IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received November 16, 2019, accepted December 4, 2019, date of publication December 20, 2019, date of current version January 6, 2020.

Digital Object Identifier 10.1109/ACCESS.2019.2960769

A Deep Convolutional Neural Network With Fuzzy Rough Sets for FER

XIANGJIAN CHEN[1], DI LI[2], PINGXIN WANG[3], AND XIBEI YANG[1]
[1]School of Computer, Jiangsu University of Science and Technology, Zhenjiang 212003, China
[2]China Shipbuilding Industry Corporation 723, Yangzhou 212005, China
[3]School of Science, Jiangsu University of Science and Technology, Zhenjiang 212003, China

Corresponding author: Xiangjian Chen (dongnan_cxj831209@163.com)

This work was supported by the National Science Foundation of China under Grant 61502211, Grant 61572242, and Grant 61702234.

ABSTRACT Existing facial emotion recognition methods do not have high accuracy and are not sufficient practical in real-time applications. We introduce type 2 fuzzy rough sets to develop a Type 2 Fuzzy Rough Convolutional Neural Network, as type 2 fuzzy rough sets form a suitable mathematical tool to characterize uncertainty of classifification. Based on the type 2 fuzzy rough sets theory, we construct an optimization objective for training CNNs by minimizing fuzzy classification uncertainty, and present the defifinition and optimization of type 2 fuzzy rough loss, which can be achieved by better performance. This method could reduce the uncertainty in terms of vagueness and indiscernibility by using type 2 fuzzy rough sets theory and specifically removing noise samples by using CNN from raw data. And finally, compared the proposed method with other feature extraction and learning techniques based on Algorithm Adaption k-Nearest-Neighbors. Experimental results demonstrate that type 2 fuzzy rough sets convolutional neural network could achieve better performances comparing with other methods.

INDEX TERMS Convolutional neural network, Type 2 fuzzy rough sets, algorithm adaption k-nearest-neighbors.

I. INTRODUCTION

The expression is a direct way of transmission of human inner emotions and plays a subtle role in interpersonal communication. The joys and sorrows can often be revealed through a subtle expression. For humans, emotion recognition maybe the basic ability of the brain. However, it is not easy for artificial intelligence systems with the corresponding emotion recognition ability. Since the artificial intelligence program does not work according to the mode of the human brain, if you want to recognize the emotions corresponding to various expressions, many supervised training data for training the model is needed. In applications such as human-computer interaction, automatic facial expression recognition technology is widely used to enhance the user experience of the system, which is an effective way for machines to understand the relationship between human expression and internal emotions.

In recent years, emotion-analysis systems based on facial expressions have been successfully applied in many application scenarios, such as human-computer interaction systems and games. The machine could interpret user emotions through emoticon recognition, which makes the algorithm more intelligent and Humanized.

Single-tag learning or multi-tag learning paradiagms have been used in facial expression recognition systems early, which is based on the characteristics of manual constructs such as Local Binary Patterns (LBP) [1], Support Vector Machines (SVM) [2] and other classifiers. A linear Softmax regression mode [3] based on emotion distribution learning can be used to predict the distribution of basic emotions for considering the complexity of expression. Compared to single-tag and multi-tag learning problems, the label space of this type of task is expanded from a limited number of points to a unit sphere composed of basic emotions, which can represent more abundant and complex expressions. Different from other expression classification methods, the fuzzy expression classification task uses fuzzy multiple labels to describe the expression, which has more important research value, because it is more closer to the actual situation.

Although the fuzzy expression classification problem can be realized by modifying the traditional k-nearest neighbor algorithm [4], but the performance of the algorithm depends on the ability of feature to distinguish expressions. Therefore,

The associate editor coordinating the review of this manuscript and approving it for publication was Khalid Aamir.

the important part to implement the fuzzy expression classification model based on K-nearest neighbor algorithm is how to select or train valid features. As an effective tool to deal with fuzzy information, fuzzy rough set theory provides a measure criterion for effectively evaluating feature performance in fuzzy classification problems, and also which is widely used in feature selection and feature weighting of classification problems. fuzzy rough set theory for feature selection and feature weighting with fuzzy expression multi-label information is more suitable for fuzzy expression classification.

On the other hand, with the introduction of the ImageNet competition 1000 picture classification tasks, the in-depth learning algorithm based on end-to-end training has been greatly developed, and has demonstrated excellent performance on many computer visual tasks. However, most of the end-to-end training Deep Network research work is based on single-tag classification tasks and has not been effectively applied to fuzzy expression classification problems.

Early facial expression recognition systems takes facial expression recognition as a classification problem, discrete emotional labels are used as model outputs, which includes binary expression classification and emotional intensity. Based on the idea of label distribution learning, Zhou *et al.* [5] constructed an expression data set containing emotional distribution labels and proposed emotional distribution learning to identify the distribution strength of various emotions in other people's facial expression pictures, when constructing the Softmax model, the weighted Jeffrey divergence is used as a loss function to obtain an experimental result that is superior to the KL divergence loss function. Considering the limited abstract ability of the linear Softmax regression model to the features, Yang *et al.* [6] expanded the conditional probability neural network (CPNN) to a fuzzy multi-signature form. Augmented Association Network (ACPNN), which is used to predict the emotions expressed by the picture and obtain better performance than the original CPNN, although ACPNN builds a multi-layer sensor model, but which still relies on features as input, and the performance of the algorithm is limited by the abstract ability of the used features to emotions.

One of the factors for the success of deep networks is the use of a large amount of supervised data for end-to-end training and automatic learning features. Taking into account that the previous label distribution learning algorithms are all based on characteristics, Gao *et al.* [7] proposed the concept of in-depth label distribution learning and used KL divergence and squared error optimization models, which has achieved success in tasks such as age classification, face attitude estimation, and image content understanding, and there are several researches focused on the convolutional neural network for FER [8]–[11]. Fuzzy rough set theory, as a mathematical tool to effectively deal with the inconsistency of label information, has been widely discussed and applied in fuzzy classification problems. It is often used for feature selection and feature weighting of fuzzy classification problems, and also provides a good idea for the training of deep convolutional neural network in fuzzy expression classification problem [12], [13]. Taking into account that there are both discrete data and continuous data in practical applications, Wang *et al.* [14] proposed a nuclear fuzzy rough set model and used kernel functions to calculate the fuzzy T equivalence relationship between samples. And the difference between fuzzy upper approximation and fuzzy lower approximation is proposed as the evaluation index of classification inconsistency, and used for attribute reduction and feature selection. In the multi-classification problem, Hu *et al.* [15] realized the feature weighted training framework based on the fuzzy rough set theory through the sum of fuzzy degrees of affiliation of the training set samples and the weight vector of the learning characteristics, and combined with convolutional neural network for FER. In addition, the fuzzy rough set model has also been successfully applied to heterogeneous attribute reduction [16], active learning [17], [18], build a robust classification method [19], fuzzy rule extraction [20], MR image segmentation [21], integrated learning [22], and Shan *et al.* [23] proposed covering-based general multigranulation intuitionistic fuzzy rough sets and corresponding applications to multi-attribute group decision-making. In other fields, it shows excellent performance. In the fuzzy expression classification problem, the fuzzy rough set theory can be used as an optimization target for feature selection and network training. But these fuzzy logic sets used in the above references are all based on the type-1, however the type-2 fuzzy sets could handle the uncertainties more flexible for more adjustable parameters contained in the structure, which could decrease the diffilculty in uncertainty representation compared to the type-1 fuzzy sets. So this paper also applied type 2 fuzzy rough set into the proposed model.

The remainder of this paper is organized as follows: our approach is presented in Section "Framework of the Proposed Approach". Experiment results and analysis are presented in Section "Experiments and Validation." Lastly, Section "Conclusions" presents our conclusions.

II. FRAMEWORK OF THE PROPOSED METHOD

The deep convolutional neural network has been broken through the graph classification questions, and the work has solved the problem of the classification of 1,000 categories of pictures in the ImageNet data set. However, the fuzzy expression classification is different from the traditional picture multi-classification problem, in fact, there is a degree of affiliation for each emotion, not a simple genus or just "in or out" of relationship. This paper mainly studies how to construct a deep convolutional neural network to predict the degree of affiliation of expression to basic emotions. By combining fuzzy rough set theory and deep convolutional network, a two-stage fuzzy rough sets convolutional neural network model is proposed here. Firstly, a fuzzy rough loss function based on the theory of fuzzy rough set is proposed to evaluate the ability of depth feature to distinguish emotion in fuzzy expression classification problem. Secondly,

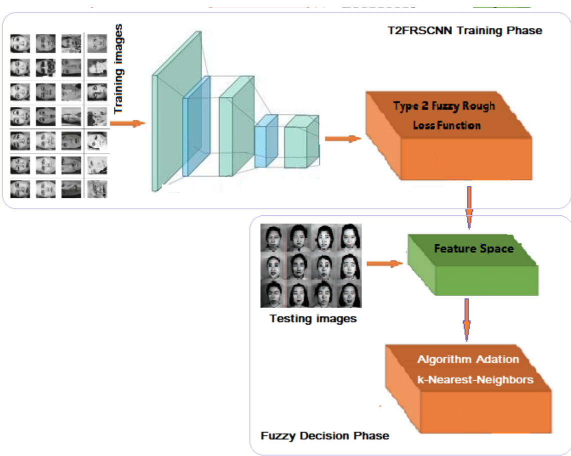

FIGURE 1. Framework of the type 2 fuzzy rough sets convolutional neural network model.

the reverse propagation processby calculating the derivative of the fuzzy rough loss function is deduced and used for pre-training in the first stage. Thirdly, Algorithm Adaption k-Nearest-Neighbor fuzzy classifiers is used for the second phase of fuzzy decision-making, and experiments on fuzzy expression classification were conducted on the Jafffe, CK+ and BU-3DFE datasets.

A. TYPE 2 FUZZY ROUGH SETS CONVOLUTIONAL NEURAL NETWORK MODEL

Fuzzy expression recognition algorithm framework based on the type 2 fuzzy rough sets convolutional neural network is proposed in Fig.1. The input of the algorithm is a face image with center clipped. Firstly the five layers of AlexNet is taken to extract the image characteristics, and then maping the feature diagram of the back and the layer to the characteristic layer by using the full connection layer.

During the type 2 fuzzy rough sets convolutional neural network training phase, the gradient descent algorithm is used to train the parameters of the convolutional layer and the full connection layer to minimize the type 2 fuzzy rough loss function(FRL). In the fuzzy decision-making stage, the type 2 fuzzy rough sets convolutional neural network was used to extract the characteristics, and a fuzzy classifier was established to predict the degree of membership of each emotion. In the process of training type 2 fuzzy rough sets convolutional neural network, the loss function uses the emotional label value and the feature vector as inputs to calculate and reduce the classification uncertainty of each type of emotion. The emotional label that enters the fuzzy rough loss function needs to be normalized to [0,1] within the interval, which is used as the degree of subordination of the image to a certain emotion. The vectors of the output need to be normalized for avoiding the convergence of the model to the meaningless trivial solutions. In the implementation of this paper, the feature vector is normalized using 2 norm. The model is optimized by gradient descent algorithm for the fuzzy rough loss function can be guided. In the fuzzy decision-making stage, the pre-trained type 2 fuzzy rough sets convolutional neural network is used as a depth feature extractor, and fuzzy expression classification based on special characteristics is constructed, which is also used for fuzzy decision-making. Therefore, in the fuzzy decision stage, the depth feature is extracted by type 2 fuzzy rough sets convolutional neural network, and the fuzzy expression recognition is realized by constructing fuzzy classifier.

B. TYPE 2 FUZZY LOSS FUNCTION

According to the theory of nuclear fuzzy rough set, fuzzy upper approximation is used to indicate the possibility of a sample belonging to a certain emotion, and fuzzy lower approximation is used to indicate the inevitability of a sample belonging to a certain emotion. In general, the classification decision of a sample has less uncertainty for the strong ability to distinguish between feature space, which means that the closer upper fuzzy approximation to the lower fuzzy approximation is the better.

For a trainable new, the characteristics of the model should have stronger classification decision-making capabilities if the uncertainty of fuzzy classification on the classification boundary can be reduced as much as possible. First, the core matrix of all sample pairs are calculated, the fuzzy upper approximation and fuzzy lower approximation of each sample are obtained, and the classification uncertainty of each sample is calculated. By adding up the classification uncertainty of all samples, the type 2 fuzzy rough loss function defined on all sample sets is obtained.

Since the upper and lower approximations in the classical rough set theory are based on the definition of strict equivalence relations, the continuous numerical characteristics directly applied to the output are discretized, not only introducing quantitative errors, but also increasing the difficulty of model optimization. By introducing the theory of type 2 fuzzy rough set, which could relax the strict equivalence relationship into the fuzzy T equivalence relationship, and defines the fuzzy upper and lower approximation, which is more suitable for processing continuous numerical data. The fuzzy equivalence relationship obtained using Gaussian functions is defined as follows:

$$f(x, y) = e^{-\frac{\left(x^{(m)}-y^{(m)}\right)^2}{\delta}} \tag{1}$$

where $x, y \in U$ are two samples in a full set, $U = (x_1, x_2, \cdots x_n)$ represents the full sample sets, x_m, y_m are the m-layer feature vectors of the type 2 fuzzy rough sets convolutional neural network output.

Fuzzy classification uncertainty based on type 2 fuzzy rough set theory is defined as follows:

$$u(x) = \sum_{d_i} (f_{LW} d_i(x) - f_{BN} d_i(x)) \tag{2}$$

The model of this paper is implemented using Tp operator and Sp operator. Therefore, the fuzzy upper approximation, fuzzy lower approximation, and fuzzy classification uncertainty of sample x can be written as follows:

$$\underline{f}_{LW} d_i(x) = \min_{y \in U}((d_i(y) - 1)e^{-\frac{(x^{(m)} - y^{(m)})^2}{\delta}} + 1)$$
$$\overline{f}_{BN} d_i(x) = \max_{y \in U}(d_i(y)e^{-\frac{(x^{(m)} - y^{(m)})^2}{\delta}}) \quad (3)$$

The uncertainty of fuzzy classification of all samples added together is defined as fuzzy rough loss, which is used as a feature evaluation index in fuzzy expression classification. The fuzzy rough loss function is defined as:

$$J_{T2FRS}(U; \theta) = \sum_{x \in U} u(x) \quad (4)$$

Then we get that:

$$\underline{f}_{LW} d_i(x) \leq d_i(x) \leq \overline{f}_{BN} d_i(x) \quad (5)$$

It can be seen that the fuzzy rough loss function has a lower bound, by optimizing the parameters θ in the type 2 fuzzy rough sets convolutional neural network model, and minimizing the fuzzy rough loss of the training set $J_{T2FRS}(U; \theta)$, you can learn better feature expression.

The type 2 fuzzy rough loss function is used as the loss function layer to train convolutional neural network in fuzzy expression classification problem. From the definition of the type 2 fuzzy rough loss function, we can see that all the calculation steps of the fuzzy rough loss function can be guided. Therefore, through the derivative of the type 2 fuzzy rough loss function on the model, the parameters can be obtained by using the chain derivative method. The following is an example of the derivative of the type 2 fuzzy rough loss function $J_{T2FRS}(U; \theta)$ to the characteristic layer vector x^m, the weight w^m of the full connection layer, and the offset b^m of the full connection layer:

The derivative of the type 2 fuzzy rough loss function $J_{T2FRS}(U; \theta)$ to x^m, w^m, and b^m show as:

$$\frac{\partial J_{T2FRS}(U)}{\partial x^{(m)}} = \sum_{y \in U} \frac{\partial u(y)}{\partial x^{(m)}}$$
$$\frac{\partial J_{T2FRS}(U)}{\partial w^{(m)}} = \sum_{x \in U} \frac{\partial J_{T2FRS}(U)}{\partial x^{(m)}} \frac{\partial x^{(m)}}{\partial w^{(m)}}$$
$$\frac{\partial J_{T2FRS}(U)}{\partial b^{(m)}} = \sum_{x \in U} \frac{\partial J_{T2FRS}(U)}{\partial x^{(m)}} \frac{\partial x^{(m)}}{\partial b^{(m)}} \quad (6)$$

Calculating the derivative of class uncertainty degree $U(x)$ to y^m of the sample x:

$$\frac{\partial u(x)}{\partial y^{(m)}} = \sum_{d_i \in D} \left(\frac{\partial \overline{f}_{BN} d_i(x)}{\partial y^{(m)}} - \frac{\partial \underline{f}_{LW} d_i(x)}{\partial y^{(m)}} \right) \quad (7)$$

where $x, y \in U$ are two samples in a full set, y^m represents the feature vector obtained after the sample y is calculated by "ConvNet".

The Gaussian function $f(x, y)$ derivative to x^m can be obtained as:

$$\frac{\partial e^{-\frac{\left(x^{(m)} - y^{(m)}\right)^2}{\delta}}}{\partial x^{(m)}} = -\frac{2\left(x^{(m)} - y^{(m)}\right) e^{-\frac{\left(x^{(m)} - y^{(m)}\right)^2}{\delta}}}{\delta} \quad (8)$$

From the above formulas, the partial derivative of the fuzzy rough loss function $J_{T2FRS}(U; \theta)$ to the vector of the sample x can be obtained:

$$\frac{\partial \underline{f}_{LW} d_i(x)}{\partial x^{(m)}} = (d_i(y) - 1) \frac{\partial k(x, y)}{\partial x^{(m)}}$$
$$\frac{\partial \underline{f}_{LW} d_i(x)}{\partial y^{(m)}} = (d_i(y) - 1) \frac{\partial k(x, y)}{\partial y^{(m)}}$$
$$y = \arg\min_{y \in U} ((d_i(y) - 1) f(x, y) + 1) \quad (9)$$

$$\frac{\partial \overline{f}_{BN} d_i(x)}{\partial x^{(m)}} = d_i(y) \frac{\partial k(x, y)}{\partial x^{(m)}}$$
$$\frac{\partial \overline{f}_{BN} d_i(x)}{\partial y^{(m)}} = d_i(y) \frac{\partial k(x, y)}{\partial y^{(m)}}$$
$$y = \arg\min_{y \in U}(d_i(y) f(x, y)) \quad (10)$$

And we can also get:

$$\frac{\partial J_{T2FRS}(U; \theta)}{\partial w^{(m)}} = \sum_{x \in U} \frac{\partial J_{T2FRS}(U; \theta)}{\partial z^{(m)}} x^{(m-1)^T}$$
$$\frac{\partial J_{T2FRS}(U; \theta)}{\partial b^{(m)}} = \sum_{x \in U} \frac{\partial J_{T2FRS}(U; \theta)}{\partial z^{(m)}} \quad (11)$$
$$\frac{\partial J_{T2FRS}(U; \theta)}{\partial \theta} = \sum_{x \in U} \frac{\partial J_{T2FRS}(U; \theta)}{\partial z^{(m)}} \frac{\partial z^{(m)}}{\partial \theta} \quad (12)$$

Gradient method could be used to optimize model parameters. In this paper, the experiment uses AlexNet, which is trained in ImageNet, as the initial value of the parameter. Since the initialization network is already a more common feature extractor, the L-BFGS algorithm is used as an optimization algorithm to fine-tune the network on the problem of fuzzy expression classification so that it converges to a minimum point.

III. EXPERIMENTS AND VALIDATION

In order to verify the effectiveness of type 2 fuzzy rough sets convolutional neural network in the fuzzy expression classification task, this article conducted experiments on three public face expression data sets(Jafffe, CK+ and BU-3DFE). type 2 fuzzy rough sets convolutional neural network is used for fuzzy expression classification tasks and compared with existing algorithms by using fuzzy expression tag training. This paper implements the type 2 fuzzy rough sets convolutional neural network model using Keras and Theano based on Python 2.7. The initialization parameters of the model are provided by the convnets-keras project. The parameters of the 5-layer convolutional layer are derived from the AlexNet pre-trained by the ImageNet dataset. The calculation of the forward propagation and backpropagation of the model is done on the GPU.

A. EXPERIMENT RESULTS ANALYSIS ON THREE DATASETS

We used Three databases in our experiments. BU-3DFE Database: The BU-3DFE multi-view facial expression database [24] contains 100 subjects of different ethnicities, including 56 females and 44 males. Six facial expressions (anger, disgust, fear, happiness, sadness, and surprise) are elicited by various manners and head poses. The Japanese Female Facial Expression (JAFFE) database has 213 facial images in total, from 10 different people displaying six kinds of expressions: anger, disgust, fear, happy, sad, surprise and neutral [25]. For each individual, there are three or four images, each 256 × 256 pixels. These face images are frontal and nearly the same size; The Extended Cohn-Kanade database (CK+) is extended From the Cohn-Kanade database (CK) [26]. Which consists of 593 sequences from 123 subjects, the sequences contain many facial expressions performed by individuals, with each image 640 × 490 pixels. This database includes facial expression from adults, in which 69% are female and 81 and 13% are Euro-American and Afro-American, respectively. Ages of participants are distributed from 18 to 50. Seven expressions are labeled in the database: anger, contempt, disgust, fear, happy, sadness, surprise.

We tested our methods on three widely used FER datasets: CK+, BU-3DFE and JAFFE. The CK+ dataset includes expressions of seven labels: contempt, disgust, fear, happy, anger, comtempt, surprise. The BU-3DFE dataset has six labels: anger, disgust, fear, happiness, sadness, and surprise. The JAFFE dataset has seven labels: fear, happy, sadness, anger, disgust, surprise and neutral. To evaluate the overall performance, the confusion matrices of our methods on three datasets are illustrated in Fig.2. Fig.2 (A)–(C) are experimental results on CK+ database, BU-3DFE dataset, JAFFE dataset implemented by proposed type 2 fuzzy rough sets convolutional neural network respectively.

From the above figures, we can see the interesting thing is that happy label has the highest accuracy in the dababases, which means that happy expression is more distinguishable than the other ones. But, the surprise and sadness expressions are likey to be confused on the trained process, because these two labels have high accuract on BU-3DFE and CK+ datasets, but fail to do well on JAFFE. For example, some instances show that their true label is angry but the classifier has misclassified as fear or surprise. So it can be difficult to recognize whether an angry expression is actually surprise or angry even for a human.

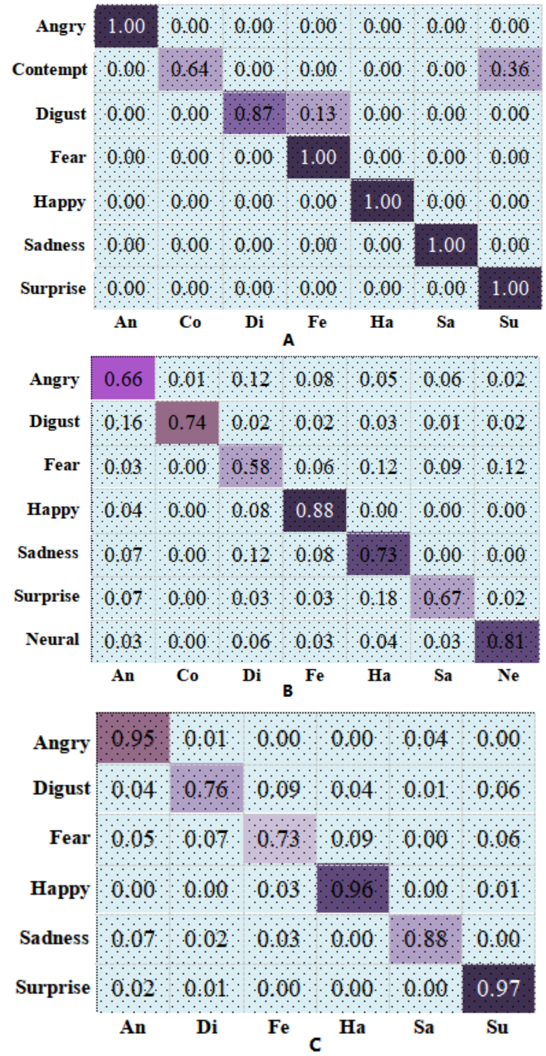

FIGURE 2. Confusion matrices for type 2 fuzzy rough sets convolutional neural network on three expression databases.

B. SIX DIFFERENT METRICS ON JAFFE AND BU-3DFE DATABASES

In this paper, six different metrics are used to evaluate the prediction results of Canberra distance, Chebyshev distance, Clark distance, KL divergence, Cosine cosine similarity and Intersection similarity for better verification of the type 2 fuzzy rough sets convolutional neural network. Detailed experimental results are given in Tables 1 and 2. The Algorithm Adaption k-Nearest-Neighbors classifier was used in the experiment for the classification test. In each subgraph, "LBP" indicates the use of the 243-dimensional LBP histogram as a feature representation of the face expression picture. "AN(FC1)" and "AN(FC2)" are the first two layers of fully connected layer output of AlexNet pre-trained by the ImageNet data set, respectively. "AN-FT(MSE)" and "AN-FT(KL)" respectively represent the fuzzy classification results obtained by AlexNet using the mean square error and KL divergence as loss functions. "type 2 fuzzy rough sets convolutional neural network" is the network structure and training method proposed in this paper. This method uses the type 2 fuzzy rough loss to fine-tune the first five layers of convolutional layer and one layer of fully connected layer of AlexNet, and the feature is extracted as a feature extractor.

Table.1 gives the experimental results for the BU-3DFE data set. The features are obtained in the type 2 fuzzy rough sets convolutional neural network training task, the fuzzy

TABLE 1. Comparison results on BU-3DFE database.

	canberra	chebyshev	clark	cosine	intersection	kldist
LBP	0.8256(+0.00%)	0.1265(+0.00%)	0.3975(+0.00%)	0.9229(+0.00%)	0.8496(+0.00%)	0.0788(+0.00%)
AN-FT(MSE)	0.5822(+29.48%)	0.0883(+30.2%)	0.2827(+28.88%)	0.9606(+4.08%)	0.8948(+5.32%)	0.0392(+50.25%)
AN-FT(KL)	0.9811(-18.83%)	0.1641(-29.72%)	0.4962(-24.83%)	0.8638(-6.4%)	0.8116(-4.47%)	0.1463(-85.66%)
AN(FCl)	0.5102(+38.21%)	0.0787(+37.80%)	0.2486(+37.46%)	0.9692(+5.02%)	0.9078(+6.84%)	0.0303(+61.52%)
AN(FC2)	0.5264(+36.24%)	0.0814(+35.67%)	0.2566(+35.45%)	0.9674(+4.83%)	0.9047(+6.48%)	0.0321(+59.32%)
FRCNN	0.4240(+50.26%)	0.0554(+45.87%)	0.1858(+49.56%)	0.9778(+2.15%)	0.9237(+3.56%)	0.0132(+57.47%)
T2FRSCNN	**0.3240(+60.76%)**	**0.0457(+63.88%)**	**0.1558(+60.81%)**	**0.9878(+7.04%)**	**0.9437(+11.07%)**	**0.0122(+84.49%)**

TABLE 2. Comparison results on JAFFE database.

	canberra	chebyshev	clark	cosine	intersection	kldist
LBP	0.8603(+0.00%)	0.1285(+0.00%)	0.4018(+0.00%)	0.9260(+0.00%)	0.8455(+0.00%)	0.0757(+0.00%)
AN-FT(MSE)	0.7139(-64.18%)	0.8285(-54.75%)	2.3493(-44.69%)	0.9866(-58.25%)	0.1715(-79.72%)	0.1659(-71.76%)
AN-FT(KL)	0.9001(-47.63%)	0.1349(-4.98%)	0.4152(-3.33%)	0.9195(-0.70%)	0.8388(-0.79%)	0.0827(-9.25%)
AN(FCl)	0.6514(+24.28%)	0.0830(+35.44%)	0.3199(+20.37%)	0.9643(+4.14%)	0.8928(+5.58%)	0.0437(+42.23%)
AN(FC2)	0.4937(+42.62%)	0.0624(+51.45%)	0.2400(+40.27%)	0.9792(+5.75%)	0.9179(+8.56%)	0.0229(+69.75%)
FRCNN	0.4531(+73.69%)	0.0541(+49.45%)	0.2031(+47.42%)	0.9572(+4.93%)	0.9148(+39.47%)	0.0202(+49.12%)
T2FRSCNN	**0.4031(+53.15%)**	**0.0518(+59.71%)**	**0.1949(+51.49%)**	**0.9858(+6.46%)**	**0.9323(+10.26%)**	**0.0152(+79.92%)**

classification results are obtained in the fuzzy expression recognition task based on Algorithm Adaption k-Nearest-Neighbors classification. It can be considered that type 2 fuzzy rough sets convolutional neural network effectively learns relevant knowledge from fuzzy multi-labels. Therefore, the fuzzy classification effect is better than "AN(FC1)" and "AN(FC2)". As shown in Table.2, using Algorithm Adaption k-Nearest-Neighbors as the fuzzy classification algorithm, under the six metrics, all deep convolutional neural networks extract features are better than "LBP". Compared with other features, type 2 fuzzy rough sets convolutional neural network achieves good fuzzy classification accuracy under various indicators. Since the performance of the Algorithm Adaption k-Nearest-Neighbors algorithm depends on the distinguishability of the feature space, it can be considered that compared to other algorithms, type 2 fuzzy rough sets convolutional neural network model maps the original picture to a space more suitable for distinguishing facial expressions, that is, a face picture with a similar expression.

C. COMPARISONS WITH THE STATE-OF-THE-ART METHODS

To evaluate the performance of the proposed algorithm with other algorithms, Tables. 3 and 4 list the accuracy of our proposed and the state-of-the-art algorithms on the CK+ and BU-3DFE databases.

TABLE 3. The accuracy (%) on CK+ dataset.

Methods	Expression	Accuracy(%)
LBP [27]	6+neutral	85.20
HOG [28]	6+neutral	84.70
Gabor filter [34]	7	87.80
Deepak [35]	6	93.10
AU-DNN [30]	6+neutral	91.05
JFDNN [32]	6	96.8
CNN [29]	6	93.1
T2FRSCNN	**7**	**95.56**

In this part, the LBP, HOG and Gabor fifilters are traditional feature descriptors in facial expression recognition and have been widely used. However, the recognition accuracy of most traditional methods are lower than that of deep learning. For the CK+ database, the accuracy of our algorithm is superior to most of the other advanced algorithms. The

TABLE 4. The accuracy (%) on BU-3DFE dataset.

Methods	Poses	Accuracy (%)
HOG [36]	5	52.64
LGBP [37]	7	74.1
JFDNN [32]	5	76.5
PCRF [38]	5	78.1
CGPR [40]	5	79.5
GSRRR [39]	5	79.90
DNN-Driven [31]	5	80.10
C-CNN [33]	1	90.07
T2FRSCNN	**5**	**90.26**

comparison results on BU-3DFE dataset are shown in Table. 4, Our best result reaches 90.26%, which is competitive with the other methods.

IV. CONCLUSION

We developed a type 2 fuzzy rough convolutional neural network model for facial expression recognition and evaluated their performances using different analyzing techniques. The results demonstrated that the proposed model has better performance on facial feature learning and emotion classification. In future work, we will try to propose one general model, and then use cross-databases training network parameters to get better generalization capabilities.

REFERENCES

[1] T. Ojala, M. Pietikäinen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[2] C. Shan, S. Gong, and P. W. Mcowan, "Robust facial expression recognition using local binary patterns," in *Proc. IEEE Int. Conf. Image Process.*, Jun. 2005, pp. 370–378.

[3] X. Geng and R. Ji, "Label distribution learning," in *Proc. IEEE 13th Int. Conf. Data Mining Workshops (ICDMW)*, Dec. 2013, pp. 377–383.

[4] X. Geng, "Label distribution learning," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 7, pp. 1734–1748, Jul. 2016.

[5] Y. Zhou, H. Xue, and X. Geng, "Emotion distribution recognition from facial expressions," in *Proc. 23rd ACM Int. Conf. Multimedia*, 2015, pp. 1247–1250.

[6] J. Yang, M. Sun, and X. Sun, "Learning visual sentiment distributions via augmented conditional probability neural network," in *Proc. AAAI*, 2017, pp. 224–230.

[7] B.-B. Gao, C. Xing, C.-W. Xie, J. Wu, and X. Geng, "Deep label distribution learning with label ambiguity," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 2825–2838, Jun. 2017.

[8] R. Jensen and Q. Shen, "New approaches to fuzzy-rough feature selection," *IEEE Trans. Fuzzy Syst.*, vol. 17, no. 4, pp. 824–838, Aug. 2009.

[9] Y. Sun, "Iterative RELIEF for feature weighting: Algorithms, theories, and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 1035–1051, Jun. 2007.

[10] Q. H. Hu, D. Yu, and W. Pedrycz, "Kernelized fuzzy rough sets and their applications," *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 11, pp. 1649–1667, Dec. 2011.

[11] C. K. Zhu, "Fuzzy expression recognition by combining fuzzy rough sets and deep convolutional neural networks," Ph.D. dissertation, Tianjin Univ., Tianjin, China, 2017.

[12] D. Chen and Y. Yang, "Attribute reduction for heterogeneous data based on the combination of classical and fuzzy rough set models," *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 5, pp. 1325–1334, Nov. 2016.

[13] R. Wang, D. Chen, and S. Kwong, "Fuzzy-rough-set-based active learning," *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 6, pp. 1699–1704, Nov. 2016.

[14] R. Wang, C. Y. Chow, and S. Kwong, "Ambiguity-based multiclass active learning," *IEEE Trans. Fuzzy Syst.*, vol. 24, no. 1, pp. 242–248, Jul. 2016.

[15] Q. H. Hu, L. Zhang, and S. An, "On robust fuzzy rough set models," *IEEE Trans. Fuzzy Syst.*, vol. 20, no. 4, pp. 636–651, Dec. 2012.

[16] S. Zhao, H. Chen, and C. Li, "A novel approach to building a robust fuzzy rough classifier," *IEEE Trans. Fuzzy Syst.*, vol. 23, no. 4, pp. 769–786, Jun. 2015.

[17] P. X. Wang and Y. Yao, "CE3: A three-way clustering method based on mathematical morphology," *Knowl.-Based Syst.*, vol. 155, pp. 55–65, Sep. 2018.

[18] H. Chen, T. Li, C. Luo, S.-J. Horng, and G. Wang, "A decision-theoretic rough set approach for dynamic data mining," *IEEE Trans. Fuzzy Syst.*, vol. 23, no. 6, pp. 1958–1970, Dec. 2015.

[19] X. Z. Wang, H. J. Xing, and Y. Li, "A study on relationship between generalization abilities and fuzziness of base classifiers in ensemble learning," *IEEE Trans. Fuzzy Syst.*, vol. 23, no. 5, pp. 1638–1654, Nov. 2015.

[20] X. Wang, W. D. Zhang, D. Liu, and X. B. Yang, "Pseudolabel decision-theoretic rough set," *Math. Problems Eng.*, vol. 2019, Mar. 2019, Art. no. 6810796.

[21] X. J. Chen, D. Li, X. Wang, H. Li, and X. B. Yang, "Rough intuitionistic type-2 fuzzy c-means clustering algorithm for MR image segmentation," *IET Image Process.*, vol. 13, no. 4, pp. 607–614, 2019.

[22] P. P. Lucey, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," in *Proc. Workshop CVPR Hum. Commun. Behav. Anal.*, 2010, pp. 94–101.

[23] P. Kumar, S. L. Happy, and A. Routray, "A real-time robust facial expression recognition system using hog features," in *Proc. Int. Conf. Comput., Anal. Secur. Trends*, 2017, pp. 289–293.

[24] A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," in *Proc. Appl. Comput. Vis.*, 2016, pp. 1–10.

[25] M. Liu, S. Li, S. Shan, and X. Chen, "Au-aware deep networks for facial expression recognition," in *Proc. IEEE Int. Conf. Workshops Autom. Face Gesture Recognit.*, Apr. 2013, pp. 1–6.

[26] T. Zhang, W. Zheng, Z. Cui, Y. Zong, J. Yan, and K. Yan, "A deep neural network-driven feature learning method for multi-view facial expression recognition," *IEEE Trans. Multimedia*, vol. 18, no. 12, pp. 2528–2536, Aug. 2016.

[27] H. Jung, S. Lee, J. Yim, S. Park, and J. Kim, "Joint fine-tuning in deep neural networks for facial expression recognition," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 2983–2991.

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

[29] M. Stewart, B. G. Littlewort, I. Fasel, and J. R. Movellan, "Real time face detection and facial expression recognition: Development and applications to human computer interaction," in *Proc. Conf. Comput. Vis. Pattern Recognit. Workshop (CVPRW)*, 2013, p. 53.

[30] D. Ghimire, S. Jeong, S. Yoon, J. Choi, and J. Lee, "Facial expression recognition based on region specifific appearance and geometric features," in *Proc. 10th Int. Conf. Digit. Inf. Manage.*, 2016, pp. 142–147.

[31] Y. Hu, Z. Zeng, L. Yin, X. Zhou, T. S. Huang, and X. Wei, "Multi-view facial expression recognition," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit.*, Sep. 2008, pp. 1–6.

[32] S. Moore and R. Bowden, "Local binary patterns for multi-view facial expression recognition," *Comput. Vis. Image Understand.*, vol. 115, no. 4, pp. 541–558, Apr. 2011.

[33] A. Dapogny, K. Bailly, and S. Dubuisson, "Dynamic pose-robust facial expression recognition by multi-view pairwise conditional random forests," *IEEE Trans. Affect. Comput.*, vol. 10, no. 2, pp. 167–181, 2016.

[34] W. Zheng, "Multi-view facial expression recognition based on group sparse reduced-rank regression," *IEEE Trans. Affective Comput.*, vol. 5, no. 1, pp. 71–85, Jan. 2014.

[35] O. Rudovic, I. Patras, and M. Pantic, "Coupled Gaussian process regression for pose-invariant facial expression recognition," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 350–363.

XIANGJIAN CHEN received the Ph.D. degree from the Department of Mechanical and Electronic Engineering, Changchun Institute of Optics, Fine Mechanics, and Physics, Chinese Academy of Sciences, China, in 2012. She is currently an Associate Professor with the Jiangsu University of Science and Technology (JUST). Her research interests include adaptive control and intelligent control.

DI LI received the Ph.D. degree from the Department of Mechanical and Electronic Engineering, Changchun Institute of Optics, Fine Mechanics, and Physics, Chinese Academy of Sciences, China, in 2012. His research interests include nonlinear control and signal processes.

PINGXIN WANG received the master's degree from the Nanjing University of Science and Technology, Nanjing, China, and the Ph.D. degree in computational mathematics from the Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing. His current research interests include matrix analysis, three-way decision, and rough set.

XIBEI YANG was born in Jiangsu, China, in 1980. He received the B.S. degree from Xuzhou Normal University (XZNU), Xuzhou, China, in 2002, the M.S. degree in computer applications from the Jiangsu University of Science and Technology (JUST), Zhenjiang, China, in 2006, and the Ph.D. degree in computer applications from the Nanjing University of Science and Technology (NJUST), Nanjing, China, in 2010. He is currently an Associate Professor with JUST and a Postdoctoral Researcher with NJUST. He has published more than 100 articles in international journals and presented at international conferences. His research interests include granular computing, rough set, and decision making.