FLGP Summary

Introduction:-

______FLGP(Feature Learning using Genetic Programming for Image Classification) is a method introduced by Ying Bi, Bing Xue and Mengjie Zhang that uses the idea of evolution based programming(Genetic programming) to increase the efficiency of image in both binary classes and multi-class field while keeping a relatively low run time. The tests conducted prove that the research was majorly successful and increased the efficiency of image classification as compared to many traditional features while keeping a relatively low runtime.

FLGP uses a classic genetic based approach using its own Program structure containing function sets and terminal sets.

Overall Algorithm:-

FLGP program starts with a basic population initialization consisting of randomly generated programs/trees based on the new program structure, that is then evaluated using the fitness evaluation process. After fitness evaluation, a selection method consisting of three genetic operators (elitism, crossover and mutation) are used to select the best performing programs that are fed into the next generation to generate more trees based on those selected trees.
This operation is performed a fixed amount of times. After that the best individual is returned. Fitness evaluation is performed by feeding the data into a linear SVM along with a vector transformed image and its labels. Linear classification is employed as it has fewer parameters as compared to other SVMs. Stratified k-fold cross validation is also used on each individual to increase generalization capability and k is set to 5 to reduce computational cost. Features are also normalized as different feature extraction methods are employed with varying max-min. Min-max normalization is used to map the features between 0 to 1.

Function Set:-

As mentioned above, as per the Genetic Programming approach, FLGP program structure uses its own custom function set and terminal set. The function set according to the new program structure consists of three major parts.

1) Region detection functions.
2) Feature extraction functions.
3) Feature concatenation functions.

Each of these will be explained separately.

Region Detection Functions:-

Region Detection functions have a basic role of detecting a region containing the needed portion from where most of the features can be extracted. This reduces noise in the image and feeds the feature extraction methods only the portion it needs to extract features from allowing features to be more cleaner to begin with. Two region extraction methods are employed, namely *Region_S* and *Region_R* that extract a square region and a rectangular region respectively. Below piece of lines are directly copied from the research paper that describes the parameters these functions take.

“The Region S function takes an image Image (Ii , the size is m × l), X, Y , and Size as inputs and returns a square region. The coordination of the top-left point of the region in Image is (X, Y) and the size of the region is Size. Thus the detected region by Region S is Image[X : min((X+Size), m), Y : min((Y + Size), l)]. Similar to Region S, the Region R function detects the Image[X : min((X +W idth), m), Y : min((Y + Height), l)] region by taking the Image, X, Y , W idth, and Height as inputs. In the two functions, the Image, X, Y , Size, W idth, and Height are terminals, which will be described in the next subsection. The values of these terminals are randomly generated from predefined ranges and can be changed by the mutation operator during the evolutionary learning process”

Feature Extraction Functions:-

FLGP uses pre coded feature extraction methods in its function set to extract varying quality of features that are used to classify the given data set. These image descriptors are:-

1) DIF
2) Hist
3) SIFT
4) HOG
5) uLBP

Features are kept as minimum as possible to avoid a big search space. In FLGP, the functions used for these image descriptors are as follows:-

1) G_DIF, L_DIF
2) G_Hist, L_Hist
3) G_SIFT, L_SIFT
4) G_HOG, L_HOG

5) G_uLBP, L_uLBP

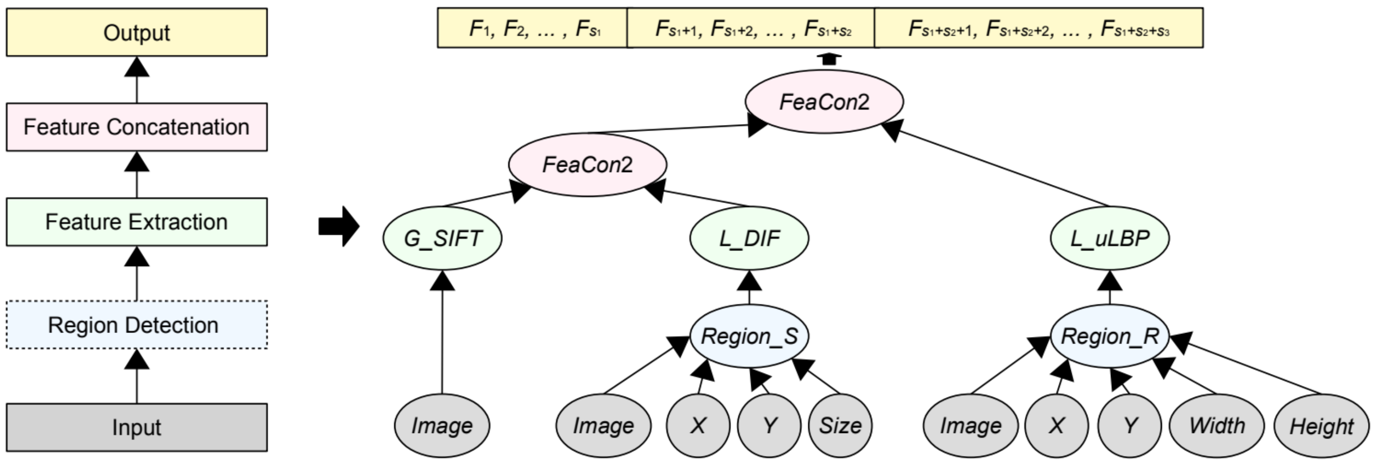
The G and L prefixes stand for global and local respectively, as the region detection methods mentioned above could either give a region within an image which is then fed into local feature extraction functions to extract features or we can extract features from entire images using the global functions.
The global functions directly use the Image terminal while the local functions use the region detection functions as their children nodes. All of these feature extraction methods extract a set(except one method) number of features, these are given below.

TABLE I
FEATURE EXTRACTION FUNCTIONS.

Methods	**Input**	**Output**	**#Features s**	**Description**
G_DIF/L_DIF	1 Image/Region	1 Vector	20	Domain independent features [15].
G_Hist/L_Hist	1 Image/Region	1 Vector	32	Histogram features of the image/region [14]. The number of bins is set to 32.
G_SIFT/L_SIFT	1 Image/Region	1 Vector	128	SIFT features. The image or detected region is considered as a keypoint [16].
G_HOG/L_HOG	1 Image/Region	1 Vector	Flexible	HOG features [5]. G_HOG/L_HOG extracts the mean value of each 20×20 / 10×10 grid with a step of 10 from a HOG image.
G_uLBP/L_uLBP	1 Image/Region	1 Vector	59	Uniform LBP histogram features [7]. In G_uLBP and L_uLBP, the radius is 1.5 and the number of neighbors is 8.

Feature Concatenation Functions:-

At the end, all the extracted features need to be combined, the feature concatenation functions are employed for this specific task .Two feature concatenation functions are used, namely *FeaCon2* and *FeaCon3* which concatenates two feature vectors and three feature vectors respectively. Both of these functions use the feature extraction functions and/or the feature concatenation functions are their children nodes. A proper structure of how this all works together is given below.



(a) Program Structure (b) Example Program

Terminal Set:-

Terminal set is basically all the parameters needed by the function set. There are a total of six different terminals(note that these terminals are the basic parameters required outside of the function set, as the function set also uses its own methods as parameters sometimes). These terminals are as follows:-

1) Image
2) X
3) Y
4) Size
5) Width
6) Height

The Image terminal is a gray-scale image, which is a two_dimensional array (mxl) with values in the range of [0, 1] (indicating the intensity of 1), the image is normalized by dividing by 255.
The other terminals are random constants pre generated as per the evolution operators. The following lines below are copied from the research paper that marks the specifics of the other five terminals.

"X and Y indicate the coordinates of the top-left point of a detected region in the image and are the parameters of the Region S and Region R functions. They are integers in the range of [0, m − 20] and [0, l − 20], respectively. The Size, W idth and Height terminals are the size or width and height of a detected region. Their values are in the range of [20, 50], which is smaller than that in [13] to narrow the search space."

Testing Phase:-

This model was tested upon eight different data sets of varying difficulty, these data sets contained both binary class data and multi-class data. The data sets contained five types of tasks:

1) Facial Expression Recognition (FEI_1 [27], FEI_2 [27], JAFFE [28])
2) Object classification (EYALE [29], ORL [30])
3) Screen classification (SCENE [31])
4) Texture classification (KTH [32])
5) Painting classification (VGDB [33])

The given reference points above are of the research paper.
Below are the specifics of all these data sets cited directly from the research paper.

"FEI 1 and FEI 2 [27] contain frontal facial images with natural or smile expression. The images in the two data sets are sampled from 200 Brazilian with different appearance, hairstyle and adorn. VGDB is to identify Vincent Van Gogh's paintings [33], which is very challenging because there are not particular objects in the images and the painting style is hard to capture. ORL [30] is to recognize faces of 40 different people from images with open or closed eyes, smiling or non-smiling, and glasses or non-glasses. JAFFE [28] has 213 images of 7 different expressions sampled from 10 Japanese females. The seven expressions are happiness, surprise, sadness, fear, anger, natural, and disgust. KTH [32] is a texture classification task of 10 classes. The images are sampled in nine scales with three poses under four illumination conditions. EYALE [29] is a face classification task, having 2424 facial images of 38 different people. The facial images are sampled under different poses and illumination conditions. SCENE [31] contains 3859 natural images in 13 groups, including the coast, forest, highway, mountain, and street. The natural images are acquired under different conditions and have high variations, which makes the task difficult."

The division of these data sets into training and testing sets and their number of classes are given in the table below.

TABLE II
DATA SET PROPERTIES.

Data set	Image size	# Classes	Training set	Test set
FEI_1	130×180	2	150	50
FEI_2	130×180	2	150	50
VGDB	200×200	2	247	83
ORL	112×92	40	280	120
JAFFE	128×128	7	140	73
KTH	100×100	10	600	210
EYALE	100×100	38	1,209	1,215
SCENE	100×100	13	1,928	1,931

Benchmark Methods:-

For testing the model, five GP-based methods, eight traditional metohds and three CNN methods were employed. These are as follows:-

1) GP-based Methods: Five GP-based methods that were used were as follows
    a) GP-GLF [13]
    b) 2TGP [34]
    c) DIF+GP [15]
    d) Hist+GP
    e) uLBP+GP [21]

    DIF+GP, Hist+GP, uLBP+GP construct features from pre-extracted features, i.e 20 DIF features, 64 Hist features and 59 uLBP features, respectively. The comparison of GP-based methods and FLGP is done using only binary image classification tasks as running the GP based methods on multi-class image classification task is too demanding(The GP based methods were originally designed for binary classification task so that is where they excel).
2) Traditional methods: Eight traditional methods using different well-known features were employed, namely:-
    a) DIF [15]

b) Hist [14]
c) GLCM [17]
d) Gabor [35]
e) SIFT [6]
f) HOG [5]
g) LBP [7]
h) uLBP [7] features.

These features are then fed into a linear SVM for classification. Given below are details of how these features are extracted, cited directly from the research paper.

"The methods for extracting these features have been introduced in Section II. The DIF, SIFT, HOG, and ulBP features are extracted using the same functions as those employed in the function set of FLGP in the global scenario. The Hist features are 256 histogram features and the LBP features are 256 LBP histogram features. The GLCM features are the statistics of each GLCM, i.e., contrast, dissimilarity, homogeneity, energy, correlation, and ASM. Each GLCM is calculated using four different orientations ($\pi\mu/4$, $\mu \in \{0, 1, 2, 3\}$). The Gobor features are the mean values of each 32×32 grid of the convolved images using different Gabor filters. Forty commonly used Gabor filters are used, involving eight different orientations ($\pi\mu/8$, $\mu \in \{0, . . . , 7\}$) at five scales ($v \in \{0, . . . , 4\}$) [35]."

3) CNN-based Methods: Three CNN based methods with different architectures are employed. Namely:-
   a) LeNet-55 [36]
   b) Five-layer CNN (CNN-5) [8]
   c) Eight-layer CNN (CNN-8)

   All of these three metohds use the rectified linear unit (ReLU) as the activation function and the softmax function for classification. Dropout is added after the pooling layer and the first fully connected layer with 0.25 and 0.5 probabilities to avoid overfitting. The loss function is cross-entropy and the adaptive subgradient method is used to train these models. Epochs is set to 500.

Parameters:-

The same parameters are used as the GP-GLF method. The few differences in the parameters are cited below:-

"The population size for the four GP-based methods is 500, while FLGP and GP-GLF use a smaller size of 100 in order to reduce the computational cost. The crossover, mutation, and elitism rates are 0.8, 0.19, and 0.01, respectively. The selection method is Tournament selection with size 7. The tree generation method is ramped-half-and-half.

The tree depth is between 2 and 6. The termination criterion for all the GP methods is reaching the maximum number of generations."

Results:-

The results are shown in the table below:-

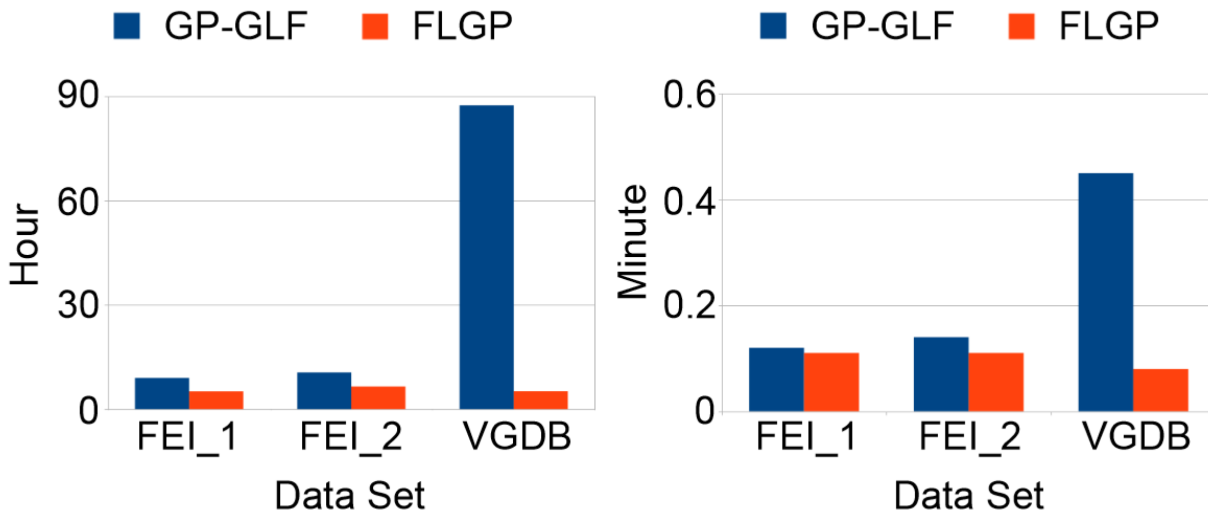
TABLE III
CLASSIFICATION ACCURACY (%) OF THE PROPOSED FLGP APPROACH AND SIXTEEN BENCHMARK METHODS ON THREE BINARY DATA SETS: FEI_1, FEI_2 AND VGDB.

	FEI_1		**FEI_2**		**VGDB**	
Methods	Max	Mean±Std.	Max	Mean±Std.	Max	Mean±Std.
2TGP	96.0	88.1±6.2+	94.0	85.5±6.0+	63.9	61.6±1.5+
DIF+GP	80.0	56.7±6.9+	72.0	60.3±8.4+	68.7	61.4±3.5+
Hist+GP	70.0	48.9±7.2+	60.0	48.8±6.1+	84.3	76.0±2.6=
uLBP+GP	66.0	50.9±7.5+	72.0	48.7±7.9+	79.5	69.4±4.4+
GP-GLF	96.0	89.1±3.9+	92.0	82.5±5.8+	74.6	65.1±4.7+
DIF	74.0	61.1±4.9+	72.0	62.8±6.1+	66.3	55.6±10.3+
Hist	54.0	48.1±3.4+	54.0	50.1±2.5+	62.7	62.2±0.8+
GLCM	50.0	49.7±0.7+	54.0	50.1±0.7+	62.7	53.3±9.8+
Gabor	82.0	71.6±7.9+	74.0	65.7±5.1+	63.9	56.0±8.3+
SIFT	82.0	82.0±0.0+	78.0	78.0±0.0+	60.2	60.2±0.0+
HOG	94.0	94.0±0.0+	88.0	88.0±0.0+	57.8	57.2±0.7+
LBP	68.0	62.5±3.5+	66.0	57.6±3.6+	84.3	80.6±3.2–
uLBP	64.0	56.9±5.2+	56.0	51.9±2.3+	81.9	71.5±8.1=
LeNet-5	98.0	94.4±2.0=	94.0	90.8±1.8+	65.1	58.1±4.8+
CNN-5	98.0	95.6±1.5=	90.0	85.0±3.0+	65.1	61.5±2.1+
CNN-8	98.0	94.2±2.1=	94.0	90.0±2.3+	61.5	56.9±4.6+
FLGP	98.0	95.8±3.2	100	93.3±3.8	81.9	74.9±3.7
Overall		**13+, 3=**		**16+**		**13+, 2=, 1–**

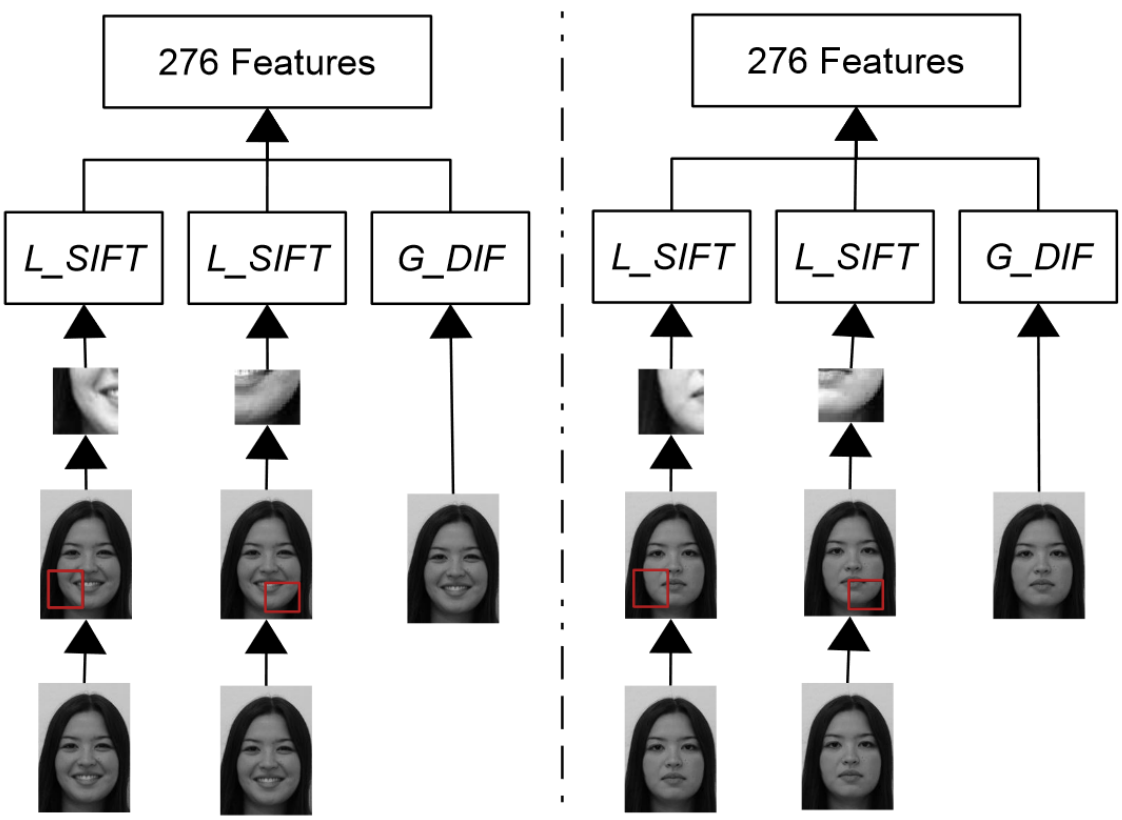
TABLE IV
CLASSIFICATION ACCURACY (%) OF THE PROPOSED FLGP APPROACH AND ELEVEN BENCHMARK METHODS ON FIVE MULTI-CLASS DATA SETS: ORL, JAFFE, KTH, EYALE, AND SCENE.

	ORL		**JAFFE**		**KTH**		**EYALE**		**SCENE**	
Methods	Max	Mean±Std.	Max	Mean±Std	Max	Mean±Std.	Max	Mean±Std.	Max	Mean±Std.
DIF	85.0	85.0±0.0+	35.6	35.6±0.0+	56.2	56.2±0.0+	26.4	26.4±0.0+	33.5	33.5±0.0+
Hist	97.5	97.5±0.0+	19.2	19.2±0.0+	51.4	51.4±0.0+	11.0	11.0±0.0+	21.2	21.2±0.0+
GLCM	2.5	2.5±0.0+	15.1	15.1±0.0+	23.3	23.3±0.0+	5.1	5.0±0.1+	13.9	13.9±0.0+
Gabor	59.2	57.1±0.9+	46.6	43.2±1.6+	44.3	42.9±0.7+	36.7	36.3±0.2+	22.7	22.3±0.2+
SIFT	98.3	98.3±0.0+	74.0	74.0±0.0+	81.4	81.4±0.0+	88.4	88.4±0.0+	63.1	63.1±0.0+
HOG	96.7	96.7±0.0+	72.6	72.6±0.0+	51.4	51.4±0.2+	74.3	74.3±0.0+	30.8	30.8±0.0+
LBP	87.5	87.5±0.0+	21.9	21.9±0.0+	87.6	87.6±0.0+	46.4	46.4±0.0+	62.5	62.5±0.0+
uLBP	94.2	94.2±0.0+	23.3	23.3±0.0+	81.0	81.0±0.0+	56.1	56.1±0.0+	66.1	66.1±0.0+
LeNet	93.3	89.9±1.9+	79.5	68.9±7.0+	78.6	72.0±6.4+	92.4	89.4±1.6+	54.2	51.1±2.4+
CNN-5	97.5	96.3±0.8+	80.8	78.9±1.3+	84.3	81.5±1.8+	99.3	98.6±0.5+	58.9	55.4±1.4+
CNN-8	96.7	94.2±1.8+	61.6	52.5±6.2+	82.4	80.7±1.5+	90.9	88.2±1.0+	69.2	66.2±2.0+
FLGP	100	99.6±0.7	91.8	81.1±4.7	95.7	94.5±0.9	99.8	99.2±0.4	77.0	75.2±0.7
Overall		**11+**		**11+**		**11+**		**11+**		**11+**

Computational Cost, FLGP vs GP-GLF:-



Example Solution:-



FeaCon3(L_SIFT(Region_S(Image, 112, 11, 50)), *L_SIFT(Region_R(Image,* 130, 64, 38, 46)), *G_DIF(Image))*

Feature Extraction Functions and their Frequency:-

After further analysis it was deduced that each of the feature extraction functions had their own uses in different data sets, and the functions being less frequently used on a certain type of data set was used heavily on another data set. The frequency of each function is given in the table below in descending order.

TABLE V

RANKING OF ALL THE FEATURE EXTRACTION FUNCTIONS IN 300 BEST-OF-THE-RUN PROGRAMS OF FLGP ON EACH DATA SET.

Function	FEI_1	FEI_2	VGDB	ORL	JAFFE	KTH	EYALE	SCENE
G_DIF	5	9	7	7	7	5	8	4
G_Hist	10	10	3	5	9	6	5	6
G_SIFT	8	5	8	1	2	2	3	2
G_HOG	3	4	10	4	5	7	9	9
G_uLBP	9	7	1	2	10	1	4	1
L_DIF	6	6	5	9	3	4	7	7
L_Hist	4	8	2	6	8	3	6	8
L_SIFT	1	1	9	10	1	10	1	10
L_HOG	7	3	4	8	4	9	10	5
L_uLBP	2	2	6	3	6	8	2	3

Proposed Changes:-

1) The parameters can be adjusted to find a better performance on relatively same computational cost(note that computational cost should be kept relatively the same or lower if possible).
2) The VGDB image data set can be studied and appropriate solutions can be found accordingly to increase the performance of FLGP on these images as well.