Knowledge-Based Systems 204 (2020) 106172

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Deep reinforcement learning for robust emotional classification in facial expression recognition

Huadong Li [a,b,c], Hua Xu [a,b,*]

[a] State Key Laboratory of Intelligent Technology and Systems, Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China
[b] Beijing National Research Center for Information Science and Technology(BNRist), Beijing 100084, China
[c] Department of Automation, Tsinghua University, Beijing 100084, China

ARTICLE INFO

Article history:
Received 4 February 2020
Received in revised form 17 June 2020
Accepted 20 June 2020
Available online 26 June 2020

Keywords:
Emotion classification
Reinforcement learning
Image selector
Deep neural network

ABSTRACT

For emotion classification in facial expression recognition (FER), the performance of both traditional statistical methods and state-of-the-art deep learning methods are highly dependent on the quality of data. Traditional methods use image preprocessing (such as smoothing and segmentation) to improve image quality. However, the results still fail to meet the quality requirements of the emotion classifiers in FER. To address the above issues, this paper proposed a novel framework based on reinforcement learning for pre-selecting useful images(RLPS) for emotion classification in FER, which is made up of two modules: image selector and rough emotion classifier. Image selector is used to select useful images for emotion classification through reinforcement strategy and rough emotion classifier acts as a teacher to train image selector. Our framework improves classification performance by improving the quality of the dataset and can be applied to any classifier. Experiment results on RAF-DB, ExpW, and FER2013 datasets show that the proposed strategy achieves consistent improvements compared with the state-of-the-art emotion classification methods in FER.[1]

1. Introduction

Emotion classification in facial expression recognition (FER) is an essential task in computer vision. It aims to detect people's emotional states and intentions and can be applied to many fields, such as social robots, medical treatment, and other human–computer interaction. Emotions can generally be divided into seven categories: anger, disgust, fear, happiness, sadness, surprise, and neutral [1]. Many traditional methods use supervised learning for emotion classification, either based on handcrafted features [2,3] or based on deep neuron networks [4,5]. Those existing classification methods strongly rely on high-quality and large-scale data, especially the deep learning method, which has achieved SOTA in emotion classification (see Fig. 1).

As the amount of data increases, the quality of data declines [6]. To address the issue of low-quality data, previous studies focus on two aspects: (1) improve the quality of datasets through image preprocessing methods [7–9] (2) improve the robustness of classification methods [10–12]. For the first aspect, we believe that high-quality image data contain all face's features in FER. Traditional image preprocessing methods (such as face alignment) can help us obtain full-face images. However the classifier may have further requirements on the quality of the image, such as whether the eyes are open or not. Existing image preprocessing methods cannot meet the specific requirements according to different classifiers. For the other aspect, the more robust the classification model, the higher the model complexity.

For the data quality problems mentioned above, this paper proposed a novel framework based on reinforcement learning for pre-selecting useful images for emotion classification in FER, which is made up of two modules: image selector and rough emotion classifier. The image selector is used to select the useful images for emotion classification through reinforcement strategy. It takes the features extracted by the classifier as input and updates the model parameters based on the reward of the classifier. Therefore, the image selector can be divided into two steps: (1) determine whether to select images or not; (2) get rewards from the emotion classifier after completing the entire selection process. Therefore, the image selector can select the images that are the most suitable for emotion classifier.

* Corresponding author at: State Key Laboratory of Intelligent Technology and Systems, Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China.
E-mail addresses: lihd19@mails.tsinghua.edu.cn (H. Li), xuhua@tsinghua.edu.cn (H. Xu).

[1] The code will be available at https://github.com/lhd777/RLPS.

https://doi.org/10.1016/j.knosys.2020.106172
0950-7051/

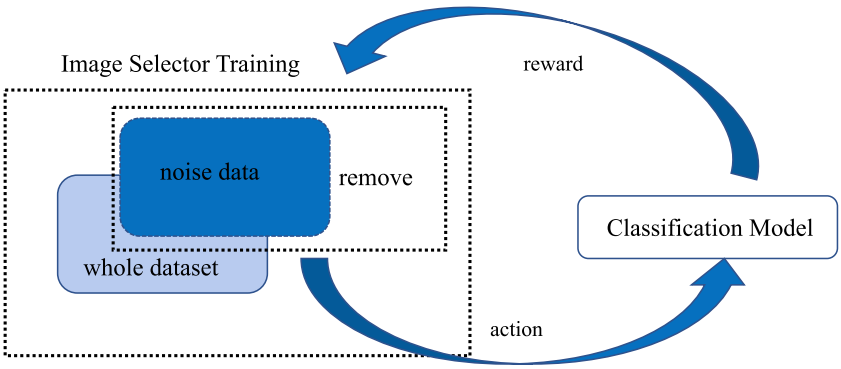


Fig. 1. Our deep reinforcement learning framework aims dynamically determining the noise data, and removing them from dataset. According to the reward from classification model, the image selector updates their parameters.

Our contribution in this paper include:

- We propose a novel emotion classification framework in facial expression recognition based on reinforcement learning, improving the classification accuracy from noise data.
- We propose a novel emotion classification framework that is model-independent. It can be used with any classifier for improving the performance without supervision or any extra cost.
- The proposed framework experiments in RAF-DB and ExpW dataset. It shows that our framework can improve classification accuracy.

2. Related work

2.1. Facial expression recognition

Facial expression recognition is a common task in computer vision. Many traditional works for emotion classification in images are based on handcrafted features [2,3,13]. However, since 2013, emotion recognition competition such as FER2013 [6] and emotion recognition datasets such as RAF-DB and ExpW [14–16] have collected sufficient data from the Internet. At the same time, due to the improvements in chip processing abilities, Studies transferred to the deep learning methods, which have achieved state-of-the-art classification accuracy [4,5].

2.2. Reinforcement learning

Deep learning enables RL to scale to decision-making problems that were previously intractable, i.e., settings with high-dimensional state and action spaces [17]. Amongst recent work in the field of DRL, Mnih developed an algorithm that could learn to play a range of Atari 2600 video games at a superhuman level, directly from image pixels [18]. This work was the first to convincingly demonstrate that RL agents could be trained on raw, high-dimensional observations, solely based on a reward signal. Li K has proposed a deep reinforcement learning for learning selecting optimization function [19]. Therefore, this proves that deep reinforcement learning can directly perform pixel-level learning based on rewards, and can be used for selecting processing.

Dewey [20] said that as reinforcement-learning based AI systems become more general and autonomous, the design of reward mechanisms that elicit desired behaviors becomes both more important and more difficult. The setting of the reward function will determine the speed and degree of convergence of the reinforcement learning algorithm [17]. Therefore, the setting of the reward function is significant for reinforcement learning.

Feng and Qin [21,22] have propose their works in relation classification, which aims to reduce the noise data from the dataset based on reinforcement learning. Both works adopt reinforcement learning to learn a selector. However, firstly both works are used in relation classification, not in facial expression recognition. Secondly, Feng [21] adopts a simple sigmoid function for the selector and cannot achieve complex actions. In contrast, we use deep fully connected neural networks in the image field to implement policy function. And in order to achieve end-to-end training and reduce computational cost, we adopt the policy function and classifier to share the feature extraction module. Thirdly Qin [22] adopt a rewarding setting that relies on the f1-score change of relation classification in epochs. In contrast, the reward in our method is reflected by the prediction probability of emotion classification of every single sample. Therefore, our reward function is more refined and can visually display the reward of every sample. Our approach is also complemented to the most studies above and can be applied to any emotion classification in facial expression recognition or any extra classification tasks.

3. Methodology

We propose a new emotion classification framework, which can be used to remove the noise data from the dataset, then get better classification performance. Our framework is based on two modules, image selector and emotion classification. We consider an RL agent to be an image selector for a robust emotion classifier. The goal of the image selector is to determine whether to retain or remove images. For reinforcement learning, the external environment and RL agent are necessary parts. They interact dynamically with each other [23]. Firstly, reinforcement learning requires the external satisfied Markov decision process(MDP). However, the images in facial expression recognition are independent of each other. Thus we cannot directly use the information of the image. To solve this problem, every image has a state s_i which contains not only the current image but also the average of already selected image's feature vector. The feature vector is gotten from the first dense layer in the emotion classifier. Secondly, in reinforcement learning, Deep Q Network(DQN) [24] is widely applied. But in our case, when one image inputs, emotion classification tasks cannot get the reward immediately. Only after finishing the training process of the classifier, the image selector can get every reward of images. Thus DQN is not suitable for our task.

In our framework, each image x_i has a corresponding action a_i to indicate whether to remove the image from the dataset. The emotion classifier uses remained dataset to train itself. Then the trained emotion classifier input the whole dataset for getting the reward of each image, even the images which have already been removed. Meanwhile, according to the reward of the classifier, the image selector updates its parameters for selecting precisely. Fig. 2 gives an illustration of our proposed framework.

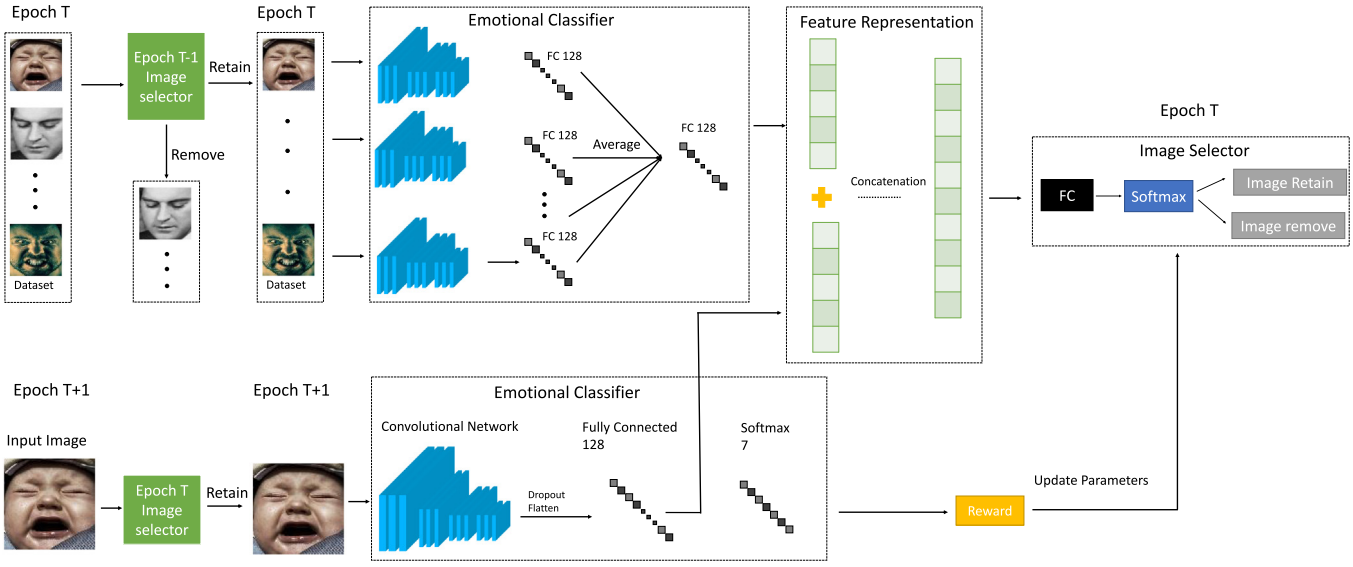


Fig. 2. The proposed image selector is based on reinforcement learning. The image selector gets the feature vector of every image from the first dense layer in emotional classifier. The feature vector is made up of two parts, one is the feature average of all pictures in the previous period, and the other is the feature vector of the current picture. The emotional classifier for each epoch will be retrained by the dataset filtered by the image selector. Then the image selector can get two actions, retain or remove. At the same time, the classifier will give a reward to guide the parameter update of the selector.

3.1. Image selector

We use reinforcement learning to achieve image selecting. In our framework, the image selector is an RL agent, which interacts with the external environment and gets feedback from the external environment. According to the current state, the image selector based on the policy network to determine the current image whether to retain or remove. We will describe our reinforcement setting (i.e., state, action, reward, optimization)

3.1.1. State

The state s_i contains not only the current image but also images that have already been selected. Instead of using the pixel of image as state, the state s_i comes from the outputs of the first dense layer of emotion classify. We represent the state as a continuous real-valued vector $F(s_i)$, that encodes two-part information: (1) the feature vector of current image which is the outputs of the first dense layer of emotion classify (2) The average of the feature vector of images which have already been selected. On the one hand, as the classify using the retained dataset to train itself, it can reflect the change of state due to the change of emotion classify. On the other hand, it can save computation cost using the features which have already been extracted by the classifier.

3.1.2. Action

We define an action $a_i \in \{0, 1\}$ to represent remove or retain. If action is 0, the image selector will remove the image. However, if action is 1, the image selector will retain the image. Therefore it is similar to a binary classification task. In our framework, the value of a_i is gotten from the policy network. The policy network adopts five fully connected layers to achieve binary classification because the inputs come from the first dense layer of emotion classifier which has been extracted features. The last layer's activation adopts sigmoid function.

3.1.3. Policy network

The input of policy network is state vector $F(s_i)$ which is list of features $F(s_i) = (f_1, f_2, \ldots, f_m)$. The features come from the first dense layer of emotion classifier. The state uses the CNN layer in the emotion classifier for extracting features. In order to obtain a high-level selecting accuracy, we adopt a DNN structure for image selector.

$$\begin{aligned}\pi_\theta(s_i, a_i) &= P_\theta(a_i|s_i) \\ &= sigmoid(DNN(F(s_i)))\end{aligned} \tag{1}$$

where the $F(s_i)$ is the state vector which is described by the emotion classifier. *DNN* is the policy network.

3.1.4. Reward

The reward comes from the prediction probability of emotion classifier, indicating the effect of how well image selector work. For each image, the image selector will return action whether to remove. However, the reward has to be delayed, because of waiting for finishing the training process of the emotion classifier.

$$\mathrm{R}_i = P\left\{e_{true}|x_i\right\} - \frac{1}{n}\sum_{j=1}^{n} p\left\{e|x_j\right\} \tag{2}$$

$$Reward = \sum_{i \in B} R_i \tag{3}$$

where x_i is the image. The B represents the retained dataset. The $P\{y|x_i\}$ is calculated by the emotion classifier which is made up of CNN. Besides, the $P\{e_{true}|x_i\}$ is prediction which predicts the true label. Compared to F_1, the accuracy can fully describe the influence of every image to the image selector. The reward can be a positive number or negative number. The higher the reward, the better the image quality is. Naturally, the value of reward is in continuous space, not in discrete space like (-1 and 1). In Eq. (1), reward of images will be constrained in [-1, 1].

The images which do not contribute to the training process of emotion classifier still get rewards to reflect their negative influence. In the selecting process, although the reward is delayed, all the actions contribute to the reward. Thus the reward can still satisfy the requirement of image selector.

3.1.5. Optimization

In our framework, optimizing the parameters in the image selector aims to maximize the value of reward. In actions, we describe that our policy network uses a five fully connected layer.

The last layer adopts sigmoid function as activation. Thus we adopt a modified binary crossentropy as cost function.

$$J(\theta) = \sum_i R_i(y_i \log[\pi\ (a = y_i|s_i;\theta)] + \ (1 - y_i) \log[1 - \pi\ (a = y_i|s_i;\theta)]) \tag{4}$$

where $\pi\ (a|s_i;\theta)$ represent the policy network. The R_i can be positive or negative. R_i is similar as direction indicator. Optimizing the image selector through the R_i. According to the policy gradient theorem [25] and the REINFORCE algorithm [26], we compute our model gradient.

3.2. Emotion classifier

In emotion classifier, we adopt a CNN architecture for predicting emotions. The classifier contains convolution layer, max-pooling layer, and dense layer.

3.2.1. CNN

In order to obtain high-level emotion recognition of each facial image, we adopt a CNN architecture for emotion classification. The structure of the classifier can be briefly shown as Fig. 3, which is made up of six convolution layers, four max-pooling layers, and two dense layers. The Fig. 4 shows the architecture of the-state-of-art model in FER2013 dataset whose code is available in Github.[2]

The inputs of the emotion classifier are not the whole dataset, but the clean dataset which image selector has filtered before. The last layer uses softmax function as activation to achieve seven basic emotion classification.

3.2.2. Loss function

Given the selected image sets $\{\hat{X}\}$ provided by the image selector, we define the cross-entropy function as the loss function of emotion classifier.

$$\mathcal{J}(\theta) = -\frac{1}{|\hat{X}|}\sum_{i=1}^{|\hat{X}|} \log p\ (e_i|x_i;\theta) \tag{5}$$

3.3. Model training

Our framework has two parts: an image selector and an emotion classifier. Two parts are highly corresponding to each other. Hence, we train two parts jointly. At the beginning, we initialize the parameters of the CNN structure of the emotion classifier and the parameters of the DNN structure of the image selector with random weights. Then we pre-train the CNN model with the whole dataset to predict emotions e_i given image x_i. we use the fixed CNN model of emotion classifier to pre-train the policy network. Finally, we can run **Algorithm** 1 to train the policy network and emotion classifier.

We assume that the reward given by the classifier can reflect the quality of the data, that is, the higher the reward given by the classifier, the higher the quality of the data. The Eq. (2) shows that if we predict the data correctly, the reward will be a positive number ($R > 0$). And Eq. (4) loss function can derivate the gradient Eq. (6). Therefore, if the input > 0 and R > 0, then $\theta > 0$ and $predict_{remain}$ would increase. If the input < 0 and R > 0, $\theta < 0$, then $predict_{remain}$ would decrease. Therefore the image selector can learn how to filter the noisy.

$$g \propto \nabla_\phi \sum^{X_i} \log \pi(a|s;\phi)\mathcal{R} \tag{6}$$

Algorithm 1 represents the details of the training process of the image selector, which is based on the reward feedback from the emotion classifier.

[2] https://github.com/WuJie1010/Facial-Expression-Recognition.Pytorch.

Algorithm 1: Image selector training process based on reinforcement learning

Data: Epoch T, given image dataset $X = \{x_1, x_2, ..., x_n\}$, the parameters ϕ of policy network and the parameters θ of CNN model.
Result: Optimize the policy network, updating ϕ

Initialize the parameters, $\phi' = \phi$, $\theta' = \theta$, $\hat{X} = \emptyset$;
pre-train the emotion classifier with whole image dataset, getting $\theta' = \theta_0$;
for *epoch* $t = 1$ *to* T **do**
  $s_i = L_{classify}(X_i)$;
  for $i = 1$ *to* N **do**
    $F(s_i) = concatenate(s_i, s^*)$;
    $a_i \sim \pi(a|F(s_i), \phi')$;
    compute $p_i = \pi(a = 1|F(s_i), \phi')$;
    if $p_i == 0$ **then**
      $\hat{X} = \hat{X} \cup x_i$
    end
  end
  train the emotion classifier based on $\hat{X}$, the parameter $\theta' = \theta_t$;
  input the whole image dataset, get *Reward* and $\hat{R} = \{R_1, R_2, ..., R_n\}$;
  Update $\phi : g \propto \nabla_\phi \sum^{X_i} \log \pi(a|s;\phi)\mathcal{R}$;
end
;

4. Experiments

The image selector is adopted as reinforcement learning agent to improve the quality of images through filtering low-quality images. The emotion classifier is adopted as the environment to give the reward and the state for the image selector. In this paper, experiments will be tested in three datasets, including RAF-DB, ExpW, and FER2013 [6,14–16]. The three datasets are the newest and largest dataset in facial expression recognition which are collected and labeled in the web environment, different from the datasets, such as CK+ and JAFFE [27,28].

4.1. Dataset

Real-world Affective Faces Database (RAF-DB) is a large-scale facial expression database with around 30 K great-diverse facial images downloaded from the Internet. RAF-DB contains: (1) 29672 number of real-world images, but only 15539 number of images with basic emotion label; (2) two different subsets: single-label subset, including seven classes of basic emotions; two-tab subset, including 12 classes of compound emotions. We choose a single-label subset for the experiment. [15] spilted the RAF-DB as the training dataset with 12271 images and test dataset with 3068 images. Then, we split the training dataset as a new training dataset with 9816 images and valid dataset with 2455 images, where the size of the new training dataset is four times larger than the valid dataset.

Expression in-the-wild(ExpW) collects facial images from the Internet. ExpW contains: (1) 91793 number of facial expression images (2) six basic emotion types and one plus neutral emotion. Without finding any official splitting proportion, we spilt the ExpW as a training dataset, valid dataset and test dataset, where the size of the training dataset is eight times larger than the valid dataset and test dataset. The training dataset has 73433 images. Valid dataset and test dataset respectively include 9180 images.

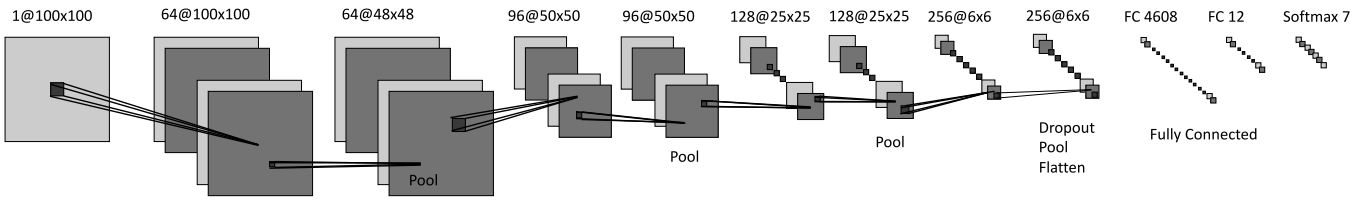


Fig. 3. It shows that the architecture of emotion classifier for RAF-DB and ExpW. In convolution layer, classifier use 3x3 convolution kernel to extract features.

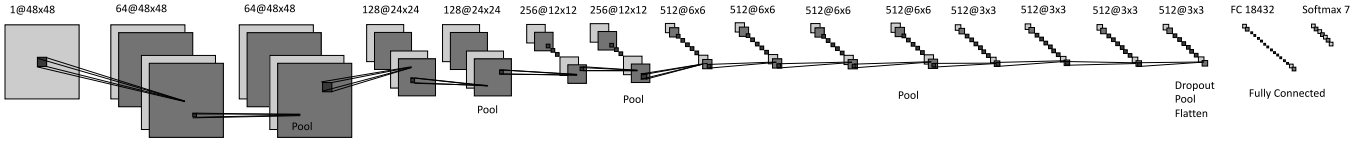


Fig. 4. It shows that the SOTA architecture of emotion classifier for FER2013. In convolution layer, classifier use 3x3 convolution kernel to extract features.

Two datasets have same seven basic emotion classification labels: anger, disgust, fear, happiness, sadness, surprise and neutral.

Facial expression recognition 2013 (FER2013) is from challenges in representation learning in ICML 2013. The data of FER2013 consists of 48 × 48 pixel grayscale images of faces. The faces have been automatically registered so that the face is more or less centered and occupies about the same amount of space in each image. FER2013 contains: (1) 28,709 images for training and 3589 images for public test, 3589 for private test (2) it is similar to above dataset with seven basic emotion classification labels. The dataset FER2013 was presented in Kaggle competition. Therefore, there are two test dataset, public test dataset and private test dataset. Competitors design their model without the private test dataset. Finally, contest organizer will re-run the participating models on the private test dataset and rank them by the new score. So we also use the score in the private test dataset to evaluate our framework.

4.2. Evaluation

We will evaluate the experiment to describe the effects of our framework from two metrics. One objective metric uses the change of emotion classification total accuracy, average accuracy and macro F1 to describe the improvement of our framework. The other subjective metric is that judging the filtered image whether is reasonable by person. Each experiment will be repeated ten times to get the average value.

Given a set of classes $C = \{C_1, C_2, \ldots, C_n\}$, we compute macro F1-score and average accuracy as the following:

$$F_{1,\text{macro}} = 2\frac{\text{recall macro} \times \text{precision}_{\text{macro}}}{\text{recall macro} + \text{precision macro}} \tag{7}$$

$$\text{average accuracy} = \frac{\sum_{i=1}^{N} \text{accuracy}_{C_i}}{N} = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_{C_i}}{TP_{C_i} + FP_{C_i}} \tag{8}$$

$$\text{precision}_{\text{macro}} = \frac{\sum_{i=1}^{N} \text{precision}_{C_i}}{N}, \tag{9}$$

$$\text{recall}_{\text{macro}} = \frac{\sum_{i=1}^{N} \text{recall}_{C_i}}{N}$$

$$\text{precision}_{C_i} = \frac{TP_{C_i}}{TP_{C_i} + FP_{C_i}}, \quad \text{recall}_{C_i} = \frac{TP_{C_i}}{TP_{C_i} + FN_{C_i}} \tag{10}$$

where C_i denotes the individual class in the set of classes C.

4.3. Baseline methods

- DCNN+Sigmoid [12].[3] We transfer the framework of the AAAI 2018 paper of [12] into facial expression recognition task. It has already achieved the SOTA performance in relation classification in NLP. It also considers the noise labeling problem. Due to the different areas, we make some changes for suiting FER task during the transferring process, such as deleting the bags setting.
- DCNN. We use the whole dataset for training the classifier without selecting. It does not consider the noise labeling problem.

[3] https://github.com/JuneFeng/RelationClassification-RL.

4.4. Experiment with original dataset

4.4.1. Experiment setting

Image Selector. The action space of image selector based on reinforcement learning includes two actions, retain or remove. We adopt five fully connected layers as a policy network, using the ReLU function as activation function. As for the inputs, it concatenates the current state and previous average state. As is shown in Fig. 3, the state, coming from the first dense layer, is a 128 dimensions' vector. Therefore the $F(s_i)$ is a 256 dimensions' vector.

Emotion Classification. In order to evaluate the actions of the image selector, we use a CNN model to get the rewards of actions. For three datasets, we adopt two different classification models. One classification model for dataset RAF-DB and ExpW is not the state-of-the-art model. The other classification model for FER2013 is a state-of-the-art model [29]. On the one hand, the simple classification model (Not SOTA) is more sensitive to the quality of training set. On the other hand, the state-of-the-art model can test our framework whether can improve performance, even if the classification model has reached the best performance before. The CNN structure of the first classification model is shown in Fig. 3, whose batch size is fixed to 128. It contains six convolution layers, four max-pooling layers and two dense layers. The activation of the last layer is softmax for seven classifications. As described in algorithm 1, we pre-train the emotion classifier with the whole dataset before training image selector. It can help the convergence of the image selector, because of high-level state vector.

We experiment with our framework in three datasets, including RAF-DB, ExpW, and FER2013. There are no studies in emotion classification which aim to filter the low-quality data through reinforcement strategy for improving the performance of emotion classification. There is some similar research in relation classification [12,22]. Therefore we adopt two baselines for comparison (see Fig. 6).

(a) 2.13% (b) 5.73% (c) 65.79%

Fig. 5. It shows the remove probability of the above images.

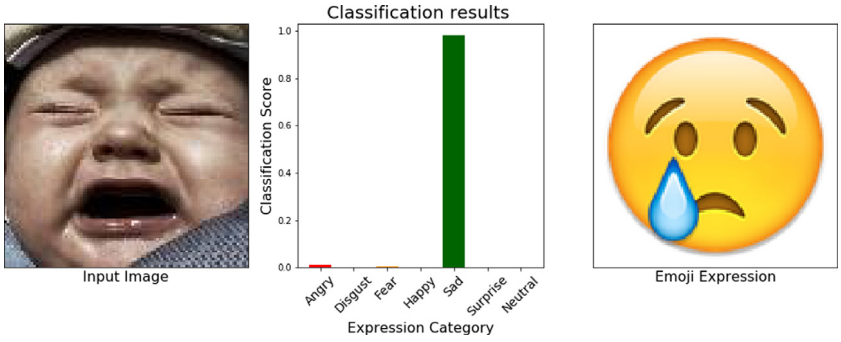


Fig. 6. It shows the predict probability of an image.

4.4.2. Result

RAF-DB & ExpW. As is shown in Table 2, our method not only improves the performance of emotion classifier but also perform much better than baselines. In RAF-DB, our method improve 6.65 percent average accuracy and 3.05 percent macro F1-score compared with the DCNN baseline. As the accuracy of each emotion class changes, we can see that our proposed method effectively alleviates the problem of unbalanced prediction. In Table 2 The accuracy of the baseline model improved by our method is close to the state-of-the-art model.

In Table 3, the model which adds our method show better performance than original baseline model, improving 10.75 percent average accuracy and 2.34 percent macro F1-score. Before using our method, some emotional features cannot be learned with the baseline model. Because the number of the dataset is too large relative to the depth of the model. After using our pre-selecting method, the model can achieve a certain classification function for disgust and fear emotion.

FER2013. For a fair comparison with above dataset, we have not adopted data-enhanced data preprocessing method to improve the classification performance. This means that we do not use translation, rotation, etc. to expand the dataset. Therefore the rebuild classification model's performance 0.7021 is little lower than the performance in paper whose accuracy is 0.7311. The experiments show that our method can improve the performance of classification model, even if the model has reach the SOTA. Both accuracy and macro f1 have been improved. As is shown in Table 1, our model(DCNN+RLPS) has achieved the state-of-the-art performance in single classifier model.

It can be seen from the experiments that our method improves the weak model better than the model that is close to SOTA. As is shown in Fig. 5, people's eyes are closed in Fig. 5(b) and 5(c). Fig. 5(c), as we thought, was removed because it lacked obvious features. However, Fig. 5(b) is retained in a high probability way. The reason we think is that although the eyes in Fig. 5(b) are closed, the feature of crying is very obvious. So the image selector tends to retain the image. In general, it tends to retain images that have distinctive features. Because if an image has enough distinguishing features, it is helpful to improve the accuracy of the classifier. Then the image selector will increase the probability of retaining this image.

Table 1
Comparison of the performance between the state-of-the-art method and our RLPS method in dataset FER2013). The metric is total accuracy.

Method	Total Acc (%)	Strategy
Gan et al.[30]	**73.73**	Ensemble
Kim et al.[31]	**73.73**	Ensemble
kim et al.[32]	72.72	Ensemble
Gan et al.[30]	71.47	Single
Winning Method in [6]	71.16	Single
Shao et al.[33]	71.14	Single
kim et al.[32]	70.58	Single
Wen et al.[34]	69.96	Ensemble
The first runner-up of [6]	69.27	Single
Agrawal et al.[35]	65.77	Single
DCNN	70.21	Single
DCNN+Sigmoid	71.23	Single
DCNN+RLPS(ours)	**72.35**	Single

4.5. Experiment with extra noise dataset

4.5.1. Experiment setting

In order to show the robustness of our framework. We add noise information in the datasets above and get new noise dataset. For the training set of RAF-DB and ExpW, we intercept 50% full face images, 25% half face images, 25% none face images. Fig. 7 shows the processed result. As for the valid set and test set, there are no changes of them. The same operation is applied to the ExpW dataset. Notice that the valid set and test set in noise dataset are still clean.

In order to prove our reward setting is technical sounding, we also compare our framework with two reward function. One is Eq. (2). The other is the change of f1 score. Figs. 8 and 9 show the visualization result between different reward setting.

4.5.2. Result

Table 4 shows that our framework reflects better results in Experiment2 and achieves better accuracy improvements. Table 5

Table 2
Comparison of the performance between the state-of-the-art method and our RLPS method in dataset RAF-DB. For compared with other paper, we use the mean diagonal value of the confusion matrix as accuracy metric).

Method	Anger	Disgust	Fear	Happiness	Sadness	Surprise	Neutral	Average Acc (%)
VGG+mSVM [14]	68.52	27.50	35.13	85.32	64.85	66.32	59.88	58.22
AlexNet+mSVM [14]	58.64	21.87	39.19	86.16	60.88	62.31	60.15	55.60
DLP-CNN+LDA [14]	77.51	55.41	52.50	90.21	73.64	74.07	73.53	70.98
DLP-CNN+mSVM [14]	71.60	**52.15**	62.16	**92.83**	**80.13**	81.16	80.29	**74.20**
DCNN	61.20	43.61	38.21	88.56	74.77	77.92	79.06	66.19
DCNN+Sigmoid	64.13	41.42	45.10	91.89	71.09	**84.72**	**82.71**	68.72
DCNN+RLPS(ours)	**72.90**	51.02	**77.52**	90.42	66.22	79.09	72.72	**72.84**

Table 3
Comparison of the performance between the state-of-the-art method and our RLPS method in dataset ExpW. For compared with other paper, we use the mean diagonal value of the confusion matrix as accuracy metric).

Method	Anger	Disgust	Fear	Happiness	Sadness	Surprise	Neutral	Average Acc (%)
HOG + SVM [16]	54.30	54.80	56.20	71.30	58.40	61.20	68.40	60.66
Baseline DCN [16]	63.50	50.00	50.00	81.90	64.00	71.00	75.00	65.06
DCN + AP [16]	**72.10**	**56.50**	**58.70**	**83.80**	69.10	74.20	76.00	**70.06**
DCNN	50.78	0.00	0.00	78.25	26.34	31.00	91.59	39.86
DCNN+Sigmoid	46.56	24.23	32.14	71.38	22.81	41.42	**86.72**	46.47
DCNN+RLPS(ours)	67.09	40.73	23.46	74.95	38.66	38.97	70.32	**50.61**

Table 4
Comparison of the performance of accuracy and macro F1 between the baselines and our RL method in noise dataset, including RAF-DB, ExpW).

Dataset	Method	Anger	Disgust	Fear	Happiness	Sadness	Surprise	Neutral	Average Acc (%)	Macro F1
ExpW	DCNN	45.30	0.00	0.00	67.27	15.25	28.14	**86.30**	34.61	0.3411
	DCNN+Sigmoid	**55.42**	24.13	28.54	66.77	18.60	22.10	83.60	42.65	0.3524
	DCNN+RLPS+F1(ours)	47.74	30.37	**33.42**	68.62	22.67	**33.23**	83.45	45.64	0.3629
	DCNN+RLPS(ours)	48.20	**30.55**	33.19	**70.11**	**24.00**	**31.65**	84.06	**45.96**	**0.3674**
RAF-DB	DCNN	43.08	12.00	24.16	90.77	62.78	66.62	74.59	52.43	0.6484
	DCNN+Sigmoid	46.77	26.43	30.90	88.05	58.62	72.94	**78.52**	57.46	0.6673
	DCNN+RLPS+F1(ours)	56.22	**34.55**	**44.05**	89.78	50.32	**68.90**	69.78	59.08	0.6733
	DCNN+RLPS(ours)	**57.00**	33.22	42.20	**91.55**	**51.81**	68.68	70.91	**59.34**	**0.6769**

(a) Full Face (b) Half Face (c) None Face

Fig. 7. The three above images show that the results of interception.

shows the probability of different noise images. It can be seen from experiments that when our dataset contains full-face, half-face, and none-face, image selector can achieve function which is similar to face recognition. Both the full-face and half-face images are benefiting to the classifier. However, the none-face images do not have the benefit to the classifier. Thus the image selector learns to select as many full-face and half-face images as possible.

We thought that our model would have better results in a very noisy environment. So we add an experiment under the adding noisy dataset. However the experiment does not show that our model has obvious advantages in the adding noisy dataset. Table 5 shows that the image selector in adding noisy dataset tend to select images with facial features and filter out images without faces. However Fig. 5 shows that the image selector tends to select images with obvious emotional features. Compared with whether there is a face, capturing more refined emotional feature may be more helpful for pre-selecting and classification.

We also compared two different reward function settings. From Figs. 8 and 9, it can be seen that both reward functions can achieve convergence. And from the Table 4 the final classifier

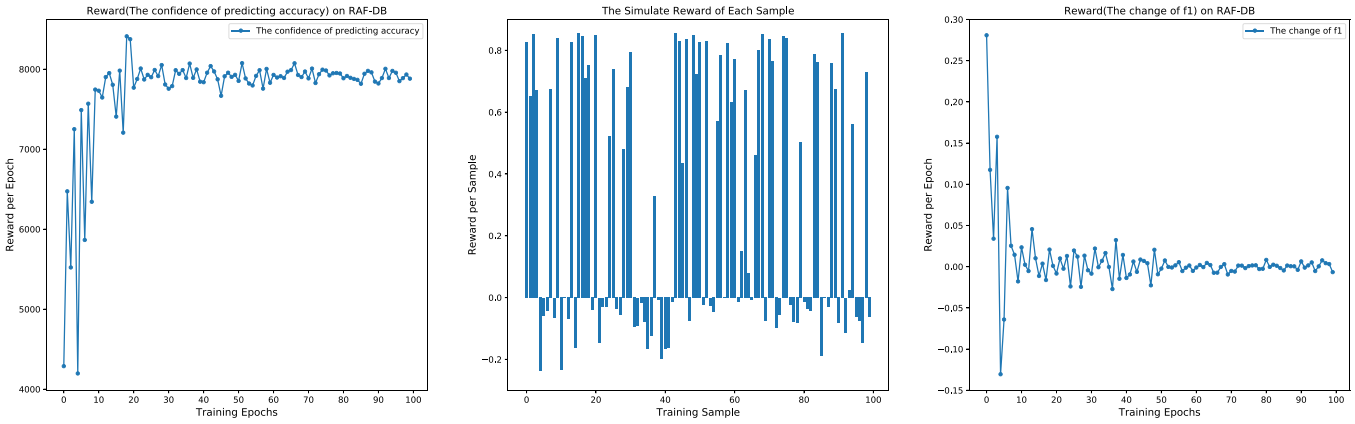
Table 5
Comparison of the selected probability in different type images.

Dataset/Prob	Total face	Half face	None face
RAF-DB	0.9618	0.8909	0.3218
ExpW	0.9786	0.8201	0.2891

results of two reward setting are similar. This is because both are based on the prediction results of the classifier for feedback. However, it can be seen from Figs. 8, 9 that our newly set reward function can have a faster convergence speed and can visually display the reward feedback of each sample.

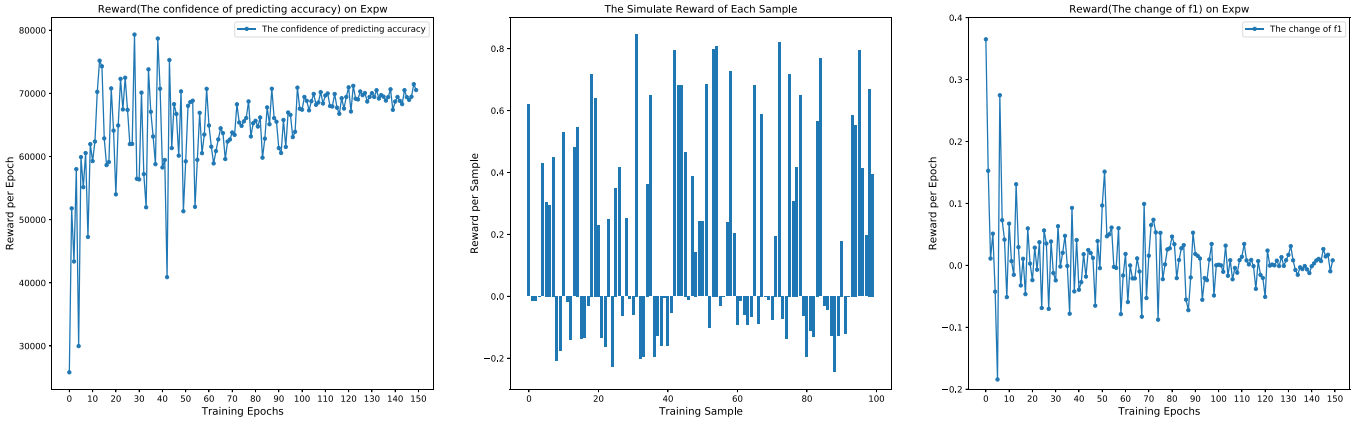
5. Conclusion and future work

In this paper, we propose a deep reinforcement learning framework for robust emotion classification in facial expression recognition. The intuition is that, unlike previous FER studies, we improve the performance of classifier through filtering noise data based on deep reinforcement learning. In relation classification,



(a) accuracy confidence curve (b) simulate reward (c) changing f1 score ccurve

Fig. 8. We compared the two reward settings with its in dataset RAD-DB. The Figure(a)(c) show the reward curve for reward per epoch. Beside, we sample 100 samples in training data and visualize those simulate reward in figure(b).



(a) accuracy confidence curve (b) simulate reward (c) changing f1 score ccurve

Fig. 9. We compared the two reward settings with its in dataset ExpW. The Figure(a)(c) show the reward curve for reward per epoch. Beside, we sample 100 samples in training data and visualize those simulate reward in figure(b).

there are some researches for selecting strategy. We transfer the SOTA selecting strategy in relation classification to emotion classification, and compare it with our framework. The experiments on RAF-DB and ExpW demonstrate that our framework can get better performance of emotion classification in FER. Besides, the proposed framework is model-independent which can be added to any classifier and improve the performance. The experiments also show that our deep reinforcement learning framework can largely improve the emotion classification from noise data.

In the future, we will transfer our framework to other tasks that live in noise environment. We will reduce the dependence on pseudo-supervisor (classifier) since the performance of our framework highly relies on its feedback (reward).

CRediT authorship contribution statement

Huadong Li: Conceptualization, Methodology, Software, Writing - original draft. **Hua Xu:** Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This paper is funded by National Natural Science Foundation of China (Grant No: 61673235) and National Key R&D Program Projects of China (Grant No: 2018YFC1707605).

References

[1] P. Ekman, W.V. Friesen, Constants across cultures in the face and emotion, J. Pers. Soc. Psychol. 17 (2) (1971) 124.
[2] C. Shan, S. Gong, P.W. McOwan, Facial expression recognition based on local binary patterns: A comprehensive study, Image Vis. Comput. 27 (6) (2009) 803–816.
[3] L. Zhong, Q. Liu, P. Yang, B. Liu, J. Huang, D.N. Metaxas, Learning active facial patches for expression analysis, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2012, pp. 2562–2569.
[4] I. Sutskever, G.E. Hinton, A. Krizhevsky, Imagenet classification with deep convolutional neural networks, Adv. Neural Inf. Process. Syst. (2012) 1097–1105.
[5] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2014, arXiv preprint arXiv:1409.1556.
[6] I.J. Goodfellow, D. Erhan, P.L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee, et al., Challenges in representation learning: A report on three machine learning contests, in: International Conference on Neural Information Processing, Springer, 2013, pp. 117–124.
[7] O.M. Parkhi, A. Vedaldi, A. Zisserman, et al., Deep face recognition, in: Bmvc, vol. 1, 2015, p. 6.
[8] Y. Sun, D. Liang, X. Wang, X. Tang, Deepid3: Face recognition with very deep neural networks, 2015, arXiv preprint arXiv:1502.00873.

[9] R. Hossain, M. Rahman, O.A. Tania, et al., Image Capturing and Automatic Face Recognition, United International University, 2019.
[10] C. Pramerdorfer, M. Kampel, Facial expression recognition using convolutional neural networks: state of the art, 2016, arXiv preprint arXiv:1612.02903.
[11] H.-D. Nguyen, S. Yeom, I.-S. Oh, K.-M. Kim, S.-H. Kim, Facial expression recognition using a multi-level convolutional neural network, in: Proceedings from the International Conference on Pattern Recognition and Artificial Intelligence, Centre for Pattern Recognition and Machine Intelligence, 2018, pp. 217–221.
[12] Y. Fan, J.C. Lam, V.O. Li, Multi-region ensemble convolutional neural network for facial expression recognition, in: International Conference on Artificial Neural Networks, Springer, 2018, pp. 84–94.
[13] R. Takanobu, T. Zhang, J. Liu, M. Huang, A hierarchical framework for relation extraction with reinforcement learning, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, 2019, pp. 7072–7079.
[14] S. Li, W. Deng, J. Du, Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2852–2861.
[15] S. Li, W. Deng, Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition, IEEE Trans. Image Process. 28 (1) (2018) 356–370.
[16] Z. Zhang, P. Luo, C.C. Loy, X. Tang, From facial expression recognition to interpersonal relation prediction, Int. J. Comput. Vis. 126 (5) (2018) 550–569.
[17] K. Arulkumaran, M.P. Deisenroth, M. Brundage, A.A. Bharath, Deep reinforcement learning: A brief survey, IEEE Signal Process. Mag. 34 (6) (2017) 26–38.
[18] V. Mnih, K. Kavukcuoglu, D. Silver, A.A. Rusu, J. Veness, M.G. Bellemare, A. Graves, M. Riedmiller, A.K. Fidjeland, G. Ostrovski, et al., Human-level control through deep reinforcement learning, Nature 518 (7540) (2015) 529–533.
[19] K. Li, J. Malik, Learning to optimize neural nets, 2017, arXiv preprint arXiv:1703.00441.
[20] D. Dewey, Reinforcement learning and the reward engineering principle, in: 2014 AAAI Spring Symposium Series, 2014.
[21] J. Feng, M. Huang, L. Zhao, Y. Yang, X. Zhu, Reinforcement learning for relation classification from noisy data, in: Thirty-Second AAAI Conference on Artificial Intelligence, 2018.
[22] P. Qin, W. Xu, W.Y. Wang, Robust distant supervision relation extraction via deep reinforcement learning, 2018, arXiv preprint arXiv:1805.09927.
[23] K. Arulkumaran, M.P. Deisenroth, M. Brundage, A.A. Bharath, A brief survey of deep reinforcement learning, 2017, arXiv preprint arXiv:1708.05866.
[24] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, M. Riedmiller, Playing atari with deep reinforcement learning, 2013, arXiv preprint arXiv:1312.5602.
[25] R.S. Sutton, D. Precup, S. Singh, Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning, Artif. Intell. 112 (1–2) (1999) 181–211.
[26] R.J. Williams, Simple statistical gradient-following algorithms for connectionist reinforcement learning, Mach. Learn. 8 (3–4) (1992) 229–256.
[27] P. Lucey, J.F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, I. Matthews, 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, IEEE, 2010, pp. 94–101.
[28] M. Lyons, S. Akamatsu, M. Kamachi, J. Gyoba, Coding facial expressions with gabor wavelets, in: Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition, IEEE, 1998, pp. 200–205.
[29] Z. Qin, J. Wu, Visual saliency maps can apply to facial expression recognition, 2018, arXiv e-prints arXiv:1811.04544.
[30] Y. Gan, J. Chen, L. Xu, Facial expression recognition boosted by soft label with a diverse ensemble, Pattern Recognit. Lett. 125 (2019) 105–112.
[31] B.-K. Kim, S.-Y. Dong, J. Roh, G. Kim, S.-Y. Lee, Fusing aligned and non-aligned face information for automatic affect recognition in the wild: a deep learning approach, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2016, pp. 48–57.
[32] B.-K. Kim, J. Roh, S.-Y. Dong, S.-Y. Lee, Hierarchical committee of deep convolutional neural networks for robust facial expression recognition, J. Multimodal User Interfaces 10 (2) (2016) 173–189.
[33] J. Shao, Y. Qian, Three convolutional neural network models for facial expression recognition in the wild, Neurocomputing 355 (2019) 82–92.
[34] G. Wen, Z. Hou, H. Li, D. Li, L. Jiang, E. Xun, Ensemble of deep neural networks with probability-based fusion for facial expression recognition, Cogn. Comput. 9 (5) (2017) 597–610.
[35] A. Agrawal, N. Mittal, Using CNN for facial expression recognition: a study of the effects of kernel size and number of filters on accuracy, Vis. Comput. 36 (2) (2020) 405–412.