# Facial Expression Recognition based on Transfer Learning from Deep Convolutional Networks

Mao Xu*, Wei Cheng*, Qian Zhao*, Li Ma*, Fang Xu†
*School of Computer Science and Engineering, Center for Robotics,
University of Electronic Science and Technology of China, Chengdu, China
†School of Science, Southwest Petroleum University, Chengdu, China

***Abstract*—It is well-known that deep models could extract robust and abstract features. We propose a efficient facial expression recognition model based on transfer features from deep convolutional networks (ConvNets). We train the deep ConvNets through the task of 1580-class face identification on the MSRA-CFW database and transfer high-level features from the trained deep model to recognize expression. To train and test the facial expression recognition model on a large scope, we built a facial expression database of seven basic emotion states and 2062 imbalanced samples depending on four facial expression databases (CK+, JAFFE, KDEF, Pain expressions form PICS). Compared with 50.65% recognition rate based on Gabor features with the seven-class SVM and 78.84% recognition rate based on distance features with the seven-class SVM, we achieve average 80.49% recognition rate with the seven-class SVM classifier on the self-built facial expression database. Considering occluded face in reality, we test our model in the occluded condition and demonstrate the model could keep its ability of classification in the small occlusion case. To increase the ability further, we improve the facial expression recognition model. The modified model merges high-level features transferred from two trained deep ConvNets of the same structure and the different training sets. The modified model obviously improves its ability of classification in the occluded condition and achieves average 81.50% accuracy on the self-built facial expression database.**

***Index Terms*—Facial expression recognition, Deep convolutional networks, Transfer learning**

## I. Introduction

Facial expression is one of the most natural and abundant communication modes among human beings. As one of the hottest points in facial expression research, facial expression recognition is always very challenging task since human beings' faces are presented in different poses, ages, illumination, and occlusions. To deal with the challenge, extracting robust features from face to avoid being interfered with intra-personal variations and natural environment factors is the kernel of facial expression recognition. It can traced back to early facial expression features extraction methods such as Gabor wavelets [1], AAM [2] and optical flow [3]. For example, optical flow expresses facial emotion features through making use of optical flow computation to identify the direction of nonrigid and rigid motions that are caused by facial expression. More recent studies have focused on 3D facial expression features extraction to obtain richer and more robust facial expression features. For example, 3D model-based features extraction [4] utilized 3D facial geometric shapes to compute shape changes caused by facial expression. However, features they extract perform less well in complex conditions (e.g., complex illumination, different facial poses, kinds of facial occlusions).

Recently, deep models such as deep belief nets and deep convolutional networks have let us have an sight into the effect on extracting robust and abstract features [5], [6], [7] and some deep models are used for facial expression recognition [8]. Susskind et al. [8] learned deep belief nets without supervision for recognizing facial action units [9] which is a kind of descriptions of facial emotion, and they demonstrated features extracted by learned deep belief nets could easily accommodate different constraints in real expressive environment. Differently, in this paper we propose to utilize transfer features from deep convolutional networks to recognize facial expression, and deep convolutional networks are more suitable for classification than deep belief nets.

We propose a effective facial expression recognition model based on transfer features from deep convolutional networks (ConvNets) for face identification. A high-level illustration of our facial expression recognition transfer learning is showed in Figure 1. The deep ConvNets is trained through performing the complex task of 1580-class face identification on the facial MSRA-CFW database [10]. 120-dimensional high-level features from trained deep ConvNets could express rich information about face through four-time convolutional features extraction and one-time fully-connecting features extraction. The facial expression recognition model transfers high-level features to classify facial patch into one of six basic emotions and neutral emotion with the seven-class SVM classifier. Compared with 78.84% accuracy based on distance features [11] about facial landmarks [12] with SVM and 50.65% accuracy based on Gabor features [13] after PCA processing with SVM, we could achieve average 80.49% accuracy on the self-built facial expression database. The database merges the CK+ facial expression database [14], the JAFFE facial expression database [15], the Karolinska Directed Emotional Faces (KDEF) [16], and the Pain expressions set from Psychological Image Collection at Stirling (PICS) [17]. As features deep ConvNets extract are robust to occlusion, we test our model in the condition of occluded face and demonstrate it could keep the ability of classifying emotion in the small occlusion case. To increase the ability in the occluded condition further, we improve the facial expression recognition model. The modified model merges 120-dimensional high-level features transferred from two trained deep ConvNets. The structures

978-1-4673-7679-2/15/$31.00 ©2015 IEEE

of two deep ConvNets are the same as deep ConvNets for face identification. One net is trained on the MSRA-CFW database, and the other is trained on MSRA-CFW database with additive occlusion samples. The modified model has obviously improved its classification ability in the occluded condition and it achieves average 81.50% accuracy on the self-built facial expression database.

This paper designs a efficient facial expression recognition model through a new idea of translate learning from deep ConvNets to extract robust features for facial expression recognition, and offers the new hybrid deep ConvNets to increase the robustness of transfer features from deep models to occlusion. The rest of this paper is organized as follows. Section II introduces related work of facial expression recognition and deep learning. In Section III, we propose facial expression recognition model based on features transferred from trained deep ConvNets for face identification and the modified model for occluded face. Experiments show effectiveness of the two proposed models and discuss some interesting phenomenons in Section IV.

## II. Related work

Studies on facial expression recognition have been lasting for three decades since 1970s. Paul Ekman et al. [18] postulated six cross-cultural basic emotions (anger, disgust, fear, happiness, sadness, and surprise) from a psychology view, and developed Facial Action Coding System (FACS) to describe facial micro-expression [9]. These significant works constitute the research basis of facial expression recognition. Liking most of studies about facial expression recognition, our work also selects the six basic emotions and neutral emotion as our standard of facial expression classification.

For facial expression recognition system, there are three basic parts: face detection, facial features extraction and facial expression recognition. In face detection, most of face detection methods can detect only frontal and near-frontal views of face. Hisele et al. [19] proposed a trainable system based on component for detecting frontal and near-frontal views in still gray images. Viola and Jones [20] utilized a set of rectangle features to detect face in real time. Following the development of 3D technology, researchers have done some attempts on 3D face acquisition. I.Ira et al. [21] utilized a single facial image of low resolution to reconstruct and acquire 3D face. About facial features extraction, three kinds of features (geometric features, appearance features and hybrid features of geometric and appearance features) are extracted for recognizing facial expression. For example, some studies used Active Shape Models to extract geometric features such as measurements among coordinates of landmarks on the face [11]. Some others studies used Gabor features [22] and local binary patterns as appearance features. A part of the remaining studies utilized the Active Appearance Model (AAM) [14] which tracks facial deformation and captures the shape of face to extract hybrid features. In facial expression recognition, there are different methods depending on spatial or spatio-temporal information. For pictures and single frames

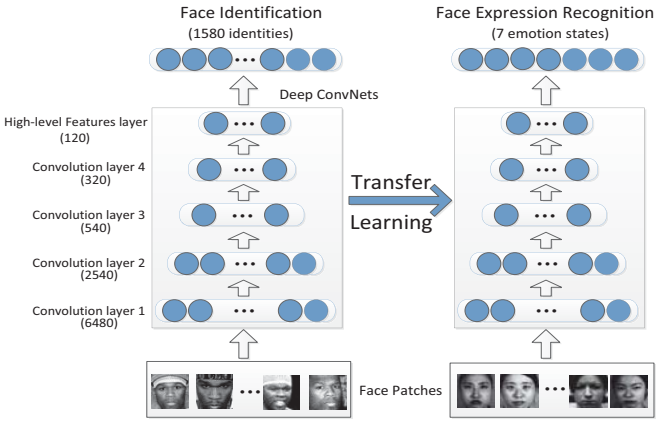


Fig. 1. A high-level illustration of facial expression recognition transfer learning

in videos, only spatial information could be used to classify emotion with diverse classifies such as Neutral Network, rule-based classifiers, Support Vector Machine (SVM) [14], [23], Bayesian Network (BN). For sequences in videos, spatial-temporal information could be utilized to classify expression with Hidden Markov Models (HMM) which is one of the most popular spatial-temporal approaches and works well on facial expression recognition [24]. Due to lack robust features, most of facial expression recognition models work poorly in the complex environment. Our work mainly focus on extracting robust features to cope with the complex environment.

In recent years, deep learning arouses academia and industrial attentions due to its magic in computer vision. Susskind et al. [8] took advantage of learned deep belief nets to classify facial action units in realistic face images. Krizhevsky et al. [5] used deep convolutional neural network to classify the 1.2 million images in the ImageNet LSVRC-2010 contest into 1000 different classes and achieved the inconceivably higher accuracy than the temporal state-of-the-art. Sun et al. [6] designed Deep hidden IDentity features (DeepID) with deep convolutional networks (ConvNets) to recognize about 1000 face identities on LFW database and achieved 97.45% vertification accuracy with only weakly aligned faces. Zhang et al. [7] utilized deep multi-task learning for detecting facial landmarks and obtained mean error 8.0% on the AFLW database which outperforms all the state-of-the-art methods (RCPR, TSPM, CDM, Luxand, and SDM). Our work are taking advantage of deep models to extract robust facial features and translate them to recognize facial emotions.

## III. Facial expression recognition model and Modified model for occluded face

### A. Deep ConvNets for face identification

Our deep ConvNets are composed by four convolutional layers with max-pooling to extract features level by level, the fully-connected high-level features layer and softmax output

layer predicting identity classes. The detail structure of deep ConvNets is showed in Figure 2. Input image is $39 \times 39$ gray facial patch which is extracted from face image on the MSRA-CFW database [10] by Viola-Jones face detection algorithm [20]. Number of features at every layer of net decreases step by step. Finally, the high-level features layer is fixed to 120 features which could express rich information of face. The last softmax output layer fully connecting high-level features predicts one of 1580 identity classes. The convolution operation in our deep model is described as:

$$y^j = \max\left(0, \sum_i x^i * k^{i,j} + b^j\right) \tag{1}$$

where $x^i$ and $y^j$ denote the $i$-th input map and the $j$-th output map respectively. $*$ notation indicates convolution, and $k^{i,j}$ is described as the kernel convolution connecting the $i$-th input map $x^i$ and the $j$-th output map $y^j$. $b^j$ represents the bias of the $j$-th output map. Active function of the convolution operation adopts the ReLU nonlinearity function ($f(x) = \max(0, x)$) which works better than sigmoid active functions [5]. To learn different regional features, weights in every layer of our deep ConvNets are locally shared. Max-pooling is expressed as:

$$y^j_{a,b} = \max_{0 \le m,n \le s} x^j_{a \cdot s+n, b \cdot s+m} \tag{2}$$

where each value in the $j$-th output map $y^j$ pools over the $s \times s$ non-overlapping region in the j-th input map $x^j$.

The high-level features layer is fully connecting to the fourth convolution layer (after ReLU nonlinearity active function). The part task is described as:

$$y^j = \sum_i x^i \times k^{i,j} + b^j \tag{3}$$

where $x^i$ and $y^j$ donate the $i$-th output value of the fourth convolution layer and the $j$-th feature in the high-level features layer respectively. $\times$ notation indicates the ordinary product. $k^{i,j}$ indicates the weight between the $i$-th output value and the $j$-th feature, and $b^j$ donates the bias of the $j$-th feature.

The output of our deep ConvNets is an 1580-way (1580 identities) softmax which predicts 1580-way probability distribution of input facial patch to identify face. The probability distribution function are expressed as:

$$y_i = \frac{exp(x_i)}{\sum_{i=1}^{1580} exp(x_i)} \tag{4}$$

where $x_i$ ($x_i = \sum_{j=1}^{120} z_j \times k_{j,i} + b_j$) linearly connects high-level features $Z$ of the last hidden layers, and $y_i$ denotes the $i$-th probability of 1580 classes. The whole deep ConvNets is learned by reducing the value of $\log y_n$ with the $n$-th sample, back-propagated by stochastic gradient descent.

### B. Facial expression recognition

After the deep ConvNets are adequately trained, we adopt the multiclass Support Vector Machine (SVM) and 120-dimensional high-level features transferred from trained deep

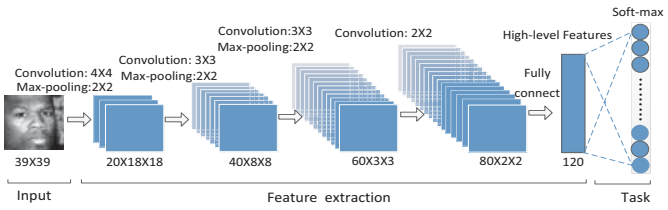


Fig. 2. Structure of deep ConvNets for face identification

ConvNets to classify seven emotion states (six basic emotions and neutral emotion). The training data of the multi-class SVM consist of $(x_1, y_1), (x_2, y_2), ..., (x_N, y_N)$ where $x_i$ denotes 120-dimensional feature vector (after normalization and whitening) and $y_i$ ($y_i \in (1, ..., 7)$) expresses the label of faical expression correspondingly. Multiclass SVM establishes seven expression functions in which each is described as $w_k \phi(x_i) + b_k$ to separates training vectors of different classes. Solving $w_k$ and $b_k$ is equivalent to the optimization problem [25] which minimizes the objective function

$$\min_{w,b,\xi} C \sum_{i=1}^{N} \sum_{k \ne l_i} \xi_i^k + \frac{1}{2} \sum_{k=1}^{7} w_k w_k^T \tag{5}$$

subject to the constraints

$$\begin{aligned} & w_{l_i} \phi(x_i) + b_{l_i} \ge w_k \phi(x_i) + b_k + 2 - \xi_i^k \\ & \xi_i^k \ge 0, i = 1, 2, ..., N, k \in 1, .., 7/l_i \end{aligned} \tag{6}$$

where $\phi(x_i)$ expresses the kernel function mapping training vector $x_i$ to a higher dimensional space which is linearly or near linearly separable. C denotes the penalty factor which penalizes the training error, and $\xi_i^k = \left[\xi_1^1, \xi_1^2, ..., \xi_N^6, \xi_N^7\right]^T$ indicates slack variable vector. $b = [b_1, .., b_7]^k$ expresses the bias vector. Lastly, the decision function is formulated as

$$f(x) = \underset{k=1,2,...,7}{argmax}\ w_k \phi(x) + b_k \tag{7}$$

where $x$ and $f(x)$ express the input feature vector and the output facial expression label respectively.

In our work, the $\phi$ kernel function selects the Radial Basis Function (RBF) which has been very widely used for classifications. After the seven-class SVM model has learned, our model could been used to recognize facial emotion.

### C. Model refinement for occluded face

Although features from deep models are robust to occlusion, they perform not well on a little bigger occlusion. To increase their robustness to occlusion further, we improve facial expression recognition model by merging high-level features of two trained deep ConvNets with the same structure. The structure of the improved model is showed as Figure 3. The modified model merges high-level features transferred from two trained deep ConvNets and utilizes the same seven-class SVM classifier with 240-dimensional high-level combined features to classify facial expression patch to one of seven emotion states. The structures of two deep ConvNets are

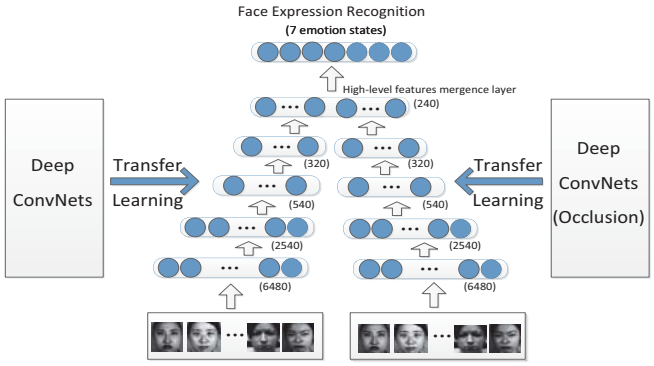


Fig. 3. Structure of the modified facial expression recognition model for occluded face

Fig. 4. Face patches on self-built facial expression database

the same as deep ConvNets for face identification, and the two deep ConvNets perform the same task of 1580-class face identification. Differently, one is trained on the MSRA-CFW database, and the other is trained on the MSRA-CFW database with additive occluded samples. The modified model is similarly trained on the self-built facial expression database.

## IV. Experiments

We evaluate our model on the self-built facial expression database which merges the CK+ facial expression database [14], the JAFFE facial expression database [15], the Karolinska Directed Emotional Faces [16], and the Pain expressions set from Psychological Image Collection at Stirling [17]. On the CK+ database, the first frame and the last frame from every video sequence are definitely marked emotion labels and the frames marked by seven emotion labels are selected. On the JAFFE facial expression database, all expressive face images are selected. On the Karolinska Directed Emotional Faces, the frontal face images from seven emotions are selected. On the Pain expressions set from Psychological Image Collection at Stirling, the frontal face images from seven emotions are also selected. The self-built facial expression database obtains 2062 face patches of seven expression states (neutral: 535, anger: 247, disguest: 261, fear: 231, happy: 272, sadness: 229, surprise 287) from the above selected face images through Viola-Jones face detection algorithm [20]. Some face patches on our self-built facial expression database are shown in Figure 2.

For facial features learning, the deep ConvNets model is trained on the MSRA-CFW database [10] which contains 202792 face images of 1583 celebrities from the Internet. Before training the deep model, we process the database through Viola-Jones face detection algorithm and artificial selection since the Viola-Jones face detection algorithm detects the area of not face on some images. We finally obtain 181730 face patches of 1580 celebrities. We randomly choose 80% face patches every celebrity to learn features and the rest to generate the validation data. The higher top-1 validation set accuracy the validation set gets, the better features learn. Finally, we achieves the top-1 validate set average accuracy 40.83% for 1580-class face identification with Caffe framework [26] which is an extensively applied framework for deep learning.

Due to the limit and imbalance of samples on the self-built database, we utilize the 5-fold cross-validation of 10 times to validate our model and select the mean of recognition rates of seven emotions to evaluate it. At the same time, we select the trained classifier which performs best on the whole self-built database from the cross-validation to construct the whole models of facial expression recognition. In training the seven-class SVM classifier, we utilize the libSVM tool [27] and set the penalty of the neutral expression is 2 and the others are all 10.

For the modified model, we train two deep ConvNets like the above training method. Differently, one deep ConvNets is trained on the MSRA-CFW database and other is trained on the MSRA-CFW database with additive occlusion samples. On the MSRA-CFW database with additive occlusion samples, each face patch on the MSRA-CFW database produces two pieces of occluded face patches through randomly setting the location of occluded block and randomly selecting the occluded ratio (the area scale of occluded block in the whole face patch) of occluded block from 0 to 0.2. The occluded database finally contains 363460 occluded face patches and 181730 face patches of 1580 celebrities. The seven-class SVM classifier is trained on the self-built facial expression database like the above method.

To show transfer features from deep ConvNets are robust to occlusion, we construct the validation sets of different occluded ratio from 0 to 0.5 and test our model on the validation sets. At the same time, we also test our modified model on the occluded validation sets. Facial expression recognition model and the modified model described in the previous segments remain unchange.

TABLE I
COMPARISON OF THREE FACIAL EXPRESSION RECOGNITION MODELS

| Method | Mean Accuracy (%) |
|---|---|
| Transfer features + SVM | 80.49 |
| Distance features + SVM | 78.84 |
| Gabor features +PCA +SVM | 50.65 |

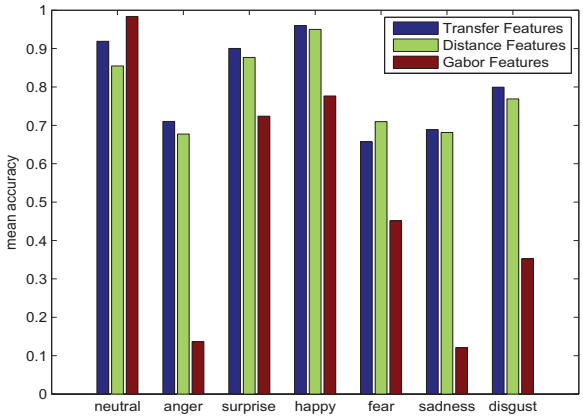


Fig. 5. Performance of three facial expression recognition models in seven emotion states

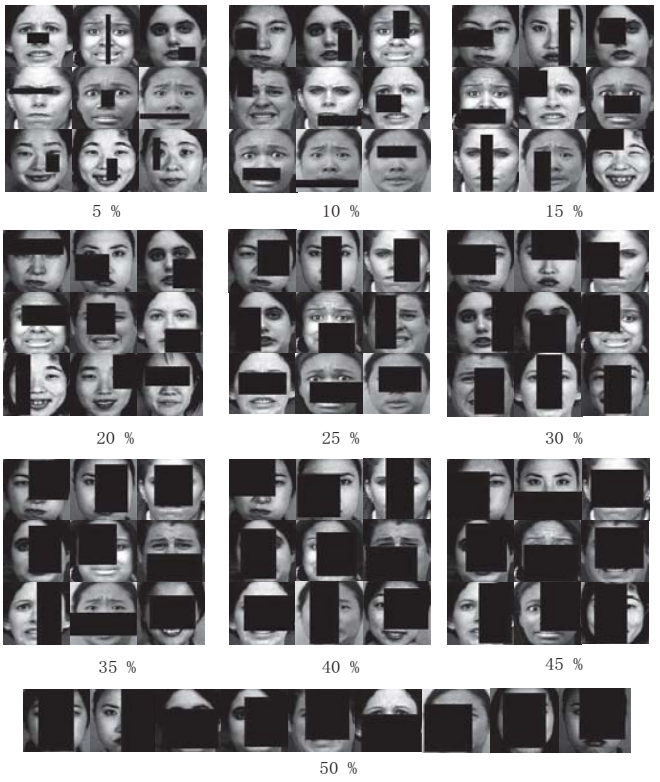


Fig. 6. Occluded face patches of different occluded rates

### A. Model comparison

The average recognition results based on three kinds of facial expression features and the seven-class SVM classifier on the validation set are listed in Table 1. Compared with the other two models, the model based on transfer features from the trained deep model has shown potential ability on the task of facial expression recognition, and it achieves average 80.49% accuracy. It demonstrates that transfer features from the deep ConvNets are easier to be distinguished than distance features and Gabor features. By observing performance of the three models in seven emotion states (Figure 5), we find that the model based on transfer learning keep the better ability of classification in all seven emotion states and performs outstandingly in emotion states (happy, surprise) which human beings could easily recognize.

### B. Model evaluation in occluded condition

In the real world, faces are sometimes occluded by objects (e.g. glasses, mask, hair and so on) and they need to been analyzed for serving human beings more comfortably. To evaluate the performance of the proposed model in the occluded condition, we construct the environment of face occlusion in which we randomly select the square block of area ratio $M$ $(5\%, 10\%, 15\%, 20\%, 25\%, 30\%, 35\%, 40\%, 45\%, 50\%)$ from facial patches on the self-built facial expression database and set it black. Finally, we obtain ten occluded facial expression datasets and some occluded face patches every occluded ratio are showed in Figure 6. We select all occluded face patches to generate the ten validate sets and observe the classification ability of our model on ten validate sets. At the same time, the model based on distance features and the model based on Gabor features are also tested on the validate sets. As showed in Figure 7, the model based on transfer features could keep its classification ability in the condition of the small occlusion. The phenomenon demonstrates the features from translate learning for deep ConvNets are robust to small occlusion. Liking the model based on translate features, the model based on distance features performs well since facial landmarks detection algorithm [12] could work normally in the small occlusion case.

To improve the classification ability of facial expression recognition model in the occluded condition, we propose the modified model in Section III. We train and test the modified model on the self-built facial expression database liking the above method, and test it on the ten validate sets of different occluded ratios. At the same time, we train and test the model based on transfer features from the deep ConvNets which is only trained on the MSRA-CFW database with additive occlusion samples, and test it on the ten validate sets. The result is illustrated in Figure 8. We find the modified model obviously improves the ability of classification on different occluded ratio, and it could achieve average 81.50% accuracy which is higher than the above proposed model on the self-built database due to additive information from the model based

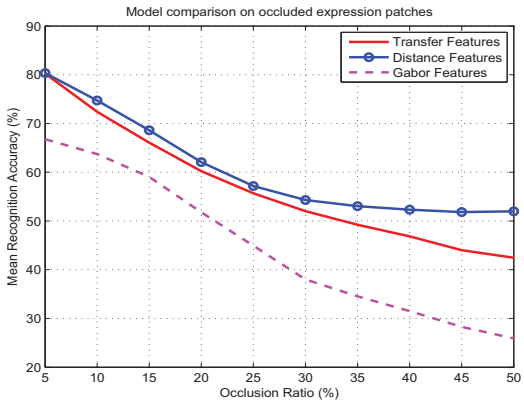


Fig. 7. Comparison model on occluded face patches

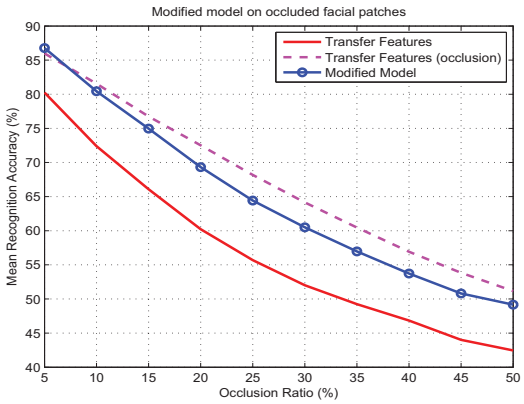


Fig. 8. Modified model on occluded face patches

on transfer features from the deep model with occlusion train. The model based on transfer features from the trained deep ConvNets with only occluded training only obtains 77.04% accuracy on the self-built database. At the same time, we also find the modified model perform a little worse than the model based on transfer features from the trained deep ConvNets with only occluded training since its robustness is not enough for occlusion and their features express interferential information for classifying facial emotion in the occlusion case.

## V. Conclusion

We have proposed a efficient facial expression recognition model based on robust transfer features from trained deep ConvNets. The deep ConvNets have been trained through the task of face identification on the MSRA-CFW database. To train and test the facial expression recognition model on larger scope, we have built a facial expression database which merges four widely used facial expression databases. Compared with 78.84% accuracy based on distance features and 50.65% accuracy based on Gabor features, we have achieved 80.49% facial expression recognition accuracy on the self-built facial expression database. At the same time, we have demonstrated the model could keep its classification ability in the condition of small facial occlusion. To increase its classification ability, we have improved the model. The modified model has obviously improved the ability of classification in the occluded case, and it have achieved average 81.50% accuracy on the self-built expression database. Future work will explore the models for different facial poses and real-time recognition.

## Acknowledgment

This work was supported by the National Natural Science Foundation of China (No. 61202045 and 60803028) and the State Scholarship Fund of China (No. 201406075055).

## References

[1] J. G. Daugman, "Complete discrete 2-d gabor transforms by neural networks for image analysis and compression," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 36, no. 7, pp. 1169–1179, 1988.

[2] J. Xiao, S. Baker, I. Matthews, and T. Kanade, "Real-time combined 2d+ 3d active appearance models," in *Computer Vision and Pattern Recognition (CVPR), 2004 IEEE Conference on*. IEEE, 2004, pp. 535–542.

[3] I. A. Essa and A. P. Pentland, "Coding, analysis, interpretation, and recognition of facial expressions," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, no. 7, pp. 757–763, 1997.

[4] J. Wang, L. Yin, X. Wei, and Y. Sun, "3d facial expression recognition based on primitive surface feature distribution," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 2. IEEE, 2006, pp. 1399–1406.

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[6] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*. IEEE, 2014, pp. 1891–1898.

[7] Z. Zhang, P. Luo, C. C. Loy, and X. Tang, "Facial landmark detection by deep multi-task learning," in *European Conference on Computer Vision (ECCV)*. Springer, 2014, pp. 94–108.

[8] J. M. Susskind, G. E. Hinton, J. R. Movellan, and A. K. Anderson, "Generating facial expressions with deep belief nets," *Affective Computing, Emotion Modelling, Synthesis and Recognition*, pp. 421–440, 2008.

[9] E. Friesen and P. Ekman, "Facial action coding system: a technique for the measurement of facial movement," *Palo Alto*, 1978.

[10] X. Zhang, L. Zhang, X.-J. Wang, and H.-Y. Shum, "Finding celebrities in billions of web images," *Multimedia, IEEE Transactions on*, vol. 14, no. 4, pp. 995–1007, 2012.

[11] M. Suk and B. Prabhakaran, "Real-time mobile facial expression recognition system–a case study," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2014 IEEE Conference on*. IEEE, 2014, pp. 132–137.

[12] X. Xiong and F. De la Torre, "Supervised descent method and its applications to face alignment," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013, pp. 532–539.

[13] W. Liu and Z. Wang, "Facial expression recognition based on fusion of multiple gabor features," in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, vol. 3. IEEE, 2006, pp. 536–539.

[14] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*. IEEE, 2010, pp. 94–101.

[15] M. J. Lyons, J. Budynek, and S. Akamatsu, "Automatic classification of single facial images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 12, pp. 1357–1362, 1999.

[16] D. Lundqvist, A. Flykt, and A. Öhman, "The karolinska directed emotional faces (kdef)," *CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet*, pp. 91–630, 1998.

[17] P. Hancock, "Psychological image collection at stirling (pics)," *Web address: http://pics. psych. stir. ac. uk*, 2008.
[18] P. Ekman, "Universals and cultural differences in facial expressions of emotion." in *Nebraska symposium on motivation*. University of Nebraska Press, 1971.
[19] B. Heisele, T. Serre, M. Pontil, and T. Poggio, "Component-based face detection," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1. IEEE, 2001, pp. I–657.
[20] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.
[21] I. Kemelmacher-Shlizerman and R. Basri, "3d face reconstruction from a single image using a single reference face shape," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 2, pp. 394–405, 2011.
[22] W. Gu, C. Xiang, Y. Venkatesh, D. Huang, and H. Lin, "Facial expression recognition using radial encoding of local gabor features and classifier synthesis," *Pattern Recognition*, vol. 45, no. 1, pp. 80–91, 2012.
[23] I. Kotsia and I. Pitas, "Facial expression recognition in image sequences using geometric deformation features and support vector machines," *Image Processing, IEEE Transactions on*, vol. 16, no. 1, pp. 172–187, 2007.
[24] G. Sandbach, S. Zafeiriou, M. Pantic, and D. Rueckert, "Recognition of 3d facial expression dynamics," *Image and Vision Computing*, vol. 30, no. 10, pp. 762–773, 2012.
[25] J. Weston and C. Watkins, "Multi-class support vector machines," Citeseer, Tech. Rep., 1998.
[26] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," *arXiv preprint arXiv:1408.5093*, 2014.
[27] C.-C. Chang and C.-J. Lin, "Libsvm: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, p. 27, 2011.