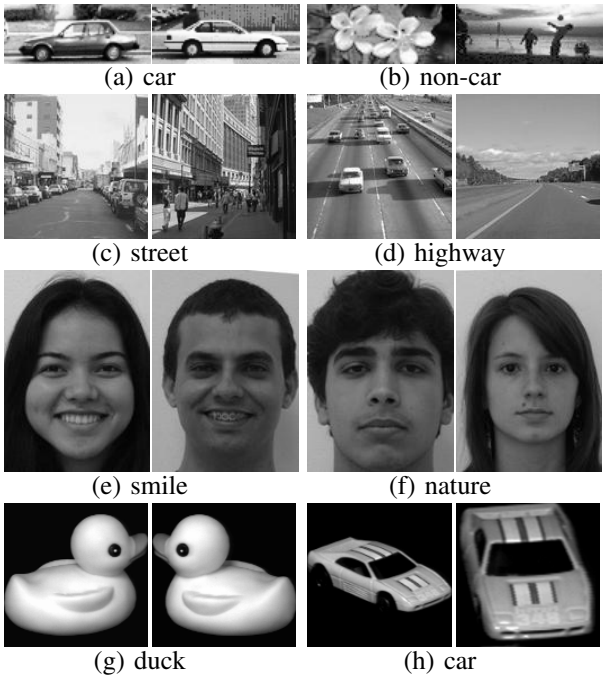
This section summarizes the details of benchmark image data sets, benchmark methods, and parameter settings of the proposed method and compared methods.

A. Data sets

Four different types of image data sets for binary image classification are employed to assess the performance of the proposed approach. They are UIUC [31], SCENE [27], FACES [32], and Columbia Object Image Library (COIL-20) [33]. For GP-based methods, each data set is split into the training set, validation set, and testing set, accounting for 50%, 25%, and 25% images respectively. For non-GP methods, the testing set keeps the same, whilst the training instances come from

the training set and the validation set used in the GP-based approaches.

UIUC is a car detection data set that has 1050 instances containing 550 car images (Fig. 2 (a)) and 500 non-car images (Fig. 2 (b)). Each instance is 100×40 pixels. The SCENE data set is used for scene classification, which consists of 292 streets and 260 highway images. Fig. 2 (c) and Fig. 2 (d) show examples of two streets and two highway images. The size of all images in this data set is 128×128 pixels. The FACES data set is from the FEI face database for facial expression classification. The original size of each image is 640×480 pixels, and are resized to 130×180 to reduce time complexity. As shown in Fig. 2 (e) and Fig. 2 (f), 100 smile images and 100 nature images are used to test the performance of the proposed approach. COIL-20 is an object classification data set that includes 1,440 grayscale images of 20 objects. In the experiment, we randomly select two types of objects for binary image classification in which there are 72 duck images (Fig. 2 (g)) and 72 car images (Fig. 2 (h)). The size of each image is 128×128 pixels.



(a) car (b) non-car (c) street (d) highway (e) smile (f) nature (g) duck (h) car

Fig. 2. Samples from UIUC, SCENE, FACES, COIL-20

B. Methods for comparisons

To demonstrate the effectiveness of the presented approach, a number of baseline methods including five GP-based and 29 non-GP are employed for comparison. The five GP-based methods are 3T-GP [15], 2T-GP [16], MLGP [21], LBP+GP, and Hist+GP. The method LBP+GP extracts image features utilizing the LBP descriptor and feeds them into the GP program for classification. The same process can be applied to Hist+GP. Furthermore, we select four popular image feature extraction methods and seven typical classification algorithms for non-GP methods. For each non-GP method, informative image features can be captured by one of the feature extraction methods and then fed into a classifier to obtain a class label. The four image feature extraction techniques include FedEx [34], Hist [6], LBP [7], and HOG [10]. 1-nearest neighbor(1NN), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), Multilayer Perceptron (MLP), Gaussian Naive Bayes (NB), and Adaptive Boosting (AdaBoost) are used to perform classification based on these extracted features. We implement these non-GP approaches based on the well-known *scikit-learn* Python package [35] with the default settings. Additionally, the performance of an image classification method based on convolutional neural networks (CNNs) [36] is reported for comparisons. It has five layers (simplified as CNN-5) that include three convolutional layers and two max-pooling layers. The parameter settings are the same as that mentioned in [36].

C. Parameter settings

To ensure the fairness of comparisons, parameters of all GP-based methods for the evolutionary process are the same as listed in Table III [15], [16], [21]. Each approach has run independently 30 times on each image data set with different random seeds.

TABLE III
PARAMETER SETTINGS FOR THE GP RUNNING

Parameter	Value
Generations	50
Population size	1024
Mutation rate	0.19
Crossover rate	0.8
Elitisim rate	0.01
Tree-depth	2-10
Selection type	Tournament
Tournament size	7

V. RESULTS AND DISCUSSIONS

We employ classification accuracy and F-measure indices for quantitative comparisons in this paper. The most common form of F-measure is F1- measure ($f1$), which is calculated as follows,

$$f1 = \frac{2TP}{2TP + FP + FN}.$$

Tables IV and VI list the classification accuracy and Tables V and VII list the F1-measure results of all approaches on four different types of data sets. Student's t-test with a 95% confidence interval is used to evaluate the performance of the presented approach compared to that of other methods. The symbols "+" and "-" indicate that the proposed approach is significantly better and worse than the corresponding method. "=" means that two methods obtain similar results. Furthermore, two evolved GP programs on the UIUC and SCENE datasets are analyzed to understand how the proposed method effectively extracts image features for classification.

A. Compared to the GP-based methods

Table IV and Table V respectively detail the classification accuracy and F1-measure achieved by the proposed method and the other five GP-based approaches sequentially on the four image data sets in terms of mean, standard deviation, and maximum.

UIUC: The proposed method reaches 91.67% mean accuracy and 96.20% maximum accuracy, which significantly outperforms other GP-based methods. The similar pattern can be seen in Table V, namely, 0.930 mean F1-measure and 0.967 maximum F1-measure. Fig. 2 (a) and Fig. 2 (b), the two classes in this data set tend to have similar background information. Therefore, it is difficult to obtain accurate classification results by utilizing simple image features (3T-GP, 2T-GP) or a single image descriptor (LBP+GP, Hist+GP). Moreover, considering that different regions in an image contain various characteristics, the proposed region adaptive image classification approach achieves better performance than MLGP.

SCENE: On this data set, the presented algorithm achieves 0.94%-9.85% and 0.012-0.106 increase in mean test accuracy and F1-measure respectively over other GP-based methods. This result occurs because the discriminative image features can be effectively captured by automatically selecting suitable image descriptors based on image content in the developed GP program, in which the right class label can be output.

FACES: The results on FACES show that the presented approach gains significantly better or similar classification accuracy than compared methods. In addition, the standard deviations of the classification accuracy and F1-measure in our method are the lowest, which means that the proposed algorithm is more stable than other GP-based methods. It would have more potential applications in the real world.

COIL-20: The majority of methods obtain 100% accuracy and F1-measure in maximum on this data set. Compared with these GP-based approaches, the proposed algorithm achieves competitive results in mean and standard deviation. But overall it is significantly worse than 3T-GP and LBP+GP, which may be because most regions in images on this data set are smooth, simple features can be employed to classify these two classes.

To summarize, the proposed region adaptive image classification approach has better classification accuracy than other algorithms on four different types of image data sets. Extracting more rich image features from an image with the flexible framework of GP contributes to the improvement of the classification accuracy.

TABLE IV
CLASSIFICATION ACCURACY(%) ON THE FOUR DATASETS FOR COMPARING WITH OTHER GP-BASED METHODS

method	UIUC		SCENE	
	mean±Std	max	mean±Std	max
3T-GP	88.71±3.45 +	94.29	85.40±1.84 +	90.58
2T-GP	88.07±3.98 +	96.20	84.44±2.90 +	89.86
MLGP	90.35±2.87 +	95.44	91.44±1.79 +	93.37
LBP+GP	81.39±1.99 +	85.93	93.09±1.43 =	94.93
Hist+GP	62.25±2.09 +	68.43	84.18±1.69 +	87.45
Ours	91.67±2.17	96.20	94.03±1.26	96.38
method	FACES		COIL-20	
	mean±Std	max	mean±Std	max
3T-GP	91.0±3.45 =	96.00	97.22±2.38 -	100.0
2T-GP	89.46±3.61 =	96.00	94.77±4.36 =	100.0
MLGP	86.46±6.80 =	100.0	92.33±4.91 =	100.0
LBP+GP	53.4±6.06 +	66.00	99.68±1.33 -	100.0
Hist+GP	60.41±7.18 +	74.00	91.44±4.18 +	97.00
Ours	90.17±3.17	96.00	93.97±3.46	100.0

TABLE V
F1-MEASURE VALUES ON THE FOUR DATASETS FOR COMPARING WITH OTHER GP-BASED METHODS

method	UIUC		SCENE	
	mean±Std	max	mean±Std	max
3T-GP	0.902±0.033 +	0.956	0.841±0.031 +	0.894
2T-GP	0.882±0.034 +	0.955	0.827±0.025 +	0.883
MLGP	0.918±0.022 =	0.958	0.922±0.024 +	0.961
LBP+GP	0.836±0.016 +	0.873	0.922±0.017 +	0.944
Hist+GP	0.643±0.018 +	0.695	0.852±0.016 +	0.892
Ours	0.930±0.021	0.967	0.934±0.016	0.960
method	FACES		COIL-20	
	mean±Std	max	mean±Std	max
3T-GP	0.891±0.049 =	0.960	0.974±0.021 =	1.00
2T-GP	0.887±0.051 =	0.960	0.948±0.038 =	1.00
MLGP	0.844±0.058 +	0.938	0.927±0.043 =	1.00
LBP+GP	0.522±0.078 +	0.653	0.996±0.013 -	1.00
Hist+GP	0.588±0.071 +	0.720	0.923±0.032 +	0.980
Ours	0.884±0.035	0.945	0.956±0.028	1.00

B. Compared to the non-GP methods

The results obtained by 29 non-GP methods on the four data sets are shown in Table VI and Table VII. Overall, the proposed approach achieves competitive performance in classification accuracy and F1-measure. On the UIUC data set, the classification accuracy of the proposed algorithm is significantly higher than most non-GP methods. The majority of the compared approaches have an accuracy under 90%, while our method reaches 91.67% mean and 96.2% maximum accuracy. In addition, the presented method considerably outperforms all classifiers with Hist features. A similar pattern can be shown on SCENE. One important point is that the F1-measure of our method is significantly higher than all non-GP approaches on the UIUC data set. For the FACES data set, almost all compared methods obtain significantly worse results including both classification accuracy and F1-measure. In total, our method gains 44 "+" and 8 "-". The performance of the evolved region adaptive GP program on COIL-20 is satisfying but is slightly worse than that of some non-GP methods.

The performance of these non-GP methods except CNN-5 largely depends on feature extraction techniques, classifiers, and their combinations. Among the four feature extraction methods, they perform differently on different types of data sets. For instance, Hist only obtains promising results on the COIL-20 data set while the classification accuracy of the other three data sets is not satisfying. That is because only using one type of features is not sufficient for various types of image data sets. Moreover, selecting the most suitable combination

TABLE VI
CLASSIFICATION ACCURACY(%) ON THE FOUR DATASETS FOR COMPARING WITH OTHER NON-GP METHODS

Feature \ Classifier	1NN	SVM	DT	RF	MLP	NB	Adaboost
	UIUC (ours: 91.67±2.17, 96.20)						
FeEX	83.65 +	52.47 +	83.26 +	89.73 +	92.77 -	88.21 +	90.11 +
Hist	49.81 +	52.47 +	58.55 +	70.34 +	59.69 +	67.30 +	67.30 +
LBP	84.41 +	52.47 +	85.55 +	88.21 +	89.35 +	87.45 +	92.78 -
HOG	90.87 +	73.76 +	80.60 +	89.35 +	92.01 =	89.73 +	92.01 =
CNN-5	mean±Std: 85.27±3.84 + ; max: 89.73						
	SCENE (ours: 94.03±1.26, 96.38)						
FeEX	88.41 +	52.89 +	84.78 +	91.30 +	84.78 +	88.41 +	89.13 +
Hist	77.53 +	52.89 +	81.88 +	84.05 +	76.81 +	82.60 +	85.50 +
LBP	89.85 +	52.89 +	92.75 +	95.65 -	94.92 =	94.20 =	94.20 =
HOG	94.20 =	73.18 +	82.60 +	92.02 +	96.37 -	96.37 -	93.47 +
CNN-5	mean±Std: 88.52±3.58 + ; max: 91.30						
	FACES (ours: 90.17±3.17, 96.00)						
FeEX	26.00 +	50.00 +	52.00 +	36.00 +	62.00 +	46.00 +	68.00 +
Hist	16.00 +	36.00 +	36.00 +	22.00 +	42.00 +	46.00 +	36.00 +
LBP	18.00 +	54.00 +	56.00 +	50.00 +	52.00 +	48.00 +	64.00 +
HOG	34.00 +	88.00 +	82.00 +	96.00 -	90.00 =	78.00 +	90.00 =
CNN-5	mean±Std: 69.00±7.47 + ; max: 80.00						
	COIL-20 (ours: 93.97±3.46, 100.0)						
FeEX	100.0 -	77.77 +	100.0 -	100.0 -	97.22 -	100.0 -	100.0 -
Hist	100.0 -	50.00 +	100.0 -	100.0 -	100.0 -	100.0 -	100.0 -
LBP	97.61 -	52.77 +	97.61 -	100.0 -	97.61 -	100.0 -	97.61 -
HOG	100.0 -	100.0 -	97.61 -	97.61 -	100.0 -	92.85 +	97.61 -
CNN-5	mean±Std: 86.42±1.64 + ; max: 90.48						

TABLE VII
F1-MEASURE VALUES ON THE FOUR DATASETS FOR COMPARING WITH OTHER NON-GP METHODS

Feature \ Classifier	1NN	SVM	DT	RF	MLP	NB	Adaboost
	UIUC (ours: 0.930±0.021, 0.967)						
FeEX	0.821 +	0.498 +	0.808 +	0.900 +	0.902 +	0.871 +	0.896 +
Hist	0.557 +	0.498 +	0.592 +	0.647 +	0.575 +	0.598 +	0.638 +
LBP	0.819 +	0.498 +	0.850 +	0.878 +	0.881 +	0.857 +	0.922 +
HOG	0.896 +	0.618 +	0.780 +	0.894 +	0.914 +	0.887 +	0.913 +
CNN-5	mean±Std: 0.862±0.035 + ; max: 0.900						
	SCENE (ours: 0.934±0.016, 0.960)						
FeEX	0.885 +	0.691 +	0.871 +	0.926 +	0.853 +	0.888 +	0.896 +
Hist	0.802 +	0.691 +	0.829 +	0.847 +	0.797 +	0.830 +	0.866 +
LBP	0.905 +	0.691 +	0.915 +	0.965 -	0.960 -	0.946 -	0.946 -
HOG	0.945 -	0.797 +	0.805 +	0.932 =	0.965 -	0.965 -	0.937 =
CNN-5	mean±Std: 0.878±0.034 + ; max: 0.902						
	FACES (ours: 0.884±0.035, 0.945)						
FeEX	0.196 +	0.548 +	0.509 +	0.409 +	0.653 +	0.517 +	0.561 +
Hist	0.315 +	0.666 +	0.339 +	0.311 +	0.500 +	0.626 +	0.483 +
LBP	0.249 +	0.327 +	0.423 +	0.541 +	0.318 +	0.454 +	0.583 +
HOG	0.352 +	0.888 -	0.901 -	0.913 -	0.980 -	0.925 -	0.888 -
CNN-5	mean±Std: 0.712±0.067 + ; max: 0.810						
	COIL-20 (ours: 0.956±0.028, 1.00)						
FeEX	1.00 -	0.714 +	1.00 -	1.00 -	0.947 +	1.00 -	1.00 -
Hist	1.00 -	0.666 +	1.00 -	1.00 -	1.00 -	1.00 -	1.00 -
LBP	0.972 =	0.405 +	0.960 =	1.00 -	0.739 +	1.00 -	0.955 =
HOG	1.00 -	1.00 -	0.952 +	0.948 +	1.00 -	0.903 +	0.952 +
CNN-5	mean±Std: 0.892±0.016 + ; max: 0.831						

of extracted features and the corresponding classifier is crucial for image classification. Taking the COIL-20 data set as an example, the result of FeEx+MLP is the worst (69.05%), while the results with other classifiers are very good (100%). HOG has a similar pattern on the FACES data set. The extensions have two aspects: the first is the consideration of various content of different image data sets, and the second consideration of properties of diverse feature extraction methods and classifiers. Domain knowledge and human intervention would be required for effective image classification. The developed GP program can obtain promising results on different types of image data sets by automatically selecting suitable image descriptors. Compared with CNN-5, the proposed method achieves significantly better performance. This maybe because image classification methods based on CNNs tend to need a large number of training instances [36] while the data sets in our experiments are relatively small. Therefore, the classification results of CNN are not satisfying.

C. Further Analysis

The GP program for image classification has good interpretability. This section will analyze two examples from SCENE and UIUC to understand how they perform high-precision classification tasks.

1) An example GP program on the SCENE data set: Fig. 3 and Fig. 4 show examples evolved by the proposed approach on the data set SCENE with 95.65% and 96.38% classification accuracy in training and testing. Two diverse edge regions in an image are detected by the evolved GP program utilizing **PrewittY**. Because these two regions in the images among the two classes have different edge structural features, different standard deviations could be gained. Further, the class label could be obtained according to the output of the GP program tree. Specifically, if the result is greater than 0, the image belongs to the highway class. Otherwise, it belongs to the street class.

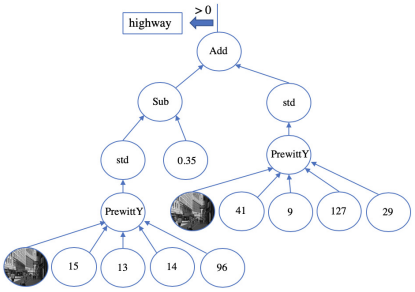


Fig. 3. An example evolved by the proposed approach on SCENE.

2) An example GP program on the UIUC data set: Fig. 5 and Fig. 6 display a evolved GP program on the UIUC that achieves 96.57% and 96.20% classification accuracy in training and testing. Two types of texture detectors (**LBP** and **FD**) and edge detectors (**PrewittX** and **PrewittY**) are employed to extract various features in different regions in an image for classification. Due to the extracted rich image features, the proposed approach achieves higher classification accuracy over compared methods.

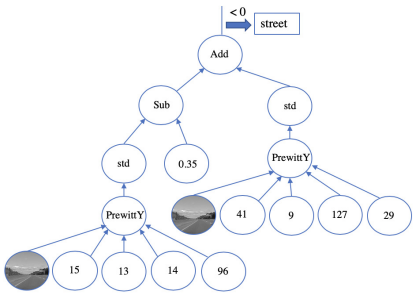


Fig. 4. An example evolved by the proposed approach on SCENE.

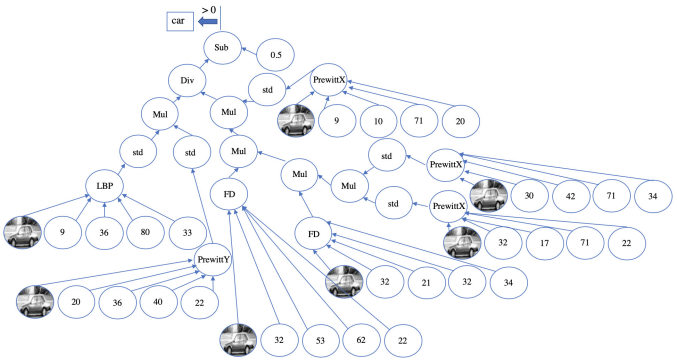


Fig. 5. An example evolved by the proposed approach on the UIUC dataset.

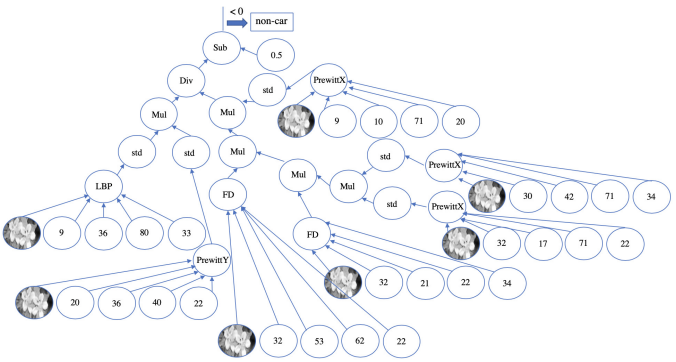


Fig. 6. An example evolved by the proposed approach on the UIUC dataset.

In summary, the proposed GP program could efficiently extract discriminative features according to the content of images. It could only use one type of edge detector for classifying the highway and street images. In addition, when the image has complex structural characteristics such as car detection, it could evolve the best combinations of various texture and edge descriptors for achieving high-accuracy image classification.

VI. Conclusions and Future Work

In this paper, we develop a new GP method for region selection, feature extraction, and image classification automatically and simultaneously, which combines diverse texture

and edge descriptors to extract various rich image features. The proposed approach can perform well on different types of image classification tasks due to the extracted adaptive features. When comparing with GP-based and non-GP based image classification methods, our approach achieves better or competitive performance. The program trees of some good individuals are analyzed to better understand how they produce useful features based on the content of images.

In the future, we would like to extend our work to multi-class image classification by utilizing useful image descriptors to extract advanced image features. A conventional classification method will be employed based on these features captured from the designed GP program. In addition, the performance of the proposed method tends to be affected by selected regions. For example, images belonging to the same class with different patterns might be classified into different classes. We aim to address this issue in the future.

REFERENCES

[1] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015, pp. 815–823.

[2] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2005, pp. 886–893.

[3] Y. Song, W. Cai, H. Huang, Y. Zhou, D. D. Feng, Y. Wang, M. J. Fulham, and M. Chen, "Large margin local estimate with applications to medical image classification," *IEEE Transactions on Medical Imaging*, vol. 34, no. 6, pp. 1362–1377, June 2015.

[4] 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, July 2014.

[5] 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.

[6] M. Hassaballah, A. A. Abdelmgeid, and H. A. Alshazly, "Image features detection, description and matching," in *Image Feature Detectors and Descriptors : Foundations and Applications*. Springer, 2016, pp. 11–45.

[7] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions." *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996.

[8] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[9] H. Bay, T. Tuytelaars, and L. V. Gool, "Surf: Speeded up robust features," in *Computer Vision – ECCV 2006*, 2006, pp. 404–417.

[10] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, June 2005, pp. 886–893.

[11] Y. Bi, M. Zhang, and B. Xue, "Genetic programming for automatic global and local feature extraction to image classification," in *2018 IEEE Congress on Evolutionary Computation (CEC)*, July 2018, pp. 1–8.

[12] J. R. Koza, "Genetic programming as a means for programming computers by natural selection," *Statistics and Computing*, vol. 4, no. 2, pp. 87–112, June 1994.

[13] H. Al-Sahaf, M. Zhang, M. Johnston, and B. Verma, "Image descriptor: A genetic programming approach to multiclass texture classification," in *2015 IEEE Congress on Evolutionary Computation (CEC)*, May 2015, pp. 2460–2467.

[14] H. Al-Sahaf, M. Zhang, A. Al-Sahaf, and M. Johnston, "Keypoints detection and feature extraction: A dynamic genetic programming approach for evolving rotation-invariant texture image descriptors," *IEEE Transactions on Evolutionary Computation*, vol. 21, no. 6, pp. 825–844, Dec 2017.

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

[16] H. Al-Sahaf, A. Song, K. Neshatian, and M. Zhang, "Extracting image features for classification by two-tier genetic programming," in *2012 IEEE Congress on Evolutionary Computation (CEC)*, June 2012, pp. 1–8.

[17] 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 *Genetic Programming*. Springer, 2016, pp. 51–67.

[18] B. Evans, H. Al-Sahaf, B. Xue, and M. Zhang, "Evolutionary deep learning: A genetic programming approach to image classification," in *2018 IEEE Congress on Evolutionary Computation (CEC)*, July 2018, pp. 1–6.

[19] A. Lensen, H. Al-Sahaf, M. Zhang, and Bing Xue, "A hybrid genetic programming approach to feature detection and image classification," in *2015 International Conference on Image and Vision Computing New Zealand (IVCNZ)*, Nov 2015, pp. 1–6.

[20] H. Al-Sahaf, M. Zhang, and M. Johnston, "Genetic programming for multiclass texture classification using a small number of instances," in *Simulated Evolution and Learning*. Springer, 2014, pp. 335–346.

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

[22] A. Almeida and R. Torres, "Remote sensing image classification using genetic-programming-based time series similarity functions," *IEEE Geoscience and Remote Sensing Letters*, vol. PP, pp. 1–5, July 2017.

[23] S. R. Price, D. T. Anderson, and S. R. Price, "Goofed: Extracting advanced features for image classification via improved genetic programming," in *2019 IEEE Congress on Evolutionary Computation (CEC)*, June 2019, pp. 1596–1603.

[24] Y. Zhang, Q. Fan, F. Bao, Y. Liu, and C. Zhang, "Single-image super-resolution based on rational fractal interpolation," *IEEE Transactions on Image Processing*, vol. 27, no. 8, pp. 3782–3797, Aug 2018.

[25] S. Novianto, Y. Suzuki, and J. Maeda, "Near optimum estimation of local fractal dimension for image segmentation," *Pattern Recognition Letters*, vol. 24, no. 1, pp. 365 – 374, 2003.

[26] Z. Guo, L. Zhang, and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1657–1663, June 2010.

[27] L. Fei-Fei and P. Perona, "A bayesian hierarchical model for learning natural scene categories," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 2, June 2005, pp. 524–531 vol. 2.

[28] D. J. Montana, "Strongly typed genetic programming," *Evol. Comput.*, vol. 3, no. 2, p. 199–230, Jun. 1995.

[29] B. Mandelbrot, W. Freeman, and Company, *The Fractal Geometry of Nature*, ser. Einaudi paperbacks. Henry Holt and Company, 1983.

[30] W.-L. Lee and K.-S. Hsieh, "A robust algorithm for the fractal dimension of images and its applications to the classification of natural images and ultrasonic liver images," *Signal Process.*, vol. 90, no. 6, p. 1894–1904, Jun. 2010.

[31] S. Agarwal, A. Awan, and D. Roth, "Learning to detect objects in images via a sparse, part-based representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 11, pp. 1475–1490, Nov 2004.

[32] C. E. Thomaz, "Fei face database," https://fei.edu.br/~cet/facedatabase.html.

[33] S. A. Nene, S. K. Nayar, and H. Murase, "Columbia object image library (coil-20)," Tech. Rep., 1996.

[34] M. Zhang and V. Ciesielski, "Genetic programming for multiple class object detection," in *Advanced Topics in Artificial Intelligence*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1999, pp. 180–192.

[35] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, and et al., "Scikit-learn: Machine learning in python," *J. Mach. Learn. Res.*, vol. 12, p. 2825–2830, Nov. 2011.

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