6th International Conference On Advances In Computing & Communications, ICACC 2016, 6-8 September 2016, Cochin, India

Automatic Facial Expression Recognition Using DCNN

Veena Mayya, Radhika M. Pai*, Manohara Pai M. M.

Department of Information & Communication Technology, Manipal Institute of Technology, Manipal University, Karnataka-576104, India

Abstract

Face depicts a wide range of information about identity, age, sex, race as well as emotional and mental state. Facial expressions play crucial role in social interactions and commonly used in the behavioral interpretation of emotions. Automatic facial expression recognition is one of the interesting and challenging problem in computer vision due to its potential applications such as Human Computer Interaction(HCI), behavioral science, video games etc.

In this paper, a novel method for automatically recognizing facial expressions using Deep Convolutional Neural Network(DCNN) features is proposed. The proposed model focuses on recognizing the facial expressions of an individual from a single image. The feature extraction time is significantly reduced due to the usage of general purpose graphic processing unit (GPGPU). From an evaluation on two publicly available facial expression datasets, we have found that using DCNN features, we can achieve the state-of-the-art recognition rate.

© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license
(http://creativecommons.org/licenses/by-nc-nd/4.0/).
Peer-review under responsibility of the Organizing Committee of ICACC 2016

Keywords: Facial expression recognition;Computer Vision;Machine Learning;Confusion Matrix;Support vector machine; JAFFE; CK+; Deep Convolutional Neural Network;

1. Introduction

Facial expression is an important part of nonverbal communication. Human expression recognition is influenced by certain context. When a subject is being investigated, the investigator might be diverted by the subject's voice tone or argument and may forget to keep track of the facial expressions. Automatic facial expression recognition systems are exempt to such contextual interference. Such systems can be beneficial in many fields, like gaming applications, criminal interrogations, psychiatry, animations etc. State-of-art approaches attempt to recognize six basic facial expressions such as anger, disgust happiness, sadness, surprise and fear.

Facial expression recognition techniques are based on either appearance features or geometry features[1]. Geometric features are extracted from the shape of the face and its components such as the eyebrows, the mouth, the nose etc. Appearance features are extracted using the texture of the face caused by expression, such as furrows, wrinkles etc. In 1970s Paul Ekman and Wallace V. Friesen, developed Facial Action Coding System (FACS)[2] which is the most

* Corresponding author. Tel.: +91-08202925362;
  E-mail address: radhika.pai@manipal.edu

widely used method for describing and measuring facial behaviors. FACS is a system designed for human observers to describe changes in facial expression in terms of observable facial muscle actions known as facial action units or AUs. FACS is demonstrated to be a powerful means for detecting and measuring facial expressions and is recently used for feature extraction in combination with other techniques such as Dynamic Bayesian Network (DBN)[3] and Local Binary Pattern (LBP)[4]. Histograms of oriented gradients(HOG)[5], Scale Invariant Feature Transform (SIFT)[6], Local Binary Pattern (LBP)[7] are few state-of-art techniques for extracting facial features. Most of the above techniques use handcrafted features for facial expression recognition, and therefore require particular efforts both in terms of computation cost and programming effort.

In recent years, deep learning using convolution neural networks(CNNs) for feature extraction of image data is becoming more popular. Their popularity stems from their ability to extract good representations from image data. DCNN's computation intensive tasks can run on GPU, which results in high performance at very low power consumption. They have also yielded high performance for some of challenges such as the CNN based model proposed by Kim et al.[8]. CNN is extensively used for facial feature extraction for determining age[9], gender[10] etc.

2. Related Work

Several methods have been reported in the literature to automatically recognize facial expressions. Lucey et al.[11] manually labeled 68 facial points in key frames and used a gradient descent Active Appearance Model (AAM) to fit these points in the remaining frames. It may not be possible to obtain accurate key points in many practical situations. A study on using LBP for facial expression recognition is proposed by Shan et al.[12]. Here expression recognition is accomplished using support vector machine (SVM) classifiers with boosted-LBP features. In[12], authors manually labeled eye positions, which is not feasible in many practical cases. Computer Expression Recognition Toolbox (CERT) is proposed by Littlewort et al.[13]. CERT convolves through the registered face image with Gabor filters to extract the facial features and uses SVM and multivariate logistic regression (MLR) classifiers to recognize facial expressions. Lyons and Akamatsu[14] proposed a system for coding facial expressions with 2D Gabor wavelets for feature extraction, having clustering for classification. These methods also require particular efforts, both in terms of computation cost and programming effort.

Deep Convolution Neural Network (DCNN) framework is widely used for extraction of features from the images. DCNN uses several layers leading accurate feature learning. Here the prelearned features are used as filters and these filters convolves through the input image and produces the features which in turn are used by other layers of the network as discussed by Krizhevsky et al.[15]. Techniques based on convolutional neural networks have been proposed for facial expression recognition such as the model proposed by Kahou et al.[16]. But they extensively train the model with other facial dataset. Sébastien[17] used Deep Convolutional Activation Feature for Generic Visual Recognition(DeCAF)[18] for facial feature extraction that does not require extensive training, but DeCAF is too slow to use it for training even the small image dataset as it does not support GPU.

3. Proposed Method

In this work, automatic facial expression recognition using DCNN features is investigated. Two publicly available datasets CK+[11] and JAFFE[20] are used to carry out the experiment. Pre-processing step involves face detection for the above two datasets. The frontal faces are detected and cropped using OpenCV[21]. Then facial features are extracted using the DCNN framework. Algorithm 1 illustrates the steps for recognizing facial expressions. Subsection 3.1 describes the facial feature extraction using Convolutional Architecture for Fast Feature Embedding(Caffe) framework[19].

3.1. Feature Extraction using Caffe

Feature extraction is performed using Caffe on Graphics Processing Unit (GPU). The convolution neural network architecture, which is used for ImageNet[15] object detection is used to extract facial features. ImageNet uses eight learned layers which includes five convolutional and three fully connected layers for object detection. In this work, features are extracted using only first five layers. These layers are combination of convolution, Rectified Linear

---

Algorithm 1 Algorithm for recognizing facial expressions using DCNN

.

    procedure RECOGNIZE-FACIAL-EXPRESSIONS
        for all image (*i*), depicting facial expressions ∈ face dataset **do**
            convert the image (*i*) to gray-scale
            detect frontal face in (*i*) and crop only the face (*c*)
            extract *POOL5* ($256 \times 6 \times 6$) features for cropped face (*c*) using DCNN
            copy predefined facial expression label for each image(*i*) as per the input dataset
    use the resulting *POOL5* vector of dimension $9216(256 \times 6 \times 6)$ for tenfold and leave-one-out cross validation with SVM classifier to recognize facial expression
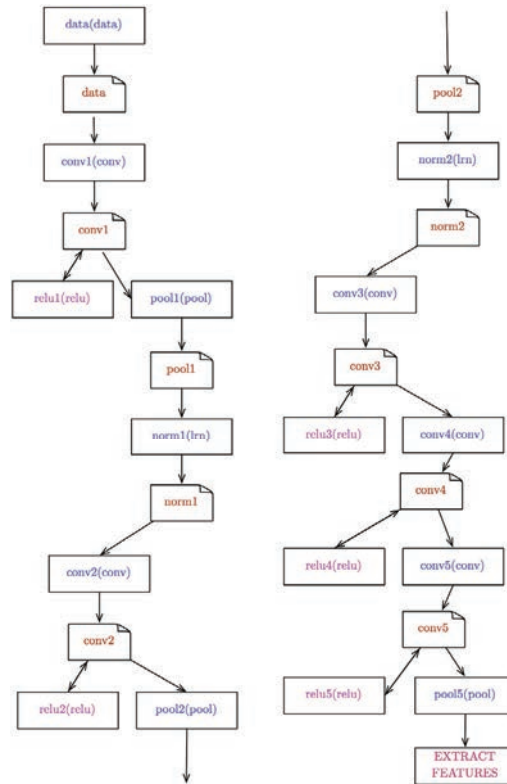

Fig. 1: The Caffe ImageNet pre-trained model

Units(RELU), Local Response Normalization(LRN) and pooling operations. The operations are repeated, to form different layers as per the model shown in Figure 1. The features extracted after pooling operation at fifth layer (POOL5) are used for recognizing facial expressions. This is the layer where features are most visible and also the POOL5 feature length($6 \times 6 \times 256$) is feasible to use in classifiers to recognize facial expressions.

The first layer is the convolution layer. This model uses 96 filters with size $11 \times 11 \times 3$. This layer extracts the low-level edge features. Figure 2 shows the sample output after first convolution filters being applied on a face image.

The second layer is ReLU (Rectified Linear Units). This layer increases the nonlinear properties of network. For any given input value x, ReLU is defined by ($f(x) = max(0, x)$).

Pooling is the next layer. In pooling layer, small rectangular blocks from the previous layer are taken and subsampled to produce a single output from that block. In this model, max-pooling is used. 3×3 rectangular blocks with an interval of 2 pixels are used for max-pooling. Figure 3a shows the facial features before performing the max-pooling
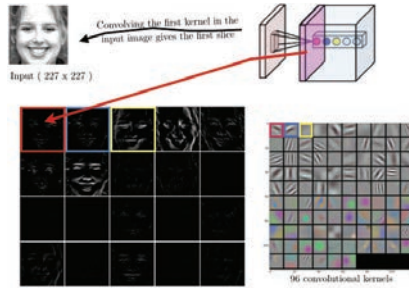
Fig. 2: Output of first convolution layer applied on face image
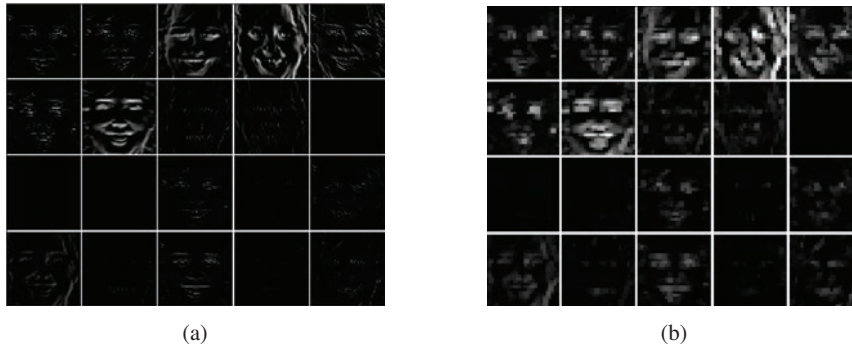

| (a) | (b) |

Fig. 3: Extracted facial features (a) before applying max-pooling; (b) after applying max-pooling

and Figure 3b depicts the facial features after performing max-pooling. It can been seen that after pooling, facial edges are retained even though the dimension of the image is reduced.

The fourth layer is LRN (Local Response Normalization). This is the brightness normalization layer given by $(1 + (\alpha/n) \sum_i x_i{}^{2\beta})$, where α and β are tuning parameters with the default values 1 and 5 respectively. This layer normalizes the brightness, so that the relevant features are more visible and irrelevant features are reduced.

Once the features are extracted, SVM is used for classification. Best kernel for SVM is estimated using grid search estimators[22]. Grid search exhaustively considers all parameter combinations on a dataset and the best combination is retained. Here 50% of the dataset is used for training and 50% for testing. The accuracy is evaluated and the estimator provided good accuracy for SVM linear kernel over other kernels. So SVM one-vs-one (SVC) and one-vs-all(LinerSVC) classifiers are used to recognize facial expressions. In SVM classification, parameter C controls the trade off between errors on training data and margin maximization (C = ∞ leads to hard margin)[23]. Best C parameter is chosen for the datasets after evaluating the accuracy for all C ranging from 1e-3 to 10 (1e-3, 1e-2 . . . 10).

Leave-one-out and tenfold cross validation methods are used to estimate the performance. In case of tenfold cross validation, the dataset is divided into ten sets. One set is used for testing and remaining sets are used for training the model. Confusion matrix is used to visualize the performance of a classifier as it identifies the nature of the classification errors, as well as their quantities. Each row of the confusion matrix represents the instances in a actual class while each column represents the instances in predicted class. The diagonal elements in the confusion matrix represent the number of correctly recognized expressions for each class. The off-diagonal elements represent the unrecognized expressions or the recognition error. Given the confusion matrix, the accuracy is calculated using Equation 1[24]. Accuracy is also measured using $F1$ score (F measure) using precision and recall as given by Equation 2. Python machine learning tools[22] are used to implement the proposed model.

$$Accuracy = \frac{\sum(DiagonalElements)}{\sum(AllElements)} \tag{1}$$

$$F1 = 2 \times \frac{precesion \times recall}{precesion + recall} \tag{2}$$
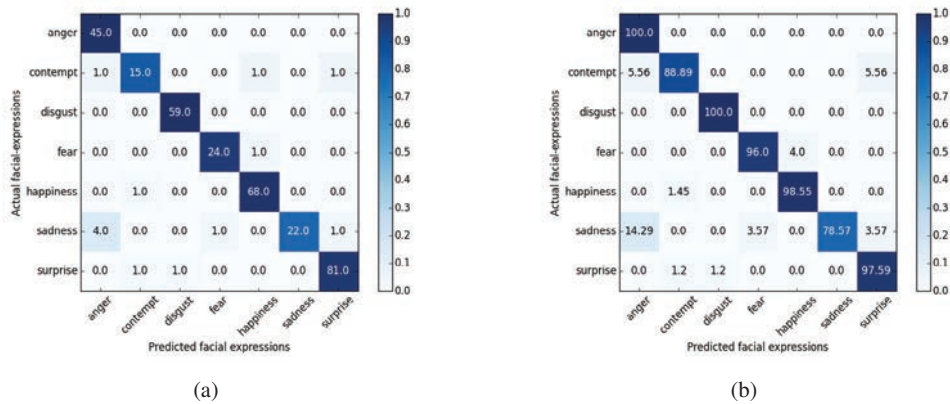
Fig. 4: Confusion matrix for CK+ dataset (a) Unnormalized; (b) Row normalized

4. Experiments and Results

Two publicly available facial expression datasets were used to evaluate the proposed method. CK+[11] dataset includes 327 video sequences acted out by 118 participants. Each sequence is labeled with one of the following emotions: anger, contempt, disgust, fear, happiness, sadness and surprise. The sequence consists of approximately 10 to 30 frames, only last frame is used to recognize facial expression. Every sequence starts with the neutral emotion and the last frame depicts the emotion which is for the corresponding label.

Japanese Female Facial Expression (JAFFE)[20] consists 213 facial expressions acted by ten subjects. It consists of 30 anger, 29 disgust, 32 fear, 31 happiness, 30 neutral, 31 sadness and 30 surprise expressions.

All 327 sequences of the CK+ dataset and 213 images from JAFFE dataset are used for evaluating the proposed model. Tesla K20Xm GPU system with compute version 3.5 is used to carry out the experiment.

4.1. Cross Validation for CK+ Dataset

Leave-one-subject-out cross validation method is used to test the model. Figure 4a shows confusion matrix for best SVM classification with the C value equal to $1e1$. The numbers in the confusion matrix represents the number of images for a particular facial expression. For example in Figure 4a, out of 18 contempt facial expression samples, 15 are rightly identified as contempt and 3 are wrongly identified as anger, happiness and surprise. In total, the facial expressions are correctly recognized for 314 images out of 327, with an accuracy of 96.02%. The normalized confusion matrix is shown in Figure 4b. It can be seen that most of the expressions are recognized correctly except for the sequences that contain sad expressions. This might be because the samples available for sadness expression are less and few of the sequences- for example S131.003, S026.002 etc. are not obvious to be recognized as sad expressions.

Since few of the state-of-art literature use only six universal expressions excluding contempt, the recognition is performed using six universal expressions (without contempt). The confusion matrix as shown in Figure 5a depicts the best mapping of true and predicted expressions with C=10.0 for SVM one-vs-all classifier. The accuracy of facial expressions recognition for the proposed method for six universal expressions is 97%. Figure 5b provides the normalized confusion matrix for six universal expressions. This is 13% more accurate than the baseline results 83.33%[11] .

4.2. Cross Validation for JAFFE Dataset

All 213 facial expressions images that are available with the JAFFE dataset were used to evaluate the proposed method. The baseline similarity rank using Gabor[20] filter is 0.679. From the baseline literature[20], it was not clear regarding the accuracy of the model, so evaluation of the proposed model is compared with the model proposed by Lyons et. al[14]. They extracted the features using 2D Gabor wavelet representation and performed ten-fold cross
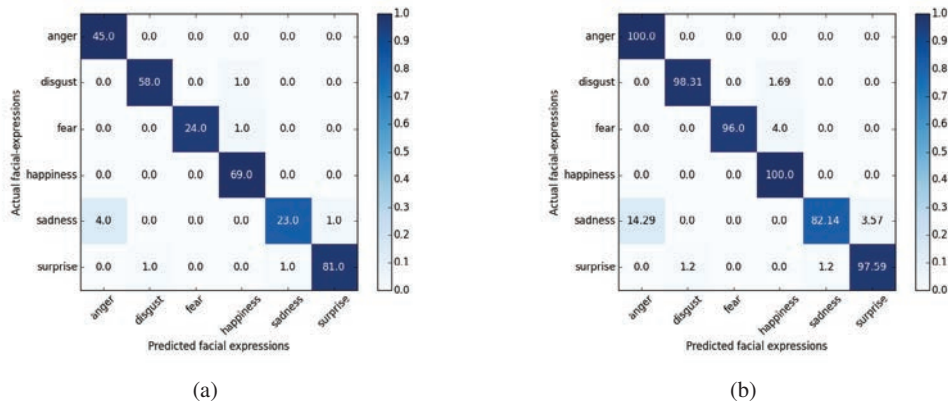
(a)   (b)

Fig. 5: Confusion matrix for CK+ dataset for only six universal expressions (a) Unnormalized; (b) Row normalized

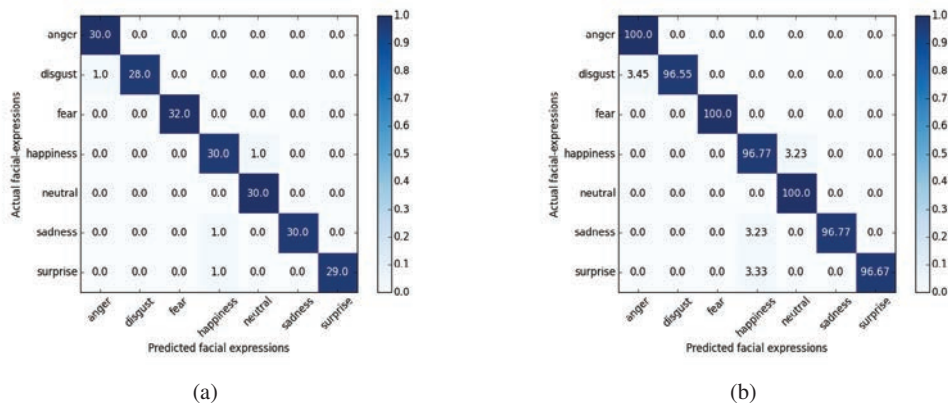(a)   (b)

Fig. 6: Confusion matrix for JAFFE dataset (a) Unnormalized; (b) Row normalized

validation which resulted 92.00% recognition accuracy. The proposed model with DCNN features and ten-fold cross validation results in 98.12% accuracy. Here the dataset is divided into ten folds using Python's StratifiedKFold machine learning tool by setting the shuffling parameter. Figure 6a depicts the confusion matrix for JAFFE dataset. The results show that out of 213 facial expressions images, the proposed model correctly recognizes expressions for 209 images with an accuracy of 98.12%. Figure 6b depicts the normalized confusion matrix. It can be seen that the accuracy for most of expressions is near 100%. Leave-one-sample-out cross validation is also performed and the no change in accuracy is found, i.e the results show the same confusion matrix as shown in Figure 6a.

Figure 7a depicts the Precision-Recall curve for the results of ten-fold cross validation. From the graph, it can be seen that recognition accuracy is 100% for fear and neutral expressions. There is slight deviation for happiness expression, as two of the expressions belonging to sadness and surprise are wrongly classified as happiness. Figure 7b provides the details of the facial expressions, for which the prediction mismatches with the corresponding labels provided in the dataset. It can be seen that the facial expressions match more with predicted classes rather than the corresponding actual labels provided in the dataset.

In information processing, F1 score (F Measure) is a measure of classifier's accuracy, as it considers both precision and recall. Table 1 summarizes F1 score for the proposed method.

Comparison of state-or-art recognition accuracy with the proposed model is shown in Table 2. It can be seen that proposed model provides the best recognition accuracy for JAFFE. For CK+, better results were obtained by other state-of-art literature. Methods proposed by Siddiqi et al.[25] and Mlakar et al.[26] results in better performance when neutral images are used along with seven basic expressions, but it can be noted that only few subjects (10, 106) were used for cross validation. For verifying the proposed method for CK+ dataset, all 118 subjects were considered
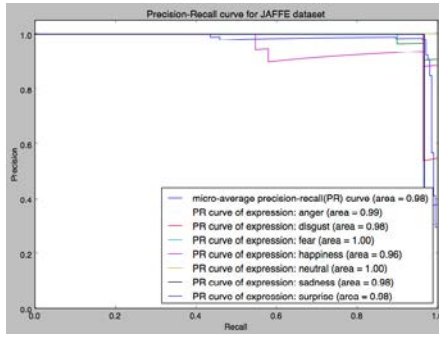
(a)

(b)

Fig. 7: (a) Precision-Recall curve for JAFFE dataset; (b) Misclassified facial expressions for JAFFE dataset

Table 1: Overall accuracy of the proposed method

Dataset	Cross validation method	F1 Score
CK+	Leave-one-subject-out(six classes)	0.9708
	Leave-one-subject-out(seven classes)	0.96024
JAFFE	Ten fold	0.9812
	Leave-one-sample-out	0.9812

Table 2: Comparison of recognition rate obtained by proposed model with state-of-art literature

Dataset	Method	Accuracy
CK+	Shan at el. (2009)(six classes) [12]	89.1
	Jeni at el. (2011) (six classes) [27]	96
	Proposed Method(six classes)	**97.08**
	Kahou et al. (2015) (seven classes) [16]	91.3
	Proposed Method(seven classes)	**96.02**
JAFFE	Lyons et al.(1999) [14]	92
	Zhao et al. (2011) [28]	81.6
	Zhang at el .(2011) [29]	92.93
	Mlakar at el. (2015) [26]	87.82
	Proposed Method(TenFold)	**98.12**
	Proposed Method(leave-one-sample-out)	**98.12**

with 327 sequences without neutral expression. The proposed method approximately requires 140-145ms to extract the facial features for an image on Tesla K20Xm GPU, which is significantly less compared to feature extraction techniques that uses only CPU (700-900 ms).

An application has been developed on Unix platform using Python machine learning tools and Caffe. This application recognizes the facial expressions from either an image or video sequence. Along with JAFFE, CK+ facial expression datasets, images from Google search are used to train the system using the above proposed model. Approximately 150 images are used for each facial expression class to train the model. The output of the classifier is saved and used for testing the recognition of facial expressions for unknown faces. For a video sequence the facial expression for every frame is recognized and result is shown using a bar graph. Figure 8 shows output for three facial expressions namely happy, sadness and surprise.

Fig. 8: Output of automatic facial expressions recognition module

5. Conclusion

Facial expressions convey the emotional state of an individual to the observers. An efficient and faster method to recognize the facial expressions is proposed in this paper. The facial features are extracted using deep convolutional neural network using Caffe on CUDA enabled GPU system. The proposed method is evaluated on two publicly available datasets and state-of-the-art results are achieved. Since GPU based Caffe module is used to conduct the experiment, the time required to extract a features is significantly reduced. The proposed model can be adopted to any generic facial expressions recognition dataset that either involves recognition in static images or video sequences. No retraining or extensive pre-processing techniques are required to adopt the proposed method for facial feature extraction. The future work involves exploring other DCNN pre-trained models such as GoogLeNet [30].

References

1. Tian, Y.L., Kanade, T., Cohn, J.F.. *Handbook of Face Recognition*; chap. Facial Expression Analysis. New York, NY: Springer New York. ISBN 978-0-387-27257-3; 2005, p. 247–275. URL: http://dx.doi.org/10.1007/0-387-27257-7_12. doi:10.1007/0-387-27257-7_12.
2. Ekman, P., Friesen, W.. *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Palo Alto: Consulting Psychologists Press; 1978.
3. Ko, K.E., Sim, K.B.. Emotion recognition in facial image sequences using a combination of aam with facs and dbn. In: *Proceedings of the Third International Conference on Intelligent Robotics and Applications - Volume Part I*; ICIRA'10. Berlin, Heidelberg: Springer-Verlag. ISBN 3-642-16583-4, 978-3-642-16583-2; 2010, p. 702–712. URL: http://dl.acm.org/citation.cfm?id=1939443.1939518.
4. Wang, L., Li, R.F., Wang, K., Chen, J.. Feature representation for facial expression recognition based on facs and lbp. *International Journal of Automation and Computing* 2015;**11**(5):459–468. URL: http://dx.doi.org/10.1007/s11633-014-0835-0. doi:10.1007/s11633-014-0835-0.
5. Dalal, N., Triggs, B.. Histograms of oriented gradients for human detection. In: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*; vol. 1. 2005, p. 886–893 vol. 1. doi:10.1109/CVPR.2005.177.
6. Lowe, D.G.. Object recognition from local scale-invariant features. In: *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*; vol. 2. 1999, p. 1150–1157 vol.2. doi:10.1109/ICCV.1999.790410.
7. Zhao, G., Pietikainen, M.. Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2007;**29**(6):915–928. doi:10.1109/TPAMI.2007.1110.
8. Kim, B.K., Roh, J., Dong, S.Y., Lee, S.Y.. Hierarchical committee of deep convolutional neural networks for robust facial expression recognition. *Journal on Multimodal User Interfaces* 2016;:1–17URL: http://dx.doi.org/10.1007/s12193-015-0209-0. doi:10.1007/s12193-015-0209-0.
9. Wang, X., Guo, R., Kambhamettu, C.. Deeply-learned feature for age estimation. In: *Applications of Computer Vision (WACV), 2015 IEEE Winter Conference on*. 2015, p. 534–541. doi:10.1109/WACV.2015.77.
10. Levi, G., Hassncer, T.. Age and gender classification using convolutional neural networks. In: *Computer Vision and Pattern Recognition Workshops (CVPRW), 2015 IEEE Conference on*. 2015, p. 34–42. doi:10.1109/CVPRW.2015.7301352.
11. Lucey, P., Cohn, J.F., Kanade, T., Saragih, J., Ambadar, Z., Matthews, I.. 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. ISBN 978-1-4244-7029-7; 2010, p. 94–101. URL: http://dx.doi.org/10.1109/cvprw.2010.5543262. doi:10.1109/cvprw.2010.5543262.
12. Shan, C., Gong, S., McOwan, P.W.. Facial expression recognition based on local binary patterns: A comprehensive study. *Image Vision Comput* 2009;**27**(6):803–816. URL: http://dx.doi.org/10.1016/j.imavis.2008.08.005. doi:10.1016/j.imavis.2008.08.005.
13. Littlewort, G., Whitehill, J., Wu, T., Fasel, I., Frank, M., Movellan, J., et al. The computer expression recognition toolbox (cert). In: *Automatic Face Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*. 2011, p. 298–305. doi:10.1109/FG.2011.5771414.
14. Lyons, M.J., Budynek, J., Akamatsu, S.. Automatic classification of single facial images. *IEEE Trans Pattern Anal Mach Intell* 1999;**21**(12):1357–1362. URL: http://dx.doi.org/10.1109/34.817413. doi:10.1109/34.817413.
15. Krizhevsky, A., Sutskever, I., Hinton, G.E.. Imagenet classification with deep convolutional neural networks. In: Pereira, F., Burges, C., Bottou, L., Weinberger, K., editors. *Advances in Neural Information Processing Systems 25*. Curran Associates, Inc.; 2012, p. 1097–1105.

URL: `http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf`.

16. Kahou, S.E., Froumenty, P., Pal, C.. *Facial Expression Analysis Based on High Dimensional Binary Features*. Cham: Springer International Publishing. ISBN 978-3-319-16181-5; 2015, p. 135–147. URL: `http://dx.doi.org/10.1007/978-3-319-16181-5_10`. doi:10.1007/978-3-319-16181-5_10.

17. Ouellet, S.. Real-time emotion recognition for gaming using deep convolutional network features. *CoRR* 2014;**abs/1408.3750**. URL: `http://arxiv.org/abs/1408.3750`.

18. Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., et al. Decaf: A deep convolutional activation feature for generic visual recognition. *CoRR* 2013;**abs/1310.1531**. URL: `http://arxiv.org/abs/1310.1531`.

19. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., et al. Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:14085093* 2014;.

20. Lyons, M., Akamatsu, S., Kamachi, M., Gyoba, J.. Coding facial expressions with gabor wavelets. In: *Proceedings of the 3rd. International Conference on Face & Gesture Recognition*; FG '98. Washington, DC, USA: IEEE Computer Society. ISBN 0-8186-8344-9; 1998, p. 200–. URL: `http://dl.acm.org/citation.cfm?id=520809.796143`.

21. Bradski, G.. The OpenCV Library. *Dr Dobb's Journal of Software Tools* 2000;.

22. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 2011;**12**:2825–2830.

23. Rychetsky, M.. *Algorithms and Architectures for Machine Learning Based on Regularized Neural Networks and Support Vector Approaches*. Germany: Shaker Verlag GmbH; 2001. ISBN 3826596404.

24. Landis, J.R., Koch, G.G.. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 1977;**33**(1):159–174. URL: `http://dx.doi.org/10.2307/2529310`. doi:10.2307/2529310.

25. Siddiqi, M.H., Ali, R., Khan, A.M., Kim, E.S., Kim, G.J., Lee, S.. Facial expression recognition using active contour-based face detection, facial movement-based feature extraction, and non-linear feature selection. *Multimedia Systems* 2014;**21**(6):541–555. URL: `http://dx.doi.org/10.1007/s00530-014-0400-2`. doi:10.1007/s00530-014-0400-2.

26. Mlakar, U., Potočnik, B.. Automated facial expression recognition based on histograms of oriented gradient feature vector differences. *Signal, Image and Video Processing* 2015;**9**(1):245–253. URL: `http://dx.doi.org/10.1007/s11760-015-0810-4`. doi:10.1007/s11760-015-0810-4.

27. Jeni, L.A., Takacs, D., Lorincz, A.. High quality facial expression recognition in video streams using shape related information only. In: *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*. 2011, p. 2168–2174. doi:10.1109/ICCVW.2011.6130516.

28. Zhao, X., Zhang, S.. Facial expression recognition based on local binary patterns and kernel discriminant isomap. *Sensors* 2011;**11**(10):9573. URL: `http://www.mdpi.com/1424-8220/11/10/9573`. doi:10.3390/s111009573.

29. Zhang, L., Tjondronegoro, D.. Facial expression recognition using facial movement features. *IEEE Transactions on Affective Computing* 2011;**2**(4):219–229. doi:10.1109/T-AFFC.2011.13.

30. Erhan, D., Szegedy, C., Toshev, A., Anguelov, D.. Scalable object detection using deep neural networks. In: *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*. 2014, p. 2155–2162. doi:10.1109/CVPR.2014.276.