International Journal of Scientific Research & Engineering Trends
Volume 6, Issue 6, Nov-Dec-2020, ISSN (Online): 2395-566X

Face Expression Recognition: A Review

M.Tech. Student Palak Jain, Assistant Professor Anamika Pyasi
Department of Computer Science Engineering
Babulal Tarabai Institute of Research & Technology,M.P.,India
enggpalakjain@gmail.com,anamikatiwari25@yahoo.com

***Abstract*- Recognition of outward appearances assumes a significant job in many mechanized framework applications like mechanical technology, instruction, man-made brainpower, and security. Perceiving outward appearances precisely is testing. Approaches for explaining FER (Facial Expression Recognition) issue can be ordered into 1) Static single pictures and 2) Image successions. Customarily, various procedures like Multi-layer Perceptron Model, k-Nearest Neighbors, Support Vector Machines were utilized by scientists for fathoming FER. These techniques removed highlights like Local Binary Patterns, Eigenfaces, Face-milestone highlights, and Texture highlights. Among every one of these techniques, Neural Networks have increased especially ubiquity and they are broadly utilized for FER. As of late, CNNs (Convolutional Neural Networks) have picked up notoriety in field of profound learning in view of their easygoing engineering and capacity to give great outcomes without prerequisite of manual component extraction from crude picture information. This paper centers around review of different face demeanor acknowledgment methods dependent on CNN. It incorporates cutting edge techniques proposed by various scientists. The paper additionally shows steps required for utilization of CNN for FER. This paper additionally incorporates examination of CNN based methodologies and issues requiring consideration while picking CNN for unraveling FER.**

***Keywords*-Classification, Survey, FER, Face Expression Recognition, Deep Learning, iCNN.**

I.INTRODUCTION

Outward appearances are common and ground-breaking signs to decipher human's passionate states and aims. These days, everything is getting robotized through PCs. Outward appearance Recognition (FER) has become well known research subject in the field of PC vision. Acknowledgment of outward appearances can be utilized in mechanical technology, neuro-showcasing, scholastics, and all the more altogether in security. We can accomplish much by precisely anticipating outward appearances of human.

This paper centers around a review and examination of two methodologies for settling FER: static pictures based and picture groupings based. Requirement for this overview is imperative in light of the fact that FER is being executed in numerous mechanical and government areas. Information is generally accessible however it needs fitting amendment. To manage such immense measure of information could be very tedious utilizing customary element based techniques. That is the reason scientists lean toward profound learning procedures, particularly CNNs for arrangement of pictures. All things considered, situations, pictures may differ as far as individual's head presents, enlightenment settings, helping conditions, and foundation. Articulation variety and impediment are significant issues. As of late, S. Li et al. [1] introduced an overview of profound learning methods like DBN (Deep Belief Network), CNN, Auto Encoders and RNN (Recurrent Neural Network). R. Ginne et al. [2] likewise introduced a review on CNN based FER systems. In the two works, just few CNN based methodologies are incorporated and not talked about inside and out.

This paper presents a wide overview on CNN based FER procedures. CNNs are demonstrated exceptionally hearty towards face related examination [3]. We have done a review on FER tending to every serious issue and their accessible arrangements like pre-handling and information growth. The paper talks about the referenced difficulties and features how those difficulties are handled by existing works. We have likewise indicated the utilization of CNN based strategies in Section II. Additionally, attributes of CNN design are likewise broke down to propose an estimated engineering according to the qualities of dataset.

This paper contains 4 Sections. Area II depicts the foundation information, use of CNN based systems, and issues in current FER. Area III presents an expansive overview on FER techniques centering single picture and picture arrangement, and

investigation of both the strategies alongside CNN qualities. Area IV presents the end.

II. BACKGROUND KNOWLEDGE

1. Facial Expressions

Outward appearances are generally effective and characteristic approach to pass on feelings and aims. S. Li et al. [4] signified that face appearances speak to intensions and enthusiastic condition of the individual. There are numerous articulations conceivable, yet P. Shaver et al. [5] inferred that lone seven articulations are the prototypic articulations, which are: Anger, Sadness, Disgust, Happiness, Fear, Surprise, and Neutral, into which every other articulation can fall.

2. Traditional feature based approach vs CNN

Generally, most analysts have proposed their strategies by utilizing classifiers like MLP (Multi-layer Perceptron Model), SVM (Support Vector Machines) and k-NN (k-Nearest Neighbors). These classifiers use handmade highlights like surface and face milestone highlights, HoG (Histogram of Oriented Gradients), angle include mapping, eigen vectors, and so forth. These highlights can be removed by systems like Gabor channels, LBP (Local Binary Patterns), Eigen Faces, LDA (Linear Discriminant Analysis), and PCA (Principal Component Analysis). Essentially, CNNs likewise utilize these highlights, however in their own particular manner. The thing that matters is, in customary techniques, we have to separate highlights (a.k.a "carefully assembled highlights") physically, while CNNs can learn such highlights consequently by their own.

3. Face expression classification using CNN

To perform FER efficiently, we concluded basic steps shown in Fig. 1.

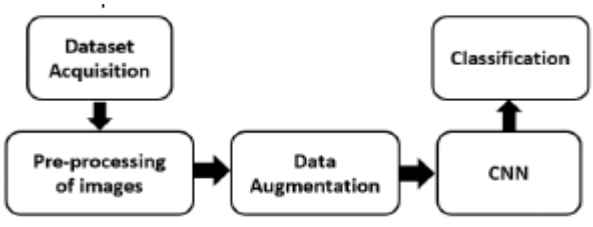

Figure 1. Basic steps to Perform CNN based Approach

4. Each step is discussed next in more detail.

4.1. Dataset Gathering

FER frameworks are issue explicit. They exceptionally rely upon which sort of pictures given as info. There are numerous datasets accessible online for different purposes like frontal face pictures, presented pictures, and unconstrained (wild setting) pictures. Nitty gritty investigation of dataset is done in Section III.

4.2. Pre-processing of images and data augmentation

Pre-processing methods are divided into three categories:

- Face recognition,
- Illumination standardization, and
- Pose standardization. After pre-preparing, information increase task is performed to get sufficient preparing tests. Information expansion is a way to deal with create integrated examples from unique pictures. Significant techniques are introduced in Table I.

Table 1. Pre-Processing Methods

Pre-processing	Method	Researchers
Face Detection	Viola-Jones	[6] [7] [8] [9] [10] [11]
	Dlib	[12]
Illumination Normalization	Histogram Equalization	[6] [9] [13]
	Discrete Cosine Transform	[12] [14]
	Zero-mean Normalization	[13] [3]
Pose Normalization	Frontalization	[11] [15]
	Face Alignment	[16] [17] [18]
	Face Normalization	[19] [20]
Data Augmentation	Data Synthesizing	[21] [7] [22]

4.3.CNN architecture

At long last, information is taken care of into CNN model for order. CNN is a class of profound neural systems. It just includes convolution activity. Henceforth, it is named "Convolutional Neural Network". For the most part three layers are remembered for CNN:

- Convolutional layer,
- Pooling layer, and
- Fully associated layer.

Going before layers separate essential shapes and succeeding layers take in progressively complex shapes from picture. Perusers can allude [23], [24] for additional subtleties on CNN. Highlight maps created by convolutional layers are diminished by utilizing pooling layers. Completely associated layer has all the associations from past layer. It processes lattice augmentation of those associations with

relating inclinations simply like counterfeit neural systems. We have to keep track about hyper parameters on the grounds that lone valuable parameters and data are required to take care of into CNN. Else, it prompts issue of overfitting, which gives higher preparing precision, yet poor test exactness. As it were, the model learns superfluous data and clamor during preparing stage.

The most widely recognized issues found in frontal presented pictures are articulation and light inconstancy, though unconstrained pictures contain serious issue of non-frontal head presents. Significant issues and arrangements proposed by various specialists are in Table 2.

Table 2. Issues In Fer.

Issue	Solution
Expression Variation	Select sequence of images from non-peak expression to peak expressions, Select multiple images of same subject (person) with various expression intensities
Illumination Variation	Histogram Equalization, InFace Toolbox, Gamma Intensity Correction, Logarithmic Transforms
Overfitting of Model	Cross-Validation, Pruning, Data Augmentation, Regularization, Early Stopping, Dropout [25]
Uncertainty of Occlusion	Feature Reconstruction, Gradient Direction [26], Sub-region based approach, 3D data based approach
Insufficient Data	Data Augmentation
Non-frontal Face Images	Frontalization [11], Frontal Face Generation

III. SURVEY ON FER

This section presents an intensive survey on both FER approaches.

1. Survey of single image based and image sequence based approaches

Table III investigates every significant work. Broadly utilized datasets are FER-13 [27], JAFFE [28], and CK+ [29]. Expanded Cohn Kanade (CK+) dataset incorporates 593 picture successions of 123 subjects. Japanese Female Facial Expression (JAFFE) dataset incorporates 213 pictures of 10 Japanese ladies. Both datasets are lab controlled datasets and has presented pictures of seven prototypic articulations which are: Happy, Sad, Disgust, Fear, Surprise, Angry and Neutral. Also, CK+ contains one more "Disdain" articulation, however analysts dispose of this articulation in tests due to assessment reason. FER-13 dataset contains 35887 unconstrained pictures gathered from Google search API and seven prototypic articulations.

V. Mavani et al. [30] introduced a novel method of visual saliency. Visual Saliency is a guide of force where higher powers show fields of most extreme consideration and lower powers show least consideration of picture. K. Liu et al. [31] proposed the ensemble method which uses multiple CNNs. They created 3 subnets (3 individual CNNs) and combined their results at the succeeding layer. Z. Yu et al. [13] used pre-trained model on FER-13 dataset and fine-tuned on SFEW 2.0 dataset, finally tested on both datasets.

Table 3. Analysis of Single Static Image Based And Image Sequence Based Approaches.

Researcher	Dataset	Samples	Subjects	Pre-processing Methods	Classes[a]	Model	Result
K. Liu et al. (2016) [31]	FER2013 [27]	35887	N/A	N/A	6 + 1 (HA, SU, DI, FE, AN, SA) + NEU	Ensemble CNN	65.03%
X. Zhao et al. (2017) [32]	Oulu-CASIA [33]	2880	80	Peak Gradient Suppression (PGS)	6 (HA, SU, DI, FE, AN, SA)	CNN	84.59%
	CK+ [29]	593	123				99.3%
	RaFD	1407	67				95.71%
Uçar et al. (2017) [34]	Cohn-Kanade	2105	182	N/A	6 + 1 (HA, SU, DI, FE, AN, SA) + NEU	CNN	98.70%
V. Mavani et al. (2017) [30]	CFEE [35]	1610	230	Visual saliency, Face cropping	6 + 1 (HA, SU, DI, FE, AN, SA) + NEU	CNN	74.79%
	RaFD	1407	67				95.71%
B. Yang et al. (2015) [36]	CK+ [29]	593	123	Rotation rectification, LBP	6 (HA, SU, DI, FE, AN, SA)	CNN	97.00%
	Oulu-CASIA [33]	2880	80				92.3%
A. Lopes et al. (2016) [21]	JAFFE [28]	213	10	N/A	6 + 1 (HA, SU, DI, FE, AN, SA) + NEU	CNN	53.57%
	BU-3DFE	2500	100				71.62%
	CK+ [29]	593	123				95.79%
Z. Yu et al. (2015) [13]	FER2013 [27]	35887	N/A	Face detection, Likelihood loss, Hinge loss	6 + 1 (HA, SU, DI, FE, AN, SA) + NEU	CNN	Around 80%
	SFEW 2.0	600	68				61.29%
K. Zhang et al. (2016) [37]	CK+ [29]	593	123	Dynamic features, Static features	6 (HA, SU, DI, FE, AN, SA)	PHRNN[e] + MSCNN[f]	98.50%
	Oulu-CASIA [33]	2880	80				86.25%
	MMI	1280 videos	43				81.18%
P. Hu et al. (2017) [15]	EmotiW 2017	1809	N/A	SSE[b], SDM[c], Face frontalization, DCT[d]	6 + 1 (HA, SU, DI, FE, AN, SA) + NEU	CNN	60.34%
A. Mollahosseini et al. (2016) [38]	MMI	1280 videos	43	Bidirectional warping, IntraFace	6 + 1 (HA, SU, DI, FE, AN, SA) + NEU	CNN	77.9%
	DISFA	4845 video frames	27				55.0%
	FERA	289	10				76.7%
	SFEW	663	95				47.7%
A. Fathallah et al. (2017) [39]	MUG	1462	86	N/A	6 (HA, SU, DI, FE, AN, SA)	CNN	87.65%
	RAFD	1407	67				99.33%
	CK+ [29]	593	123				99.33%
E. Ijjina et al. (2014) [26]	EURECOM kinect face dataset	N/A	52	Facial depth by kinect sensor, Gradient direction	6 (Occlusion paper, Occlusion mouth, Left profile, Open mouth, Right profile, Neutral)	CNN	87.98%
K. Shan et al. (2017) [6]	JAFFE [28]	213	10	Face detection, Histogram equalization	6 + 1 (HA, SU, DI, FE, AN, SA) + NEU	CNN	76.7442%
	CK+ [29]	593	123				80.303%
X. Chen et al. (2017) [40]	JAFFE [28]	213	10	Image normalization	6 + 1 (HA, SU, DI, FE, AN, SA) + NEU	CNN	87.735%
	CK+ [29]	593	123				99.16%
W. Li et al. (2015) [7]	CIFE	10595	N/A	Face alignment and rectification, Image cropping	6 + 1 (HA, SU, DI, FE, AN, SA) + NEU	CNN	81.5%
	CK+ [29]	593	123	Face alignment and Rectification			83%
R. Kumar et al. (2017) [9]	FER-2013 [27]	35887	N/A	Viola- Jones algorithm,	6 + 1 (HA, SU, DI, FE, AN, SA) + NEU	CNN	Around 90%
	CK+ [29]	593	123				
H. Li et al. (2017) [41]	BU-3DFE	SS[g] 1: 1200 SS 2: 2400	100 in both subsets	Nose detection, Face cropping, Re- sampling, 3D Face normalization	6 (HA, SU, DI, FE, AN, SA)	DF-CNN	SS 1- 86.20% SS 2- 81.33%
	Bosphorus 3D	360 (SS)	60				80.00%

E. Ijjina et al. [26] proposed a procedure which utilized just a profundity information of a picture rather than RGB data in light of the fact that dissimilar to RGB data, profundity data is obtuse toward brightening conditions. Profundity information (bit profundity) is the complete number of bits for each shading patch of single pixel. They utilized kinect profundity sensor to get profundity information.

K. Zhang et al. [37] proposed Part-based HierarchicalBidirectional Recurrent Neural Network (PHRNN) toremove facial highlights from fleeting groupings. Forthe most part two terms are included: Spatialhighlights and Temporal highlights. Spatial highlightsare the information spoken to with explicit area andpersonality though Temporal highlights allude to theinformation spoke to in some part of time.Consequently, their PHRNN extricates transienthighlights from successive casings and Multi-SignalConvolutional Neural Network (MSCNN) separatesspatial highlights from still edges to acquire the stillappearance data.

2. Analysis of CNN architecture

Table IV shows an escalated examination of the papers so as to propose a productive engineering for picked issue. After examination, we found that when we manage presented pictures, CNN requires just 1 or 2 convolutional layers which are trailed by pooling layers. Generally max pooling is utilized by analysts. Convolutional layers require 3x3 to 7x7 bit size (for the most part odd numbers practically speaking). Number of parts increment bit by bit from going before to succeeding layers. All these first layers are trailed by a couple of completely associated layers, which contain neurons running from around 100 to 3072 (relies upon issue). Convolutional layers remove valuable highlights and max pool layer is utilized to diminish measurements of highlight maps created by convolutional layer.

Unconstrained pictures require progressively number of layers to encourage learning of complex shapes. CNN expects 2 to 6 convolutional layers followed by pooling layers. Be that as it may, it can't the case to continue pooling layer directly after convolutional layer. A few specialists [40], [31] utilized two convolutional layers successively which are then trailed by one pooling layer. Expectation behind doing so is to learn model all the more over and again and afterward lessen the measurements utilizing pooling layer. Progressively number of layers lead to an issue of enormous number of hyper-parameters. Z. Yu et al. [13] utilized dropout method [25] which haphazardly drops neurons. This arbitrariness is useful to decrease the danger of overfitting of model. It is utilized when we have high number of preparing parameters. Information growth is additionally utilized by [21], [7], [22], to limit issue of overfitting.

We conclude the following points related to CNN architecture:

A concise CNN with less number of layers and hyper- parameters is preferred for frontal face images.

A complex CNN with more number of layers and parameters is preferred for spontaneous faces.

A moderate CNN with balanced number of layers and parameters with proper pre-processing is preferred for occluded images.

Table 4. Reasoning on CNN Architecture

Dataset	Researcher	CNN Architecture (Layer Name, Size of kernel, No. of kernels, Stride)	Result
CK+ [29] (Posed, Spontaneous)	A. Lopes et al. (2016) [21]	(I/P, 32x32), (Conv1, 5x5, 32), (MaxPool, 2x2, 32), (Conv2, 7x7, 64), (MaxPool, 2x2, 64), (FC1, 256)	95.79%
	X. Chen et al. (2017) [40]	(I/P, 227x227), (Conv1, 255x255, 96), (Conv2, 29x29, 128), (MaxPool, 14x14, 128), (Conv3, 12x12, 156), (Conv4, 6x6, 256), (MaxPool, 3x3, 256), (FC, 512)	99.16%
	W. Li et al. (2015) [7]	(I/P, 64x64), (Conv1, 7x7,32), MaxPool(2:1), (Conv2, 7x7, 32), (MaxPool, 2:1), (Conv3, 7x7, 64), (FC1, N/A)	83%
	R. Kumar et al. (2017) [9]	(I/P, 48x48), (Conv1, 5x5, 64), (MaxPool, 3x3, St-2), (Conv2, 5x5, 64), (MaxPool, 3x3), (Conv3, 4x4, 128), (FC, 3072)	Around 90%
	K. Shan et al. (2017) [6]	(I/P, N/A), (Conv1, 5x5, 6), (MaxPool, 2x2), (Conv2, 5x5,12), (MaxPool, 2x2), (FC1, N/A)	80.303%
	K. Zhang et al. (2016) [37]	(I/P, 64x64), (CROP, 60x60), (Conv1, 10), (MaxPool), (Conv2, 20), (MaxPool), (Conv3, 40), (MaxPool), (Conv4, 40), (MaxPool), (FC1, 80)	98.50%
	A. Fathallah et al. (2017) [39]	(I/P, 165x165), (Conv1, 4x4), (MaxPool, 2x2), (Conv2, 3x3), (MaxPool, 2x2), (Conv3, 3x3), (MaxPool, 2x2), (FC1, 160)	99.33%
Oulu-CASIA [33] (Posed)	K. Zhang et al. (2016) [37]	(I/P, 64x64), (CROP, 60x60), (Conv1, 10), (MaxPool), (Conv2, 20), (MaxPool), (Conv3, 40), (MaxPool), (Conv4, 40), (FC1, 80)	86.25%
JAFFE [28] (Posed)	A. Lopes et al. (2016) [21]	(I/P, 32x32), (Conv1, 5x5, 32), (MaxPool, 2x2, 32), (Conv2, 7x7, 64), (MaxPool, 2x2, 64), (FC1, 256)	53.57%
	Uçar et al. (2017) [34]	(I/P, N/A), (Conv1, 5x5), (MaxPool, 3x3, St-2), (Conv2, 5x5), (MaxPool, 3x3, St-1), (Conv3, 5x5), (MaxPool, 3x3, St-2), (Conv4, 2x2), (Conv5, 1x1), (FC1-N/A)	96.10%
	K. Shan et al. (2017) [6]	(I/P, N/A), (Conv1, 5x5, 6), (MaxPool, 2x2), (Conv2, 5x5,12), (MaxPool, 2x2), (FC1, N/A)	76.7442 %
	X. Chen et al. (2017) [40]	(I/P, 227x227), (Conv1, 255x255, 96), (Conv2, 29x29, 128), (MaxPool, 14x14, 128), (Conv3, 12x12, 156), (Conv4, 6x6, 256), (MaxPool, 3x3, 256), (FC1, 512)	87.74%
FER-2013 [27] (Spontaneous)	Z. Yu et al. (2015) [13]	(I/P, 48x48), (Conv1, 5x5), (Stochastic Pool, 3x3, St-2), (Conv2, 3x3, 64), (Conv3, 3x3, 64), (StochasticPool, 3x3, St-2), (Conv4, 3x3, 128), (Conv5, 3x3, 128), (StochasticPool, 3x3, St-2), (FC1, 1024), (FC2, 1024)	Around 80%
	R. Kumar et al. (2017) [9]	(I/P, 48x48), (Conv1, 5x5, 64), (MaxPool, 3x3, St-2), (Conv2, 5x5, 64), (MaxPool, 3x3), (Conv3, 4x4, 128), (FC1, 3072)	Around 90%
	K. Liu et al. (2016) [31]	Subnet-1 (Conv1, 3x3, 64), (MaxPool, 2x2, St-2), (Conv2, 3x3, 128), (MaxPool, 2x2, St-2), (Conv3, 3x3, 256), (MaxPool, 2x2, St-2), (FC1, 4096), (FC2, 4096) Subnet-2 (Conv1, 3x3, 64), (MaxPool, 2x2, St-2), (Conv2, 3x3, 128), (MaxPool, 2x2, St-2), (Conv3, 3x3, 256), (Conv4, 3x3, 256), (MaxPool, 2x2, St-2), (FC1, 4096), (FC2, 4096) Subnet-3 (Conv1, 3x3, 64), (MaxPool, 2x2, St-2), (Conv2, 3x3, 128), (Conv3, 3x3, 128), (MaxPool, 2x2, St-2), (Conv4, 3x3, 256), (Conv5, 3x3, 256), (MaxPool, 2x2, St-2), (FC1, 4096), (FC2, 4096)	65.03%

IV. CONCLUSION

This paper concentrated on adequacy of convolutional neural system for outward appearance arrangement. FER is valuable in numerous applications like instruction, HMI frameworks, and security. This paper passed on helpful foundation information to comprehend FER area alongside various pre-preparing methods. It additionally introduced contrast between conventional element based and CNN based methodologies for fathoming CNN. It is reasoned that customary component extraction based techniques became tedious for choosing suitable highlights for learning of model. CNN naturally learns those highlights proficiently and that is the reason, CNN can turn out to be exceptionally advantageous for true situations. We did a

comprehensive study and investigation of FER strategies utilizing CNN. Moreover, we dissected CNN structures of various works and proposed design according to the attributes of dataset.

REFERENCES

[1] S. Li and W. Deng, "Deep Facial Expression Recognition: A Survey," Computer Vision and Pattern Recognition, 2018.

[2] R. Ginne and K. Jariwala, "Facial Expression Recognition using CNN: A Survey," International Journal of Advances in Electronics and Computer Science, vol. 5, no. 3, 2018.

[3] C. Garcia and M. Delakis, "Convolutional Face Finder: A Neural Architecture for Fast and Robust Face Detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 11, 2004.

[4] S. Z. Li and A. K. Jain, "Handbook of Face Recognition," in Handbook of Face Recognition, Springer, 2004.

[5] P. Shaver, J. Schwartz, D. Kirson and G. O'Connor, "Emotion Knowledge: Further Exploration of a Prototype Approach," Personality aid Social Psychology, vol. 52, no. 6, pp. 1061-1086, 1987.

[6] K. Shan, J. Guo, W. You, D. Lu and R. Bie, "Automatic Facial Expression Recognition Based on a Deep Convolutional-Neural- Network Structure," IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA), pp. 123-128, 2017.

[7] W. Li, M. Li, Z. Su and Z. Zhu, "A Deep-Learning Approach to Facial Expression Recognition with Candid Images," 14th IAPR International Conference on Machine Vision Applications (MVA), pp. 279-282, 2015.

[8] H. W. Ng, V. D. Nguyen, V. Vonikakis and S. Winkler, "Deep Learning for Emotion Recognition on Small Datasets Using Transfer Learning," Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, pp. 443-449, 2015.

[9] R. Kumar, R. Kant and G. Sanyal, "Facial Emotion Analysis using Deep Convolution Neural Network," International Conference on Signal Processing and Communication (ICSPC), pp. 369-374, 2017.

[10] P. Viola and M. Jones, "Rapid Object Detection Using a Boosted Cascade of Simple Features," Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2001, pp. I-I, 2001.

[11] T. Hassner, S. Harel, E. Paz and R. Enbar, "Effective face frontalization in unconstrained images," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4295-4304, 2015.

[12] M. Shin, M. Kim and D.-S. Kwon, "Baseline CNN structure analysis for facial expression recognition," 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), 2016.

[13] Z. Yu and C. Zhang, "Image based Static Facial Expression Recognition with Multiple Deep Network Learning," Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, pp. 435-442, 2015.

[14] C. Weilong, E. M. Joo and W. Shiqian, "Illumination compensation and normalization using logarithm and discrete cosine transform in logarithm domain," ICARCV 2004 8th Control, Automation, Robotics and Vision Conference, vol. 1, pp. 380-385, 2014.

[15] P. Hu, D. Cai, S. Wang, A. Yao and A. Yao, "Learning Supervised Scoring Ensemble for Emotion Recognition in the Wild," ICMI '17, pp. 553-560, 2017.

[16] D. Chen, S. Ren, Y. Wei, X. Cao and J. Sun, "Joint cascade face detection and alignment," Computer Vision – ECCV, pp. 109-122, 2014.

[17] X. Cao, Y. Wei, F. Wen and J. Sun, "Face Alignment by Explicit Shape Regression," International Journal of Computer Vision, vol. 107, no. 2, p. 177–190, 2014.

[18] K. Zhang, Z. Zhang, Z. Li and Y. Qiao, "Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks," IEEE Signal Processing Letters, vol. 23, no. 10, 2016.

[19] C. Zhang and Z. Zhang, "Improving multiview face detection with multi-task deep convolutional neural networks," IEEE Winter Conference on Applications of Computer Vision, pp. 1036-1041, 2014.

[20] V. Blanz, P. Grother, P. J. Phillips and T. Vetter, "Face Recognition based on Frontal Views generated from Non-Frontal Images," IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), pp. 454-

461, 2015.

[21] A. T. Lopes, A. F. D. S. E. de Aguiar and T. O-Santos, "Facial Expression Recognition with Convolutional Neural Networks: Coping with Few Data and the Training Sample Order," Pattern Recognition.

[22] S. Yang, P. Luo, C. C. Loy and X. Tang, "Faceness-Net: Face Detection through Deep Facial Part Responses," IEEE Transactions on Pattern Analysis and Machine Intelligenc, vol. 40, no. 8, pp. 1845 - 1859, 2017.

[23] M. A. Ponti, L. S. F. Ribeiro, T. S. Nazare, T. Bui and J. Collomosse, "Everything you wanted to know about Deep Learning for Computer Vision but were afraid to ask," 30th SIBGRAPI Conference on Graphics, Patterns and Images Tutorials (SIBGRAPI-T), pp. 17-41, 2017.

[24] K. Nogueira, O. A. B. Penatti and J. A. d. Santos, "Towards Better Exploiting Convolutional Neural Networks for remote sensing scene classification," Pattern Recognition, vol. 61, no. C, pp. 539-556, 2016.

[25] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," The Journal of Machine Learning Research, vol. 15, no. 1, pp. 1929-1958, 2014.

[26] E. P. Ijjina and C. K. Mohan, "Facial expression recognition using kinect depth sensor and convolutional neural networks," 13th International Conference on Machine Learning and Applications, pp. 392-396, 2014.

[27] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee, Y. Zhou, C. Ramaiah, F. Feng and R. L. e. al., "Challenges in Representation Learning: A Report on Three Machine Learning Contests," International Conference on Neural Information, p. 117–124.

[28] M. J. Lyons, S. Akemastu, M. Kamachi and J. Gyoba, "Coding Facial Expressions with Gabor Wavelets," 3rd IEEE International Conference on Automatic Face and Gesture Recognition, pp. 200-205, 1998.

[29] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih and Z. Ambadar, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," Computer Vision and Pattern Recognition Workshops (CVPRW), p. 94–101, 2010.

[30] M. Viraj, R. Shanmuganathan and M. K. P, "Facial Expression Recognition using Visual Saliency and Deep Learning," IEEE International Conference on Computer Vision Workshops (ICCVW), 2017.

[31] K. Liu, M. Zhang and Z. Pan, "Facial Expression Recognition with CNN Ensemble," 2016 International Conference on Cyberworlds (CW), pp. 163-166, 2016.

[32] X. Zhao, X. Liang, L. Liu, T. Li, Y. Han, N. Vasconcelos and S. Yan, "Peak-Piloted Deep Network for Facial Expression," Computer Vision – ECCV 2016, pp. 425-442, 2016.

[33] G. Zhao, X. Huang, M. Taini, S. Z. Li and M. PietikäInen, "Facial expression recognition from near-infrared videos," Image and Vision Computing, vol. 29, no. 9, pp. 607-619, 2011.

[34] A. Uçar, "Deep Convolutional Neural Networks for Facial Expression Recognition," IEEE International Conference on INnovations in Intelligent SysTems and Applications (INISTA), pp. 371-375, 2017.

[35] S. Du, A. M. Martinez and Y. Tao, "Compound facial expressions of emotion," 2014.

[36] Y. Biao, C. Jinmeng, N. Rongrong and Z. Yuyu, "Facial Expression Recognition using Weighted Mixture Deep Neural Network Based on Double-channel Facial Images," IEEE Access, vol. 6, pp. 4630-4640, 2018.

[37] K. Zhang, Y. Huang, Y. Du and L. Wang, "Facial Expression Recognition Based on Deep Evolutional Spatial-Temporal Networks," IEEE Transactions on Image Processing , vol. 26, no. 9, pp. 4193- 4203, 2016.

[38] A. Mollahosseini, D. Chan and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1-10, 2016.

[39] A. Fathallah, L. Abdi and A. Douik, "Facial Expression Recognition via Deep Learning," IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA), pp. 745-750, 2017.

[40] X. Chen, X. Yang, M. Wang and J. Zou,

"Convolution Neural Network for Automatic Facial Expression Recognition," International Conference on Applied System Innovation (ICASI), pp. 814-817, 2017.

[41] H. Li, J. Sun, Z. Xu and L. Chen, "Multimodal 2D+3D Facial Expression Recognition with Deep Fusion Convolutional Neural Network," IEEE Transactions on Multimedia , vol. 19, no. 12, pp. 2816 - 2831, 2017.