International Journal of Network Dynamics and Intelligence Scilight

Article

Conditional Generative Adversarial Net based Feature Extraction along with Scalable Weakly Supervised Clustering for Facial Expression Classification

Ze Chen [1], Lu Zhang [2], Jiaming Tang [3], Jiafa Mao [3], and Weiguo Sheng [1,*]

[1] Department of Computer Science, Hangzhou Normal University, Hangzhou 311121, China
[2] China Telecom Hangzhou Branch, Hangzhou 310016, China
[3] School of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310027, China
[*] Correspondence: w.sheng@ieee.org

Received: 28 September 2023
Accepted: 30 June 2024
Published: 24 December 2024

Abstract: Extracting proper features plays a pivotal role in facial expression recognition. In this paper, we propose to extract facial expression features via a conditional generative adversarial net, followed by an algorithmic optimization step. These refined features are subsequently integrated into a scalable weakly supervised clustering framework for facial expression classification. Our results show that the proposed method can achieve an average recognition rate of 85.3%, which significantly outperforms related methods. Further, by employing a residual-based scheme for feature extraction, our method shows superior adaptability compared to algorithms based solely on weakly supervised clustering. Additionally, our method does not require high accurate annotation data and is robust to the noise presented in data sets.

Keywords: facial expression recognition; neutral expression; feature extraction; weakly supervised clustering

1. Introduction

Facial expression recognition, which is an important branch of face recognition, has received much concern within fields such as computer vision and human-robot interaction. Facial expression encompasses various muscle movements [1] with at least 21 distinct types. However, only a subset of 6 expressions can be intuitively recognized by humans, as depicted in Figure 1. It has been well known that facial expression recognition is inherently challenging. Existing methods for facial expression recognition are generally based on the supervised approach (i.e., training on fully annotated data) or semi-supervised approach (i.e., training on partially annotated data). When confronted with a large volume of inaccurate annotations, recognition accuracy could be severely compromised. The process of annotating facial expression image datasets (e.g., CK+ [2], SEMAINE [3] and AM-FED [4]), often presents difficulties in discerning the correctness of expert-provided labels through preprocessing models. Hence, there is a pressing need to investigate facial expression recognition from weakly annotated data.

Facial expression recognition typically encompasses a four-step process: image acquisition, face detection, feature extraction and classification [5]. Among these, feature extraction holds a pivotal role in determining the performance of recognition systems. Existing feature extraction algorithms can be broadly classified into two categories: the static and dynamic methods. The static method can be further divided into the global and local methods, whereas dynamic methods consist of the optical flow method, model-based method and geometric method [1]. Along with the advancement of neural network technology, it becomes feasible to employ such technology for facial expression feature extraction [2]. In this paper, we tend to first employ a conditional Generative Adversarial Net (cGAN) [6] to generate a neutral expression image from the original expression image. Then, we extract feature values from both the generated neutral expression image and the original expression image. The discrepancies between these two sets of feature values are then employed as the feature representation for the original expression image.

Figure 1. Common facial expressions (from top to bottom and left to right, the images show expressions of joy, sadness, surprise, disgust, anger and fear).

cGAN is based on the Generative Adversarial Networks (GANs) by introducing conditions to the model. These conditions serve as guidance for cGAN, directing the network towards generating images aligned with the specified conditions. Despite the capabilities of the generator, there may still exist certain disparities between the generated image and the target image. To mitigate the impact of discrepancies in image details, we incorporate a geometric approach for obtaining specific feature values. Our emphasis is on capturing geometric information pertaining to the facial key points rather than the texture details. By doing so, we aim to effectively address the challenges arising from the low accuracy of images generated by cGAN.

The contribution of our work is two-fold:

1) A novel method is proposed to learn facial expressions from weakly annotated data. This method first trains a generative model to generate a neutral facial image based on the input image and subsequently extracts geometric features from pairs of such images. The extracted features are then integrated into a scalable weakly supervised clustering algorithm for facial expression classification. The utilization of neutral expressions in this process could lead to significant enhancement in recognition accuracy.

2) In the proposed method, geometric feature differences observed in facial key points between the expression image and its corresponding neutral expression image are employed to characterize facial expressions. By doing so, issues related to identity-related variations and the impact stemming from differences in image details could be effectively mitigated.

An overview of the proposed method is shown in Figure 2. The proposed method involves four primary steps. First, we employ a generator to generate corresponding neutral expression images from original facial expression images. In this step, facial expression images are served as conditions for the generator. Then, neutral expression images and their corresponding facial expression images are used in pairs as inputs to train discriminator. In each paired images, images will compete with each other to generate good-quality neutral expression images. The details of above step will be described in Section 3. In second and third steps, the active appearance method (AAM) will be employed to label key points of the face. This is followed by employing a geometric method to obtain two sets of feature values from facial expression images and neutral expression images based on relative positions of key points. Then, differences between these two sets will be calculated and optimized using an information gain rate scheme. The details of these two steps will be presented in Section 4. Finally, in the fourth step, a weakly supervised clustering will be employed for face expression classification, which will be described in Section 5. The experimental results and conclusions will be shown in Sections 6 and 7, respectively.

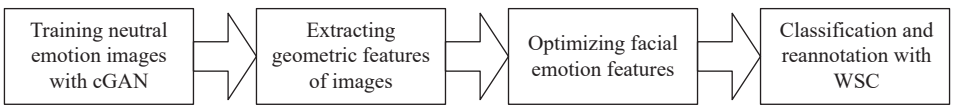


Figure 2. Overview of the proposed method.

2. Related work

Facial expression recognition is an important research topic in pattern recognition. Many facial expression recognition methods have been proposed. These methods typically employ a directional histogram (HOG) with discrete wavelet transform (DWT) for facial feature extraction, and then employ methods such as cropping and standardization for preprocessing. Finally, a support vector machine (SVM) classifier is employed for classification [7].

Recently, neural networks have been widely employed for facial expression recognition [8, 9]. For example, Sun et al. [10, 11] proposed a convolutional neural network (CNN) based method with 11 layers for facial expression recognition. This method obtains convolution properties from facial images for classification. Studies by Kim et al. [12] and Zafeiriou et al. [13] showed that the usage of neutral face images can improve the efficiency of facial expression recognition. The facial expression image, which subtracts its corresponding neutral expression image at the pixel level or feature level, can reduce the intra-class variation while highlighting the facial expression. Bazzo et al. [14] obtained a good recognition rate by using Gabor wavelet to recognize facial expressions. Zafeiriou et al. [13] applied the sparse facial expression representation to the image, which is obtained by subtracting the neutral image from the expression image. The results show that the usage of neutral images can emphasize the moving part of face.

Extracting facial features are mainly relevant to identifying positions of facial components [15−17]. Based on local feature representation and Bezier curves, Hong et al. [18] tracked and described facial components such as eyebrows and noses with Bezier control points to extract facial features. This method can achieve a good recognition rate. Ghahari et al. [19] applied Canny Edge Detector method to local face image after face positioning step. In this method, a hierarchical clustering-based scheme is used to strengthen the search area of extracted high-texture face clusters to construct expression feature vector. Yang et al. [20] proposed a De-expression Residue Learning network (DRL) based on a GAN. The idea is that, during the process of generating neutral expression images from expression images, expressive components of facial expressions will remain in the network framework. In this method, the authors try to extract expressive components of facial expressions as feature values.

Weakly supervised learning methods employ inaccurate, incomplete or inaccurate annotations (i.e., weak annotations) and has proved its effectiveness in solving computer vision tasks. For example, Bilen et al. [21] applied a SVM with convex clustering to locate windows with high probability of objects in the image set from noisy and incompletely annotated complex images. Prest et al. [22] proposed a method of weakly supervised learning to study the incomplete label of "action" in the process of human-object interaction. These methods can be used to learn from weakly annotated images (i.e., inaccurate annotations). At the same time, they can alleviate noisy annotations.

Additionally, Yuan et al. [23] leveraged knowledge graph technology for visual representation, offering users the reference and theoretical foundation in selecting methods for facial expression recognition research. Luo et al. [24] introduced an enhanced CNN model focusing on face image preprocessing, feature extraction, test sample training, feature acquisition, expression classification and face image restoration. Shiomi et al. [25] proposed explicit and implicit methods. The explicit method employs a classifier to identify expression types and regressors to determine intensity. In contrast, the implicit method assigns zero or non-zero values to regressors based on input image correctness for each facial expression type. Jin et al. [26] conducted a comparative analysis of three classic emotion recognition algorithms utilizing CNNs. Additionally, an image enhancement algorithm is devised integrating super resolution generative adversarial networks (SRGAN) and adaptive grayscale normalization, and such an algorithm is thus tailored to the dataset and Efficient Convolutional Neural Networks (MobileNet) characteristics.

3. Neutral expression image generation

Neutral expressions, also called natural expressions, are expressions when people do not have any emotion. Any facial expression can be deemed as a combination of the neutral expression and the expression component. In other words, facial expressions can be broken down into neutral expressions and expression components.

To obtain a neutral expression image corresponding to a facial expression, we use an expression image and a neutral expression image as a pair of inputs to train a cGAN. Generally, a cGAN consists of two networks: a generator network G and a discriminator network D. The image generated by generator is discriminated by the discriminator. The two networks are continuously optimized in the confrontation. The final image generated by the generator will be close to the target image. Specifically, for an input image data I_{input}, the generator network will obtain $G(I_{input})$. The generator network structure is shown in Figure 3.

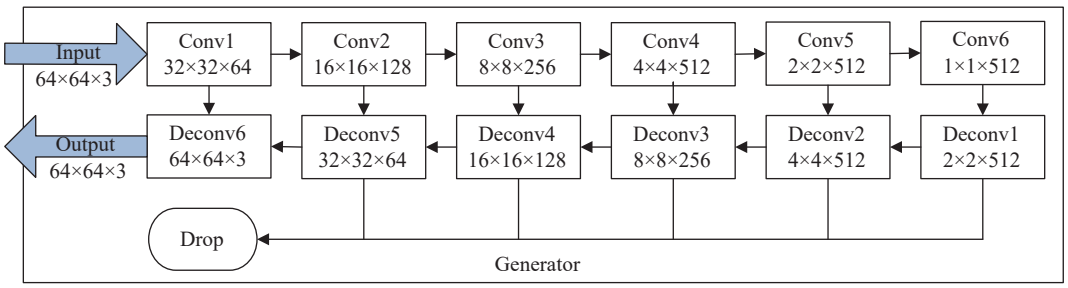


Figure 3. The diagram of generator network structure.

The generator network is based on a UNet architecture and is composed of 6 convolutional as well as 6 deconvolutional layers. The input is the image data with conditional parameters. In the process of deconvolution, the general output will be randomly dropped, and then connected to the same structured result of the output of convolution layer. The deconvolution layer is used to ensure the quality of network-generated image. After the output $G(I_{input})$ of generator network is connected to original input image I_{input}, and the target image I_{target} is connected to original input image I_{input}, the connected input pair will be added to discriminator network for identification. That is, the discriminator will be trained to output "yes" for input $<I_{target\,t}, I_{input}>$ and "no" for input $<G(I_{input}), I_{input}>$.

During training, we define the objective function of discriminator as:

$$L(D) = \frac{1}{N}\sum_{i=1}^{N}\left\{\log D\left(I_{target}, I_{input}\right) + \log\left(1 - D\left(G(I_{input}), I_{input}\right)\right)\right\} \tag{1}$$

where *N* denotes the number of input images. The objective function of generator is defined as:

$$L(G) = -\frac{1}{N}\sum_{i=1}^{N}\left\{\log D\left(G(I_{input}), I_{input}\right) + \theta_1 \| I_{target} - G(I_{input}) \|\right\} \tag{2}$$

The generator aims to generate image, which is similar to both input and target images. The parameter θ_1 in equation (2) is used to control the similarity between generated and target images. In summary, the final objective function of cGAN is defined as

$$L = \arg\left(\min_{G}\left(\max_{D}(L(D) + \theta_2 L(G)\right)\right) \tag{3}$$

The neutral expression image generated by cGAN is shown in Figure 4.

Figure 4. cGAN training results. The first column denotes the original input image, the second column is training output and the third column is target image.

4. Feature value extraction and optimization

The neutral expression images generated by cGAN may have certain difference with the target image in details. Consequently, such images are unsuitable for texture feature extraction. To deal with this issue, we consider geometric features of facial key points to describe expression images. The active appearance method (AAM) [27] is a framework that combines the shape, texture and other factors to identify facial key points. The framework can be used to identify a total of 68 key points of face. Among them, the eyebrows (5 points), inner eyebrow (4 points), eyes (6 points), nose (9 points) and mouth (20 points), as shown in Figure 5, are key points identified by AAM.

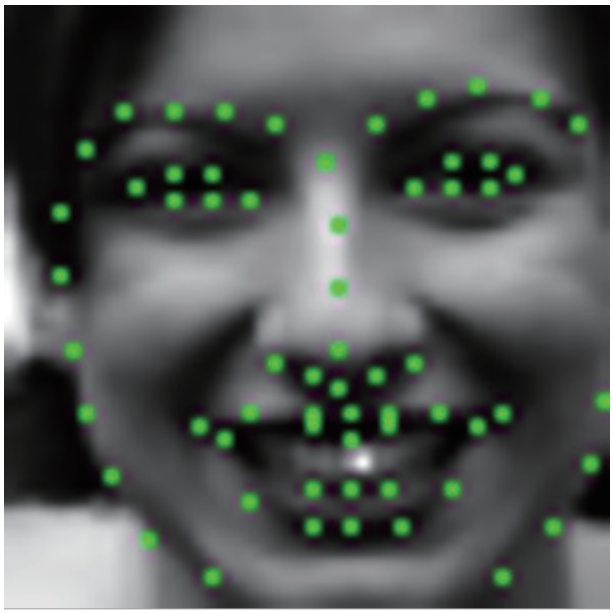


Figure 5. Face key point annotation via AAM.

4.1. Geometric feature value extraction

For the key points at different parts in face, different strategies have been adopted to extract geometric feature values. In the following subsections, we should describe extraction strategies used for different parts.

A) Upper part including eyebrows and eyes

For eyebrows and eyes, the degree of bending usually indicates the "violence" degree of the expression. We define two geometric features F1 and F2 to represent the eccentricity of fitting ellipse of facial points on left and right eyebrows, respectively. Eccentricity is measured by dividing the distance between the center of ellipse and the focus from the center to apex. Circle is an ellipse with zero eccentricity. A lower eccentricity means more curved shape of the object. The eccentricity is calculated as:

$$F1\,(\text{or}F2) = c_{eyebrow}/a_{eyebrow} \tag{4}$$

where $a_{eyebrow}$ is the semi-major axis of ellipse of the eyebrow point and $c_{eyebrow}$ is the semi-focal distance, which is the square difference between the semi-major axis a and semi-minor axis b of ellipse, that is, $c^2=a^2+b^2$.

While using key points of the eyebrow part to fit the ellipse, only the upper half of the ellipse is convered. Consequently, the fitting ellipse is not unique. To address this issue, point P_n will be added to the lower part of the ellipse based on original key points according to the following equation:

$$Pn = P1 + P5 - P3 \tag{5}$$

Here, $P1$, $P3$ and $P5$ are the first, middle and last point, respectively, of the fitting points. The symbols "+" and "-" here denote operations between point coordinates, that is, the corresponding point Pn of P3 is calculated using the midpoint of P1 and P5 as the symmetric point, and Pn is in the lower half of fitting ellipse.

The key points of the eye part can also be used to extract eccentricity features F3 and F4 in a similar way as the eyebrow part. Specifically, they are calculated as:

$$F3\,(\text{or}F4) = c_{eye}/a_{eye} \tag{6}$$

In addition to the eccentricity of the fitting ellipse, the distance of key points on the face is another measure of the "violence" of the expression. Here, we use the distance above and below as the eye socket. Generally, to extract facial expression features, rather than the distances, the ratio between distances is adopted as the ratio between distances largely includes the differences between individual faces. In addition the expression information, also contains individual identity information. In this work, we extract the ratio based on the neutral expression image and expression image. By using the difference between the distances of these two images, the influence of individual identity information contained in the distance information can be eliminated to a certain extent. Specifically, the following equation is employed to calculate the distances of left and right eye sockets, i.e., F5 and F6:

$$F5\,(\text{or}F6) = 2b_{eye} \tag{7}$$

B) Center part including nose

When the mood changes, key points of the bridge of nose basically do not change. By contrast, the lower end of the nose will change obviously. This change is usually reflected by the degree of bending. Here, the eccentricity of the fitted ellipse will be used to describe this change, which is defined as

$$F7 = c_{nose}/a_{nose} \tag{8}$$

In addition, we define feature F8 to describe the average distance between the lower end of the nose and the upper end of the mouth. This feature could be used to discriminate expressions with mouth changes such as anger and surprise. This feature is defined as

$$F8 = \arg\left(D(P_{nose}, P_{out\_mouth})\right) \tag{9}$$

Here, P_{nose} and P_{out_mouth} denote the points in the lower end of the nose and the upper end of outer mouth, respectively. The value of F8 reflects the Euclidean distance of corresponding points.

C) Lower part including mouth

The key points of mouth marked by AAM are mainly divided into outer and inner mouth contours. We consider the contour of the mouth as an ellipse, and calculate the eccentricity as a measure of the degree of curvature. The eccentricities of the outer and inner mouth contours are calculated as

$$F9 = c_{out\_mouth}/a_{out\_mouth} \tag{10}$$

$$F10 = c_{in\_mouth}/a_{in\_mouth} \tag{11}$$

The opening distance of mouth contour is also a key part of expression features. We denote the opening distance of the outer and inner mouth contour as F11 and F12, respectively, and calculated them as

$$F11 = 2b_{out\_mouth} \tag{12}$$

$$F12 = 2b_{in\_mouth} \tag{13}$$

The feature value of expression image as well as its neutral expression image will be extracted separately. Then, the calculation of difference is performed to minimize the influence of individual differences of expression features.

4.2. Feature optimization

Different features have different levels of impacts on decisions of facial expressions. We allocate a weight to each feature to enhance the influence of more effective attributes within the dataset during decision-making. Features with higher weights wield a more substantial impact on the decision-making process. We leverage the concept of feature selection, akin to a decision tree, where we treat each extracted feature as a potential decision and the individual features as decision options. Subsequently, the potential outcomes within the dataset, i.e., the expressive results requiring classification, can be characterized using the notion of information entropy. The calculation is defined as

$$H(D) = -\sum_{K=1}^{K} \frac{|C_K|}{|D|} \log \frac{|C_K|}{|D|} \tag{14}$$

Here, K represents the number of tags of image classification in data set, D represents the number of images and C_K is the number of images of a specific tag. We quantify the information contained in all images into information entropy H(D). According to equation (14), the less frequently a label appears within the dataset, the greater the amount of information a label holds.

When one feature is selected to partition the entire data set, the information entropy contained within the data will adjust accordingly. The disparity between the information entropy before and after this feature-based division serves as a measure of the information gain. A higher information gain indicates that the dataset's "purity" increases more rapidly when the feature set is employed for data partitioning. The information gain is calculated as

$$H(D, A) = H(D) - H(D|A) \tag{15}$$

Here, H(D|A) represents the information entropy resulting from the partitioning of the dataset by feature A. Similar to equation (14), we can derive it using the following equations:

$$H(D|A) = \sum_{V=1}^{V} \frac{|D_V|}{|D|} H(D_V) \tag{16}$$

$$H(D_V) = -\sum_{K=1}^{K} \frac{|C_{VK}|}{|D_V|} \log \frac{|C_{VK}|}{|D_V|} \tag{17}$$

where V represents various values of the features, and $H(D_V)$ represents the information entropy of the feature values at different levels. To calculate $H(D_V)$, it's necessary to determine the count of images with different labels corresponding to different values of feature A, denoted as C_{VK}. This enables the calculation of the information entropy of the dataset after feature division, thereby yielding the information gain. It's worth noting that the method for calculating the information gain has a limitation. When there are numerous feature values, i.e., a large number of V, regardless of wshat specific value V takes, the value of C_{VK} will invariably be smaller, resulting in $H(D_V)$ consistently being a larger value. Consequently, the computed information gain may not accurately reflect the quality of the feature values. To address this issue, we contemplate employing information gain ratios to characterize feature values.

In contrast to the information gain, the information gain rate incorporates a penalty factor. This penalty factor is inversely proportional to the feature value. It decreases as the feature value increases and vice versa. The information gain rate is derived by multiplying the information gain by this penalty factor, which effectively mitigates the issue of imprecise description stemming from a high number of feature values. The specific definitions of these penalty parameters can be written as

$$P_{parameter} = \frac{1}{H_A(D)} = \frac{1}{-\sum_{i=1}^{n} \frac{|D_i|}{|D|} \log_2 \frac{|D_i|}{|D|}} \tag{18}$$

Here, $H_A(D)$ represents the entropy of random variable when the data set D is partitioned by feature A. $|Di|$ denotes the count of occurrences where feature A within the dataset is D_i. In other words, the penalty parameter is essentially the reciprocal of the information entropy determined by partitioning dataset D using the feature as a random variable and subsequently partitioning dataset D based on this feature.

The information gain rate will serve as a metric to convey the extent of influence that features exert on the recognition of expressions. Features boasting a high information gain rate will wield more substantial influence on expressions, whereas those with a low information gain rate will exert comparatively modest influence. The information gain rate associated with each feature is subject to normalization, and the resultant value represents the feature weight within dataset D.

5. Weakly supervised clustering method and re-annotation

The weakly supervised clustering method (WSC) employs weakly supervised spectral clustering to solve an embedding space in the feature space. This embedding space maintains consistency in visual similarity and weak annotation (inaccurate supervision), and reduces the dimension of the feature space. Different from traditional methods, WSC takes into account the credibility of weak annotations and seeks a proper balance between the visual similarity and weak annotation [28]. The embedding space is solved as

$$\begin{cases} \min\limits_{W \in R^{N \times K}} f(W,L) + \dfrac{\alpha}{|\delta|} \varphi(W,\delta) \\ \text{s.t.} W^T W = I_K \end{cases} \tag{19}$$

Here, W is the embedding space to be solved, N is the number objects in the data, and K is the dimension of the embedding space. L is a Laplacian matrix calculated from the initial feature based on Euclidean distance and then calculated from the distance matrix. f(W,L) is a solution obtained from the minimum of $\mathrm{Tr}\left(W^T L W\right)$ under the condition that $W^T W$ is the identity matrix and equation (19) is satisfied. $|\delta|$ denotes the number of weak annotation clusters, and $\varphi(W,\delta)$ is used as a regularizer to encourage images with similar weak annotations to approach in the learning embedding space. $\alpha \gg 0$ is a parameter used to balance visual similarity and weak annotation. A larger α means a stronger of weak annotation while a smaller α leads to a stronger visual consistency. The process of deriving equation (19) can be found in 0.

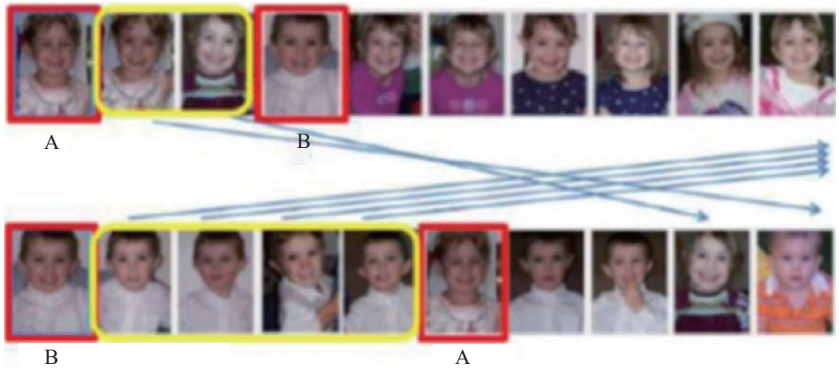
After obtaining embedding space, the original weakly annotated data needs to be re-annotated according to the embedding space. We first construct a k-nearest neighbor matrix based on the embedding space W and the rank-order distance between images [29]. Then, a hierarchical clustering method is used to perform clustering on the embedding space, and the data in the same cluster is modified based on the principle of the labeling minority. Our experiments show that this method can improve the accuracy of weak annotation labels.

In the process of label re-annotation, the number of clusters produced by hierarchical clustering has great influence on the result. So, we use modularity [30] to describe the quality of clustering. The modularity is defined as

$$Q = \frac{1}{2m}\sum_{c=1}^{n}\left[2lc - \frac{dc^2}{2m}\right] \tag{20}$$

where m is the total number of connected edges in the original neighboring matrix, c is the cluster number, lc is the number of edges included in cluster c, and dc is the number of nodes included in cluster c. Q is the modularity with a value between [−0.5, 1]. Our results show a value between 0.3 and 0.7 to calculate a proper Q that could lead to good performance.

After performing hierarchical classification using order distances, the objective of order clustering for the original sample data is accomplished. Subsequently, based on annotations within the classification results, the label data undergoes reannotation following the 'minority follows majority' voting principle, resulting in the final sample classification and improved annotations. Following the utilization of the WSE algorithm that helps to acquire the embedding space for weakly annotated data, a necessary step involves reannotating the data. This process comprises two fundamental steps: 1) utilize order clustering for embedding space classification, and 2) enhance annotations within each classification. Implementing order clustering algorithms requires constructing an undirected graph within the embedding space utilizing the order distance. This principle underscores that samples from the same classification often exhibit significant similarities, whereas samples from different classifications tend to have substantial differences. Figure 6 illustrates the principle of order distance. In Figure 6(1), A and B originate from different classifications (entities). Despite their small absolute distance, their top-level neighbors exhibit significant differences. In contrast, in Figure 6(2), A and B belong to the same category, showcasing notably similar top-level neighbors.



(1) A's top-level neighbors are different from B's

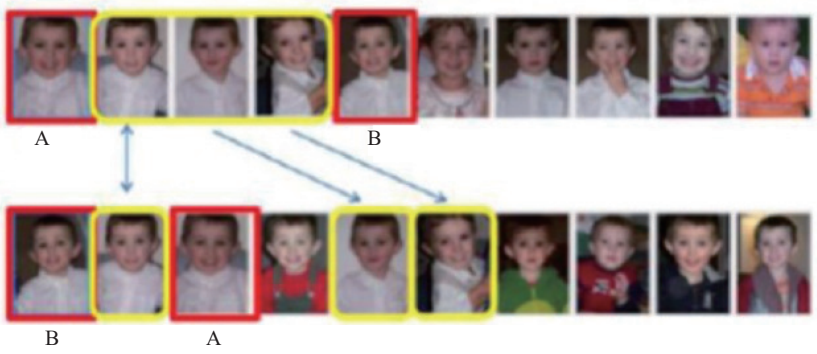


Figure 6. Schematic diagram of order distance principle.

6. Experiments

In this section, we will first describe our experimental data sets. Then, the performance of the proposed method is accessed and compared with related methods. Finally, we analyze the advantages of the proposed method.

6.1. Data sets

The experiments are conducted on CK+ [2], JAFFE [31] and Oulu-CASIA [32] datasets. The CK+ dataset includes 123 subjects with 593 image sequences, among which 327 sequences are labeled with expressions of neutral, anger, contempt, aversion, fear, joy, sadness and surprise. The JAFFE dataset contains 213 images from 10 Japanese female students. Each person has seven kinds of expressions, including angry, strange, fear, happy, sad, surprise and neutral. The Oulu-CASIA facial expression dataset contains videos of six typical expressions (happiness, sadness,

surprise, anger, fear and disgust) from 80 subjects captured with two imaging systems, NIR (Near Infrared) and VIS (Visible light), under three different illumination conditions. In this paper, the video screenshots taken under normal indoor illumination are used.

Since not every subject has a neutral label image and the image sequence is composed of a series of changing expression images, we choose the first or second image of the changing expression as the target neutral expression image and the last 1 to 3 images as the corresponding expression images. The formed data pair will be fed to cGAN for training. Take the CK+ dataset as an example. For this data set, 3589 pairs of image data are used to train cGAN, including 1668 positive expression images (mainly labeled as happy or surprise) and 1921 negative expression images (mainly labeled as disgust, fear, and sadness). Following this, we choose 1367 previously unused images to constitute the test set and apply the cGAN training model to create neutral expression images from expressive ones, encompassing 679 positive expression images and 688 negative expression images. Ultimately, we handpick 1301high-quality images for feature extraction. The detailed image numbers used for training and testing in different datasets are shown in Table 1.

Table 1 Number of images used for training and testing in different datasets

		Training	Test	selected
CK+	Positive	1668	679	664
	Negative	1921	688	637
	All	3589	1367	1301
JAFFE	Positive	93	35	30
	Negative	90	32	29
	All	183	67	59
Oulu-CASIA	Positive	801	263	248
	Negative	672	184	171
	All	1473	447	419

6.2. Face key point annotation

The neutral expression images and initial expression images obtained are used to perform facial key point annotation and geometric feature value extraction according to the method described in Section 4.

As illustrated in Figure 7, we annotated facial key points using AAM for both the initial expression image and generated neutral expression image. The top row represents the initial expression image, while the bottom row portrays the neutral expression image. The first two columns display positive expression images, while the third and fourth columns exhibit negative expression images. It is noteworthy that the annotated facial key points effectively convey expression information, despite some texture gaps and other detail differences between the generated neutral expression image and the actual image

Figure 7. Facial key point annotation.

6.3. Results of re-annotation and weakly supervised clustering

We then proceed by extracting the image's geometric features and subsequently optimizing them. The resulting feature matrix is then employed to resolve the embedding space and perform re-annotation in accordance with the methodology outlined in Section 5. Notably, the initial label data undergoes random inversion at a rate of 0.3, implying that the original correct labels generate weak annotations with accuracy of 0.7. The outcomes pertaining to the CK+ dataset are shown in Figure 8.

The accuracy between the re-annotated labels and the actual labels is represented by the blue line, while the red line reflects the variation in modularity levels. The *x*-axis corresponds to the number of clusters employed in hierarchical clustering during re-annotation. When the number of clusters reaches 2^7, the modularity attains its peak at 0.68, signifying excellent clustering performance within the 0.3 to 0.7 range. At this juncture, the accuracy rate (ACC)

reaches 83.7%, which has the mere 1.6% deviation from the highest accuracy rate of 85.3%. This relationship can be attributed to the fact that elevated modularity coincides with heightened accuracy. Consequently, modularity can serve as a reliable criterion for determining the optimal number of clusters in hierarchical clustering. Specifically, one should select the number of clusters that yields the highest modularity.

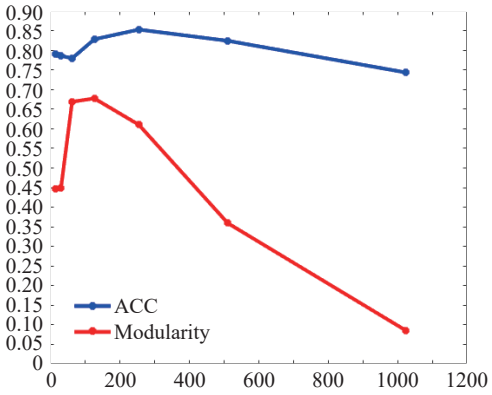


Figure 8. Re-annotate results of CK+ dataset.

By employing WSC approach to solve the embedding space, the influence of weak annotations can be adjusted by changing the parameter α. According to the results in Section 5, smaller α values tend to cluster visually similar images together due to their proximity in the feature space. Conversely, a larger α biases the algorithm more heavily towards the weak annotations, resulting in a cleaner clustering outcome. An optimal α value falls into where the embedding space effectively retains images that close in the feature space and similar in annotations.

Figure 9 visually portrays the distribution of the embedding space when α equals an intermediate value. In this representation, red dots represent images with positive facial expressions as their true annotations, while blue dots represent images with negative facial expressions as their genuine annotations. At this juncture, the embedding space accurately captures the facial expression information, which greatly enhances accuracy during the re-annotation stage.

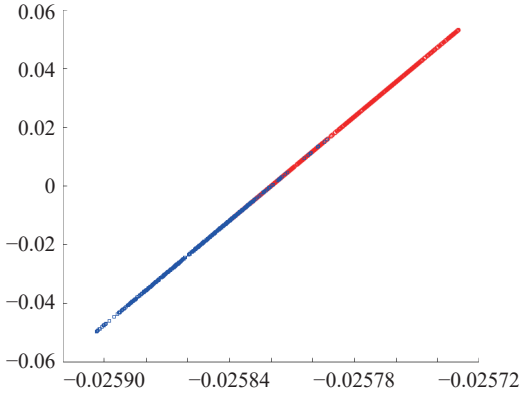


Figure 9. The distribution of embedding space (CK+) with $\alpha = 2^{(-4)}$.

Additionally, as shown in Figure 10, we explored the impact of using varying proportions of weakly annotated data on accuracy within the CK+ dataset. It becomes evident that when the proportion of correctly annotated data falls below 0.7, the method's accuracy suffers substantial decline. However, when the proportion of correctly annotated data exceeds 0.7, the improvement in accuracy become less pronounced. As a result, in subsequent experiments, we will maintain an annotation data ratio to 0.7.

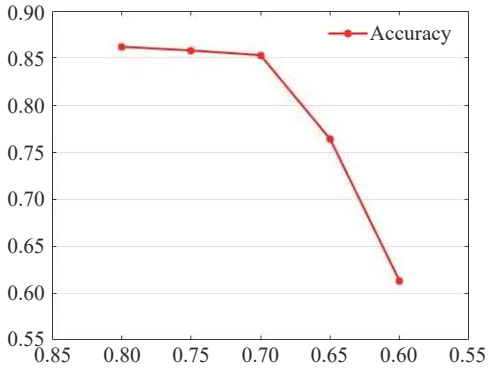


Figure 10. The change of accuracy under different amounts of weak annotated data. The ordinate represents accuracy rate, and the abscissa represents proportion of correct data. For example, if the abscissa is 0.7, the proportion of data using correct annotation is 0.7, and the proportion of data using weak annotation (wrong annotation) is 0.3.

6.4. Facial expression recognition based on feature optimization methods

According to the feature optimization theory in Section 4, we selected similarity evaluation indicators and conducted facial expression recognition experiments on three feature optimization methods: the selection threshold, random discard rate, and Pearson correlation difference. The experimental results are shown in Table 2.

Table 2 Results of feature optimization verification experiments

Selected threshold	Random drop rate	Average Pearson correlation difference	Recognition accuracy
0.175	0.7	0.037196 0.046481	0.841
0.150	0.7	0.059815 0.060255	0.862
0.125	0.7	0.078500 0.081274	0.913
0.100	0.7	0.082734 0.076061	0.883
0.150	0.5	0.026140 0.019860	0.804
0.125	0.5	0.037192 0.038695	0.83

The average Pearson correlation difference includes the correlation difference between positive and negative samples (the first row) and the correlation difference between similar samples (the second row). When the selection threshold is 0.125 and the random discard rate is 0.7, the weakly supervised clustering algorithm has the highest accuracy in feature space of 0.913. At this time, according to the theory in Section 4, this is also the optimal selection threshold and random discard rate.

Comparing the experimental results with a selection threshold of 0.125, a random discard rate of 0.7, a selection threshold of 0.100, and a random discard rate of 0.7, the average Pearson correlation between positive and negative samples is better in the second case, while the average Pearson correlation between similar samples is better in the first case. Overall, the feature space optimization effect of the two cases is similar, and from the perspective of recognition accuracy, the two cases are also very similar. However, in terms of feature space dimension, the first scenario has a smaller feature dimension and is easier to classify when information needs are satisfied.

Comparing the experimental results with a selection threshold of 0.150, a random discard rate of 0.7, a selection threshold of 0.125, and a random discard rate of 0.7, according to the discussion in section. 4, the optimal solution was selected for optimizing the parameters, but the second case was better in terms of the recognition rate.

6.5. Comparison with related methods

The CK+ dataset will be used to validate the effectiveness of our feature extraction approach. Table 3 presents the re-annotation accuracy achieved under various cluster numbers of weakly supervised clustering, utilizing different feature extraction methods. The first row corresponds to the usage of PCA for extracting principal component features, the second row involves FaceNet, and the third row relies solely on geometric features without any optimization, meaning that the neutral expression images generated by cGAN are not incorporated. Given that the CK+ dataset comprises expression images captured under varying lighting conditions, the variations in brightness are quite pronounced. Consequently, PCA-based facial feature extraction tends to categorize images with similar brightness

levels more effectively by the weakly supervised clustering algorithm, albeit at the expense of accurately reflecting emotions. On the other hand, FaceNet-based facial feature extraction predominantly emphasizes identity information associated with the face, making it less proficient at expressing facial expression details. Our feature extraction approach, in conjunction with the weakly supervised clustering algorithm, exhibits notable adaptability. The optimization method we employ mitigates the impact of non-essential features, significantly enhancing the accuracy of re-annotated labels.

Table 3 Accuracy of different feature extraction methods

	2^4	2^5	2^6	2^7	2^8	2^9
PCA	0.56	0.56	0.59	0.66	0.68	0.65
Facenet	0.55	0.56	0.57	0.57	0.66	0.67
Notopt-imized	0.71	0.73	0.78	0.78	0.77	0.75
Ours	**0.79**	**0.78**	**0.76**	**0.83**	**0.85**	**0.82**

Table 4 presents a comparative analysis of our method with related methods on various datasets, where "UD" denotes the utilization of unannotated data, and "PN" indicates the pruning of noisy annotations. Our initial experiments are conducted on the CK+ dataset. The results demonstrate that our method achieves recognition rates exceeding 85% in both scenarios and surpassing the performance of related methods. Notably, both the FCM (Fuzzy C-means clustering) and the FIS (Fuzzy Inference System) directly employ facial landmark coordinates without transforming them into geometric features. While unannotated data can be utilized in semi-supervised learning, it is essential to acknowledge that incorrect and noisy data within the dataset cannot be rectified. Moreover, our method consistently exhibits superior accuracy compared to these two techniques. SVM is widely recognized as an effective algorithm for facial expression recognition and has a high recognition rate. However, it necessitates the use of annotated data, and it is unable to rectify noisy data. When employed with weakly annotated data, SVM's recognition accuracy drops significantly from 93.1% to 68.7%. Other SVM based algorithms, such as LapSVM [33] and TSVM [34], can employ part of annotated data and part of unannotated data for semi-supervised learning. However, they cannot prune the noise data either.

Table 4 Comparison results of our method and related facial expression recognition methods in different datasets, including CK+, JAFFE and Oulu-CASIA

	Method	ACC	UD	PN		Method	ACC	UD	PN
CK+	FCM [4]	80.7%	√	×		Ours	83.1%	√	√
	FIS [4]	72.0%	√	×		CNN [27]	80.78%	×	×
	SVM [32]	93.1%	×	×		LBP-TOP [36]	68.1%	×	×
	Ours	85.3%	√	√		STM-Exp [37]	74.6%	√	×
JAFFE	FERF [40]	82.0%	√	×	Oulu-CASIA	Atlases [35]	75.5%	√	×
	ISR [39]	83.5%	×	×		PPDN [38]	84.6%	×	×
	CNN [39]	82.54%	×	×		Ours	81.4%	√	√

Experiments were carried out on JAFFE dataset. On this data set, our method achieved accuracy of 83.1%, which is comparable with related methods. The fuzzy emotion recognition framework (FERF) employs a fuzzy approach to analyze the facial components for emotion type determination. Specifically, this method extracts semantic features by transforming face key points into facial components using geometric model analysis. These parameters are then used to establish facial component rules, which are employed to classify facial expressions. The accuracy of FERF is slightly lower than our method and, further, it lacks the capability to handle noisy annotations. The improved sparse representation method (ISR) leverages sparse representation to achieve high recognition accuracy with limited training samples, with accuracy improved as the quantity of training samples increased. However, ISR is not suitable for handling unannotated data and noisy annotations, despite that its accuracy reaches 83.5%.

Finally, we compared the performance of various methods on Oulu-CASIA dataset. The Oulu-CASIA dataset contains data captured under three different illumination conditions using two types of cameras. In our experiments, the data captured under strong illumination conditions with the VIS are utilized. Similar to CK+ dataset, each video sequence starts from a neutral facial expression and ends with a peak facial expression. Hence, we handled the Oulu-CASIA dataset in a manner analogous to the CK+ dataset. The results show that our method is able to achieve higher accuracy than most of the related methods, including longitudinal atlases construction based method (Atlases [35]) and manually constructed feature-based methods (LBP-TOP [36] and STM explet [37]). PPDN [38], which is based on CNNs, does demonstrate superior performance to our method. However, it is not able to handle unannotated data and noisy data. It is worth noting that both PPDN [38] and our method employ static images for facial expression recognition, while others incorporate temporal information from video sequences.

7. Conclusions

In this work, we propose a method to learn facial expressions from data with weak annotations. The proposed method employs the residual technique to derive features by employing cGAN and geometric modeling. This method exhibits excellent compatibility with weakly supervised clustering algorithms. Unlike conventional facial expression recognition methods, our proposed method can effectively utilize weak annotations, enhancing recognition accuracy by incorporating neutral facial expression images to mitigate inter-individual variations. Further, an optimization method has also been designed to enhance the accuracy of re-annotation. The performance of the proposed method has been evaluated on various datasets, including CK+, JAFFE and Oulu-CASIA, and compared with related methods.

The evaluation results show that our method outperforms related methods and is promising future for practical applications. Our proposed method can effectively handle weakly annotated data and rectify noisy data while maintaining robust recognition performance. However, it should be noted that during the process of obtaining a neutral image using cGAN, certain discrepancies between the obtained neutral image and target image may arise due to network constraints and lower image resolution. Further, in our method, extraction of geometric features operates at the pixel level, potentially resulting in the loss of some facial expression information. Despite efforts to enhance the importance of key features through optimization, there remains a discernible gap in recognition accuracy compared to certain facial expression recognition technologies.

For the future work, one potential direction is to enhance the performance of cGAN by acquiring high-quality neutral expression images, thereby closing the gap between neutral image and target image. Further, experiments on multi-label weakly annotated data and applications to facial expression recognition of video streams could be carried out to verify the performance of the proposed method.

Author Contributions: **Ze Chen:** Methodology, Investigation, Software, **Writing Lu Zhang:** Methodology, Data collection and analysis **Jiaming Tang:** Software, **Writing Jiafa Mao:** Investigation, Validation, Writing – Editing and Reviewing **Weiguo Sheng:** Conceptualization, Writing – Editing and Reviewing, Supervision.

Funding: This work is supported by the Scientific Research Fund of Zhejiang Provincial Education Department (No. Y202147393), the "Pioneer" and "Leading Goose" R&D Program of Zhejiang Province (No. 2023C01022), the National Natural Science Foundation of China (No. 62176237), the Zhejiang Province Natural Science Foundation of China (No.LY20F020022) and the National Key R&D Program of China (No. 2018YFB0204003).

Data Availability Statement: The data is available upon request.

Conflicts of Interest: The authors have no relevant financial or non-financial interests to disclose.

Acknowledgements: This work is supported in part by the "Pioneer" and "Leading Goose" R&D Program of Zhejiang Province (No. 2023C01022), the Scientific Research Fund of Zhejiang Provincial Education Department (No. Y202147393), the National Natural Science Foundation of China (No. 62176237), the Zhejiang Province Natural Science Foundation of China (No. LY20F020022) and the National Key R&D Program of China (No. 2018YFB0204003).

References

1. Chu, W.S.; De la Torre, F.; Cohn, J.F. Selective transfer machine for personalized facial action unit detection. In *2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013*; IEEE: New York, 2013; pp. 3515–3572. doi:10.1109/CVPR.2013.451
2. Lucey, P.; Cohn, J.F.; Kanade, T.; *et al*. The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition – Workshops, San Francisco, CA, USA, 13–18 June 2010*; IEEE: New York, 2010; pp. 94–101. doi:10.1109/CVPRW.2010.5543262
3. Valstar, M.F.; Almaev, T.; Girard, J.M.; *et al*. FERA 2015 - second facial expression recognition and analysis challenge. In *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Ljubljana, Slovenia, 4–8 May 2015*; IEEE: New York, 2015; pp. 1–8. doi:10.1109/FG.2015.7284874
4. McDuff, D.; El Kaliouby, R.; Senechal, T.; *et al*. Affectiva-MIT facial expression dataset (AM-FED): Naturalistic and spontaneous facial expressions collected "in-the-wild". In *2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Portland, OR, USA, 23–28 June 2013*; IEEE: New York, 2013; pp. 881–888. doi:10.1109/CVPRW.2013.130
5. Xia, Y.F.; Yu, H.; Wang, X.; *et al*. Relation-aware facial expression recognition. IEEE Trans. Cogn. Dev. Syst., **2022**, *14*: 1143−1154. doi: 10.1109/TCDS.2021.3100131
6. Mirza, M.; Osindero, S. Conditional generative adversarial nets. arXiv preprint arXiv: 1411.1784, 2014.
7. Nigam, S.; Singh, R.; Misra, A.K. Efficient facial expression recognition using a histogram of oriented gradients in wavelet domain.

Multimed. Tools Appl., **2018**, *77*: 28725−28747. doi: 10.1007/s11042-018-6040-3
8. Ge, H.L.; Zhu, Z.Y.; Dai, Y.W.; *et al*. Facial expression recognition based on deep learning. Comput. Methods Programs Biomed., **2022**, *215*: 106621. doi: 10.1016/j.cmpb.2022.106621
9. Zhang, H.F.; Su, W.; Yu, J.; *et al*. Identity–expression dual branch network for facial expression recognition. IEEE Trans. Cogn. Dev. Syst., **2021**, *13*: 898−911. doi: 10.1109/TCDS.2020.3034807
10. Sun, W.Y.; Zhao, H.T.; Jin, Z. A visual attention based ROI detection method for facial expression recognition. Neurocomputing, **2018**, *296*: 12−22. doi: 10.1016/j.neucom.2018.03.034
11. Michael Revina, I.; Sam Emmanuel, W.R. Facial expression recognition via modified GAD features with PSO-KNN. In *2018 International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 13–14 December 2018*; IEEE: New York, 2018; pp. 145–149. doi:10.1109/ICSSIT.2018.8748697
12. Kim, Y.; Yoo, B.; Kwak, Y.; *et al*. Deep generative-contrastive networks for facial expression recognition. arXiv preprint arXiv: 1703.07140, 2017.
13. Zafeiriou, S.; Petrou, M. Sparse representations for facial expressions recognition via l1 optimization. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition – Workshops, San Francisco, CA, USA, 13–18 June 2010*; IEEE: New York, 2010; pp. 32–39. doi:10.1109/CVPRW.2010.5543148
14. Bazzo, J.J.; Lamar, M.V. Recognizing facial actions using Gabor wavelets with neutral face average difference. In *Proceedings of Sixth IEEE International Conference on Automatic Face and Gesture Recognition, Seoul, Korea (South), 19 May 2004*; IEEE: New York, 2004; pp. 505–510. doi:10.1109/AFGR.2004.1301583
15. Rawal, N.; Stock-Homburg, R.M. Facial emotion expressions in human-robot interaction: A survey. Int. J. Soc. Robot., **2022**, *14*: 1583−1604. doi: 10.1007/S12369-022-00867-0
16. Zhang, X.; Zhang, F.F.; Xu, C.S. Joint expression synthesis and representation learning for facial expression recognition. IEEE Trans. Circuits Syst. Video Technol., **2022**, *32*: 1681−1695. doi: 10.1109/TCSVT.2021.3056098
17. Yaermaimaiti, Y.; Kari, T.; Zhuang, G.H. Research on facial expression recognition based on an improved fusion algorithm. Nonlinear Eng., **2022**, *11*: 112−122. doi: 10.1515/nleng-2022-0015
18. Bao, H.; Ma, T. Feature extraction and facial expression recognition based on Bezier curve. In *2014 IEEE International Conference on Computer and Information Technology*, *Xi'an*, *China*, *11–13 September 2014*; IEEE: New York, 2014; pp. 884–887. doi:10.1109/CIT.2014.140
19. Ghahari, A.; Fatmehsari, Y.R.; Zoroofi, R.A. A novel clustering-based feature extraction method for an automatic facial expression analysis system. In *Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Kyoto, Japan, 12–14 September 2009*; IEEE: New York, 2019; pp. 1314–1317. doi:10.1109/IIH-MSP.2009.38
20. Yang, H.Y.; Ciftci, U.; Yin, L.J. Facial expression recognition by de-expression residue learning. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018*; IEEE: New York, 2018; pp. 2168–2177. doi:10.1109/CVPR.2018.00231
21. Bilen, H.; Pedersoli, M.; Tuytelaars, T. Weakly supervised object detection with convex clustering. In *2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015*; IEEE: New York, 2015; pp. 1081–1089. doi:10.1109/CVPR.2015.7298711
22. Prest, A.; Schmid, C.; Ferrari, V. Weakly supervised learning of interactions between humans and objects. IEEE Trans. Pattern Anal. Mach. Intell., **2012**, *34*: 601−614. doi: 10.1109/TPAMI.2011.158
23. Bo, Y.; Fan, J.J.; Zhuang, J. Visual analysis of facial expression recognition research based on knowledge graph. In *Proceedings of the 4th International Conference on Machine Learning for Cyber Security, Guangzhou, China, 2–4 December 2022*; Springer: Berlin/Heidelberg, 2022; pp. 350–357. doi:10.1007/978-3-031-20102-8_27
24. Luo, Y.; Wu, J.X.; Zhang, Z.H.; *et al*. Design of facial expression recognition algorithm based on CNN model. In *Proceedings of the 3rd IEEE International Conference on Power, Electronics and Computer Applications, Shenyang, China, 29–31 January 2023*; IEEE: New York, 2023; pp. 580–583. doi:10.1109/ICPECA56706.2023.10075779
25. Shiomi, T.; Nomiya, H.; Hochin, T. Facial expression intensity estimation considering change characteristic of facial feature values for each facial expression. In *Proceedings of the 23rd ACIS International Summer Virtual Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, Kyoto City, Japan, 4–7 July 2022*; IEEE: New York, 2022; pp. 15–21. doi:10.1109/SNPD-Summer57817.2022.00012
26. Jin, X.F.; Liu, J.Y.; Yue, D. The research and improvement of facial expression recognition algorithm based on convolutional neural network. In *Proceedings of the 26th ACIS International Winter Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, Taiyuan, China, 5–7 July 2023*; IEEE: New York, 2023; pp. 166–170. doi:10.1109/SNPD-Winter57765.2023.10224044
27. Wang, L.; Li, R.F.; Wang, K. A novel automatic facial expression recognition method based on AAM. J. Comput., **2014**, *9*: 608−617.
28. Zhao, K.L.; Chu, W.S.; Martinez, A.M. Learning facial action units from web images with scalable weakly supervised clustering. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018*; IEEE: New York, 2018; pp. 2090–2099. doi:10.1109/CVPR.2018.00223
29. Zhu, C.H.; Wen, F.; Sun, J. A rank-order distance based clustering algorithm for face tagging. In *CVPR 2011, Colorado Springs, CO, USA, 20–25 June 2011*; IEEE: New York, 2011; pp. 481–488. doi:10.1109/CVPR.2011.5995680
30. Newman, M.E.J. Modularity and community structure in networks. Proc. Natl. Acad. Sci. USA, **2006**, *103*: 8577−8582. doi: 10.1073/pnas.0601602103
31. Lyons, M.; Akamatsu, S.; Kamachi, M.; *et al*. Coding facial expressions with Gabor wavelets. In *Third IEEE International Conference on Automatic Face and Gesture Recognition, Nara, Japan, 14–16 April 1998*; IEEE: New York, 1998; pp. 200–205. doi:10.1109/AFGR.1998.670949
32. Zhao, G.Y.; Huang, X.H.; Taini, M.; *et al*. Facial expression recognition from near-infrared videos. Image Vision Comput., **2011**, *29*: 607−619. doi: 10.1016/j.imavis.2011.07.002
33. Melacci, S.; Belkin, M. Laplacian support vector machines trained in the primal. J. Mach. Learn. Res., **2011**, *12*: 1149−1184.
34. Joachims, T. Transductive inference for text classification using support vector machines. In *Proceedings of the Sixteenth International Conference on Machine Learning, Bled, Slovenia, 27–30 June 1999*; Morgan Kaufmann Publishers Inc.: San Francisco, 1999; pp. 200–209.

35. Guo, Y.M.; Zhao, G.Y.; Pietikäinen, M. Dynamic facial expression recognition using longitudinal facial expression atlases. In *12th European Conference on Computer Vision, Florence, Italy, 7–13 October 2012*; Springer: Berlin/Heidelberg, 2012; pp. 631–644. doi:10.1007/978-3-642-33709-3_45
36. Zhao, G.Y.; Pietikainen, M. Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Trans. Pattern Anal. Mach. Intell., **2007**, *29*: 915−928. doi: 10.1109/TPAMI.2007.1110
37. Liu, M.Y.; Shan, S.G; Wang, R.P.; *et al*. Learning expressionlets on spatio-temporal manifold for dynamic facial expression recognition. In *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014*; IEEE: New York, 2014; pp. 1749–1756. doi:10.1109/CVPR.2014.226
38. Zhao, X.Y.; Liang, X.D.; Liu, L.Q.; *et al*. Peak-piloted deep network for facial expression recognition. In *14th European Conference on Computer Vision, The Netherlands, 11–14 October 2016*; Springer: Berlin/Heidelberg, 2016; pp. 425–442. doi:10.1007/978-3-319-46475-6_27
39. Liu, S.G.; Li, L.J.; Peng, Y.L.; *et al*. Improved sparse representation method for image classification. IET Comput. Vision, **2017**, *11*: 319−330. doi: 10.1049/iet-cvi.2016.0186
40. Liliana, D.Y.; Basaruddin, T. The fuzzy emotion recognition framework using semantic-linguistic facial features. In *2019 IEEE R10 Humanitarian Technology Conference*, *Depok*, *West Java*, *Indonesia*, *12–14 November 2019*; IEEE: New York, 2019; pp. 263–268. doi: 10.1109/R10-HTC47129.2019.9042442

Citation: Chen, Z.; Zhang, L.; Tang, J.; *et al*. Conditional Generative Adversarial Net based Feature Extraction along with Scalable Weakly Supervised Clustering for Facial Expression Classification. *International Journal of Network Dynamics and Intelligence*. 2024, 3(4), 100024. doi: 10.53941/ijndi.2024.100024