IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Vol.-6, Issue-12, Dec 2018 E-ISSN: 2347-2693

A Bird View on Deep Learning Facial Expression Recognition Approaches for Thermal and Infrared Images

Reshma Nehlani[1*], Devang Pandya[2]

[1]Department of Computer Engineering, Silver Oak College of Engineering, Gujarat Technological University, Ahmedabad, India
[2]Department of Information and Communication Technology, Adani Institute of Infrastructure Engineering, Ahmedabad, India

[*]Corresponding Author: reshma.nehlani@gmaill.com, Tel.: +91-99742-17433

Available online at: www.ijcseonline.org

Accepted: 16/12/2018, Published: 31/Dec/2018

Abstract— With the capability to self-learn and succeed to achieve favourable results in various classification problem, deep learning techniques are increasingly used for Automatic Facial Expression Recognition (AFER). In this paper, we provide brief survey on deep learning technique particularly Convolutional Neural Networks (CNN) for Facial Expression Recognition (FER) and newly introduced Infrared based FER dataset. This review is focused on various CNN techniques applied in almost last half decade for FER on Infrared and Visible light images. There are certain unique advantages of using thermal and infrared images which can make FER techniques robust. Paper describes the standard flow of deep facial expression recognition and suggested methods based on research conducted specifically in this area. Later, review of existing novel deep neural networks and implementations for still images and video-based FER for Infrared Images is provided which subsequently follows glimpses of available well-known datasets. Since all types of cameras experience price reduction over the years, in near future integration and usage of such cameras would be common also because of its illumination invariant characteristic. It becomes evident at the end of the paper that there is a definite scope of developing promising and robust FER with use of Infrared and Thermal images.

***Keywords*—**Convolution Neural Networks, Deep Learning, Facial Expression Recognition, Infrared Images

I. Introduction

Facial expressions are important medium of interaction between human beings as they are expressed before we speak. It is one of key research area in computer vision because of its potential application in human computer interaction. Other applications of FER include medical diagnosis, robotics, security applications and driver-fatigue surveillance, virtual reality.

Ekman et al. [1] defined Facial action coding system (FACS) based on facial muscle change and identified six basic emotions anger, disgust, fear, sad, surprise and happy based on cross-culture study. Contempt, a subsequent addition as one of basic emotion is proposed in [2]. Based on feature representation facial expression recognition (FER) can be divided into two different categories: Static or Frame-based methods and dynamic or image sequence methods. In static or frame-based methods, the feature maps are obtained using current single frame providing spatial information, whereas dynamic methods consider the temporal relation among contiguous frames in the input sequence.

Further FER methodologies can be broadly divided as conventional methods and deep learning methods. The majority of the conventional methods have used features defined manually (e.g., local binary patterns (LBP) [3], LBP on three orthogonal planes (LBP-TOP) [4], SIFT, Facial landmarks) for FER. Figure 1. shows major steps using conventional methods.

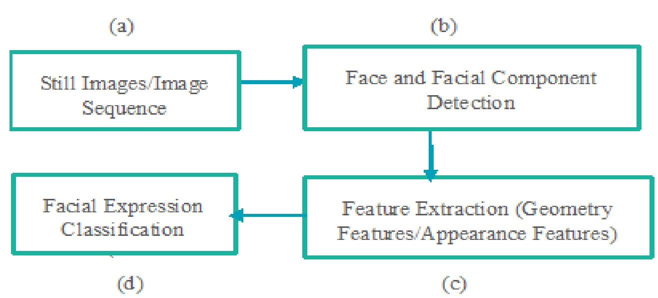

Figure 1. Block Diagram of Standard FER process

In difference to conventional approaches based on engineered or hand-crafted methods, deep learning has become prominent general approach to machine learning. It is yielding state-of-the-art results in many computer vision studies with the availability of big data and GPU. Deep-learning-based FER approaches relieves the dependence on human expertise for feature extraction by possessing intrinsic self-learning capability directly from the input images. Among several other deep learning models, Convolutional Neural Network (CNN) is the most popular network model. In this survey, techniques based on CNN is provided. Mali et al. [5] applied CNN for developing music system based on mood detection.

Further, researchers have started to focus on images which are captured through infrared or thermal sensors. Varying illumination conditions affect the performance of FER carried out on images captured in visible spectrum. Infrared spectrum offers advantage of being invariant to illumination changes, which means expression changes due to inhomogeneous light conditions would not affect the performance of the system. Applications include study of physiophsychological effects visible in the IR domain, extracting respiratory rate from thermal recordings [6], surveillance and driver drowsiness.

Most current FER systems capture images/videos in the visible light spectrum (VIS) (380 nm–750 nm). Near-infrared (NIR) (780 nm–1100 nm) and LWIR (Long wave infrared) or thermal cameras operate in wavelengths (7 - 14 μm) are alternatives for facial expression recognition to provide robustness against environmental conditions. But thermal sensors are expensive and there are no specialized algorithms developed for thermal images are limiting factors. The first limitation has been addressed by using affordable sensors such as microbolometers array sensors [7]. The purpose of carrying out the survey is to highlight the work carried out for facial expression recognition using infrared images so far and future scope of deep learning techniques in further achieving the task with availability of more thermal datasets.

This paper is organized as follows: - Section I contains Introduction of characteristics, methodologies of FER, Section II briefs pre-processing techniques, data augmentation, Convolutional neural networks, Section III introduces datasets captured in Infrared spectrum and section IV covers literature survey on FER using deep learning techniques and Infrared datasets, Section V concludes with future direction of facial expression recognition.

II. DEEP FACIAL EXPRESSION RECOGNITION TECHNIQUES

II.I Pre-processing

Pre-processing is required to reduce the variations such as different backgrounds, head poses deflection which are irrelevant for the task.

The first step in FER is to detect the face and remove non-face areas. S.He, S.Wang, W.Lan, H.Fu, and Q.Ji [8] adopted Otsu threshold to binarize infrared thermal images . Though Face detection is requisite step, further face registration can substantially enhance the FER performance. Based on the facial landmarks co-ordinates, faces can be registered into a predefined uniform template with an affine transformation. The Viola-Jones (V&J) face detector is a classic and widely adopted method for face detection [9]. *IntraFace* has been employed for face registration in deep facial expression recognition systems [10,11]. It is a supervised descent method and provides 49 accurate facial landmark points including the two eyes, the nose, the mouth, and the two eyebrows. Rotation rectification is implemented with the help of landmarks such as eyes. Other effective open-source algorithm is Dlib C++ library [12] implemented in OpenCV capable of locating 68 facial landmarks.

II.II Data Augmentation

Deep neural networks require large training data to ensure generalizability to a given recognition task and alleviate overfitting. Data augmentation can be done in two ways: offline data augmentation and online or on-the-fly data augmentation.

Numerous offline data augmentation techniques have been contributed for deep FER. The most frequently used methods include translation, horizontal flips, rotation, scaling and sheer for random transformation proposed by Simard et al [13]. Lopes et al. [14] used 2D gaussian distribution for adding random noise in location of eyes. Few other techniques such as applying appearance filters [15], adding noise such as salt and pepper [16] are used for data augmentation.

On-the-fly data augmentation is often embedded in deep learning toolkits to mitigate overfitting problem. ***ImageDataGenerator*** API in Keras framework is one such implementation available. During the training step, the input samples are rotated, translated, flipped horizontally, scaling, zoomed to increase dataset size online.

II.III Feature Learning with Convolution Neural Network (CNN)

Deep learning attempts to capture high-level abstractions through hierarchical architectures of multiple nonlinear transformations and representations. Among various deep learning techniques such as Recurrent neural network, deep belief network the Convolutional Neural Network (CNN) is one of the most successful deep models for visual tasks.

A CNN consists of three types of different layers: **convolutional layers, pooling layers, and fully connected layers** as illustrated in Fig 2**.** The convolutional layer has a set of learnable filters to convolve through the whole input image and produce various specific types of activation feature maps. These feature maps are fully connected to dense layers and based on maximum probability obtained by SoftMax function expression is recognized.

The advantages of CNN are: little pre-processing required, independence from prior expert knowledge for feature extraction. Some distinct features of CNN are local receptive fields which learns correlations among neighbouring pixels; weight sharing in the same feature map reduces the number of the parameters to be learned and shift-invariance to the location of the object as each filter is replicated across entire visual field.

Second the pooling layer is used to down sample size of the feature maps and the computational cost of the network. Average pooling and max pooling are the two most commonly used nonlinear down-sampling strategies for translation invariance. The fully connected layer is usually included at the end of the network to ensure that all neurons in the layer are fully connected to activations in the previous layer and to enable the 2D feature maps to be converted into 1D feature maps for further feature representation and classification. In addition to these three typical layers, various follow-up strategies are widely used in CNNs.

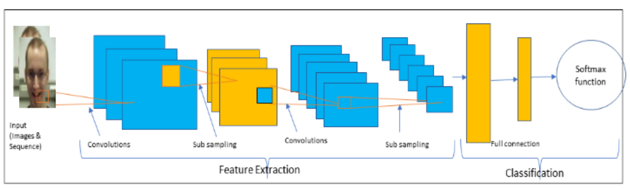

Figure 2: A general architecture of CNN (Face Images [31])

Rectified linear unit (ReLu) [17] is the most common activation function which increases the nonlinear properties of the network without suffering from the vanishing gradient problem. Other variants of ReLu are Leaky ReLu [18] to overcome vanishing gradient problem caused by hard zero activations in ReLu, Parametric ReLu (PReLu) [19].

Batch normalization (BN) [20] was proposed to regularize the model by reducing internal covariate shift and improve the convergence speed. Dropouts [21] is another universal strategy to reduce over fitting by randomly dropping a portion of the feature detectors usually applied to fully connected regions. Convolutional layers are also able to significantly reduce the complexity of the model through the optimization of its output. These are optimized through three hyperparameters, the depth, the stride and setting zero-padding [25].

Further some of well-known CNN models that have been successful for image recognition are summarized in Table 1. VGGNet is widely used pre-trained model in FER.

Table 1. CNN Models

CNN Models	**# of Convolution layers**	**# of Dense Layers**
AlexNet [17]	5	3
VGG-16/VGG-19[22]	13/16	3
GoogleNet[23]	21	1
ResNet[24]	151	1

One of the problems of CNN is over-fitting which can happen if the size of the network and the size of the database were not well matched. Peng et al. [26] found that a medium-sized network is more suitable for medium size dataset. Therefore, choosing right network size, hyperparameters and regularization techniques are vital for applying CNN.

II.IV Facial Expression Classification

After learning the deep features, the final step of FER is to classify the given face into one of the basic expression categories. Feature extraction and classification are independent step in deep neural networks which is not the case in conventional approach. Specifically, a loss layer is added to the end of the network to regulate the back-propagation error; then, the prediction probability of each sample can be directly output by the network.

Another alternative is to employ a deep neural network (particularly CNN) as a feature extraction tool and then apply additional classifiers, such as support vector machine (SVM) or Random Forest, to the extracted image representations. This technique of applying is known as Feature learning.

II.V Fusion Techniques

Many works have been reported by fusing different visual modalities describing the face or, more commonly, by using other sources of information (e.g. audio or physiological data). The fusion approaches can be grouped into three main categories: early, late and sequential fusion Early fusion merges the modalities at the feature level, while late fusion does so after applying expression recognition i.e. at the decision level. Fusing multiple modalities provides complementary information and thereby robustness against varying environment and type of data. Early fusion results in higher-dimensional feature space, increasing the likelihood of over-fitting. On the other hand, late fusion increases the amount of available data and applied by merging results obtained from multiple classifiers to predict final outcome.

A fusion function can be defined as f : x_t^a *,*x_t^b*,*$\rightarrow y_t$ *fuses two feature maps* $x_t^a \in R^{W*H*D}$ *and* $x_t^b \in R^{W'*H'*D'}$*, at time t, to produce an output map* $y^t \in R^{W''*H''*D''}$*, where W, H and D are the width, height and number of channels of the respective feature maps*[27]. Common fusion methods are concatenation, product, sum, subtract and max. Zhang et al. performed multimodal fusion taking joint representation by considering texture and landmark modality of features [28]. Yang et al. [34] authors use weighted summation of features extracted from grayscale images and LBP images using 10-fold cross validation while authors in [41] applies concatenation at feature level to fuse global and local features, while in. The former both are early fusion techniques.

III. INTRODUCTION TO FER DATASET

Here we briefly introduces some popular FER datasets captured in visible light and infrared spectrum.

The Extended Cohn-Kanade Dataset (CK+) [29]: CK+ is widely used in FER system containing 593 video sequences on both posed and non-posed (spontaneous) emotions. With 123 subjects, age ranging from 18 to 30 years, most of who are female. The images have pixel resolutions of 640×480 and 640×490 with 8-bit precision for gray-scale values. Details of access is available at ***http://www.consortium.ri.cmu.edu/ckagree/***

Japanese Female Facial Expressions (JAFFE) [30]: The JAFFE database contains 213 images of seven facial emotions (sad, angry, happy, surprise, disgust, fear and neutral) posed by ten different female Japanese models. Each image was rated based on six emotional adjectives using 60 Japanese subjects. The original size of each facial image is 256 pixels×256 pixels. Details of access is available at ***http://www.kasrl.org/jaffe.html***

Oulu-CASIA [31]: - Oulu-CASIA Near Infrared (NIR) facial expression database includes videos captured in three different illumination conditions: normal, weak, and dark. Database is collected in experiment room and contain image sequence captured both in visible light and infrared spectrum. Most of subjects are Chinese or Finnish. The videos of the database were captured by using a USB 2.0 PC Camera (SN9C201&202) that includes a Visible light spectrum (VIS) camera and a Near Infrared (NIR) camera. The NIR camera receives NIR light at around 850 nm by using 80 NIR light emitting diodes. Videos are captured with 25 frames per second imager with each image is of size 320 pixels * 240 pixels. The database contains 80 subjects between 23 and 58 years posing for 6 expressions with total of 480 video sequence (80*6). Details of dataset is available at ***http://www.cse.oulu.fi/CMV/Downloads/Oulu-CASIA***

Natural Visible and Infrared facial expression (USTC NVIE) [32]: - The database collected both spontaneous and posed expressions captured for more than 100 subjects simultaneously using a visible and an infrared thermal camera. The USTC-NVIE database [9] has images of 215 subjects (157 males and 58 females) ranging from 17 to 31 years old, collected at the University of Science and Technology of China. The spontaneous expressions consist of image sequences from onset to apex, collected by a visible and an infrared thermal camera at the same time under front, left, and right illumination. The posed database also includes expression image sequences with and without glasses. Details of dataset is available at ***https://nvie.ustc.edu.cn/***

Thermalfaceproject [33]: - A High resolution data at 1024 x 768 pixels database with a wide range of head poses instead of the usually fully frontal recordings provided .68 facial landmark points are manually annotated. All images for database were recorded using an Infratec HD820 high resolution thermal infrared camera with a 1024 x 768 pixel-sized microbolometer sensor equipped with a 30 mm f/1.0 prime lens. The database contains 2500 images of 90 subjects. Details of dataset is available at ***https://github.com/marcinkopaczka/thermalfaceproject***

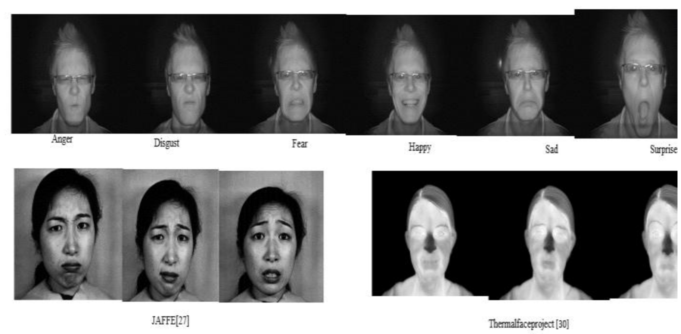

Figure 3: Dataset Samples from Oulu-CASIA [31] (Top row), Jaffe [30] (bottom row-left) and Thermalfaceproject [33] (Bottom-row Right)

IV. LITERATURE SURVEY

The advantage of deep networks is its end to end learning architecture just by feeding the input images without need of much expertise required in contrast of hand-crafted features. However, pre-processing techniques such as face detection, face alignment and head pose rectification have been applied to offer better performance which is already discussed in section II. As most FER is performed on RGB still images or videos, illumination variation remains a challenge. The performance of FER degrades in weak light condition or varying illumination conditions. In this context, He *et al*. [8] employed a DBM model on USTC- NVIE thermal infrared images. The thermal images are not sensitive to illumination changes and rely on heat radiations emitted by object.

Kopaczka et al. [33] proposed detection and tracking of facial landmarks in Long Wave Infrared (LWIR) spectrum using active appearance model (AAM) which are robust *against out-of-plane rotation and occlusions.* They obtained results leaving one-subject-out-(LOSO) cross-validation. In this type of validation all images of a given subject from the database are removed and algorithms are trained on all remaining subjects and tested their performance on the subject previously removed from the database. A fully annotated thermal face database is introduced in literature [33] which can be used for thermal FER. To evaluate the capacity of database different combinations of feature descriptor (LBP, HOG, DSIFT) and classifiers (SVM, Random Forest (RF), Linear Discriminant Analysis (LDA), K-NN) were tested. Out of various test performed DISFT and SVM obtained optimal results for the dataset.

Yang et al. [34] proposed the weighted mixture CNN architecture by fusing grayscale images and LBP images. They performed 10-fold cross validation, but they do not mention if there were images of the same subject in more than one-fold.

Fathallah et al. [35] have used VGG model with fine tuning strategy to overcome overfitting problem of deep networks. Hasani et al. [36] proposed spatio-temporal feature model that applies CNN for extracting spatial information and Long Short Term memory (LSTM), a type of Recurrent Neural Network(RNN) for extracting temporal characteristics to address expression intensity variations. LSTM has been widely adopted for extracting temporal features in FER.

Hernández et al. [37] defined regional descriptors of thermal infrared images using the gray level co-occurrence matrix (GLCM). In [38] a combination of thermal statistical features (StaFs), 2D-DCT and GLCM features is used, extracting both local and global information. Shen et al. [39] used infrared thermal videos by extracting horizontal and vertical temperature difference from different facial sub-regions. For FER, the Adaboost algorithm with the weak classifiers of k-Nearest Neighbour is used. In [40] authors integrated thermal infrared data with visible spectrum images for spontaneous facial expression recognition. They applied decision-level and feature-level fusion strategy using Bayesian network and SVM.

Wu *et a*l. [41] proposed three stream 3D convolution network to extract spatio-temporal features from video sequences using Oulu-CASIA Near Infrared datasets achieving accuracy of 78.42 %. The architecture fused global features and local features extracted using three Convolutional network streams in a fully connected layer. They used 10-fold cross-validation. With total 80 subjects in dataset, 72 people as the training set and 8 people as the test set. In all experiments, there are no overlapping images between the training sets and test sets

Trujillo et al. [42] generated eigen images from each region of interest from thermal images and uses the principal component values as features. In [43] authors used local parts eyes and lips for facial expression recognition using Bezier curve.

Although facial expression recognition has been studied in the literature, few works perform fair evaluation avoiding mixing subjects while training and testing the proposed algorithms.

Few literatures perform the cross-database experiments, where network was trained with few databases and accuracy evaluation is carried out using different FER datasets. A practical and fair evaluation method for facial expression recognition should guarantee no overlap of subjects in both training and testing sets at the same time.

V. CONCLUSION AND FUTURE WORK

A large number of FER deep learning techniques have been experimented in recent years mostly using visible light spectrum datasets. However, there are other modalities coming up making FER an active research topic. One of them is Infrared images which can be fused with visible face images or alone can be experimented for FER. While most of the work on thermal images is based on conventional approaches, there is still scope to do empirical study using deep learning techniques. The robustness of thermal infrared images against varying illumination condition and further advancement in thermal sensor technology paves path for exciting research in this area. Future work will be an attempt to fuse appearance and geometry features using thermal datasets which will be useful in identifying efficiency of Convolutional Neural Network using different modalities for same. Definitely, there is a bright chance of obtaining very stable and robust FER by fusing advantages of Deep learning techniques and Infrared and Thermal source images.

REFERENCES

[1] P. Ekman ; W.V Friesen,. "*Facial Action Coding System: Investigator's Guide*", 1st ed.; Consulting Psychologists Press: Palo Alto, CA, USA, pp. 1–15, 1978., ISBN 9993626619.

[2] D. Matsumoto, "*More evidence for the universality of a contempt expression*," Motivation and Emotion, vol.16, no.4, pp.363–368,1992.

[3] Happy, S.L.; George, A.; Routray, " *A real time facial expression classification system using local binary patterns*". In Proceedings of the 4th International Conference on Intelligent Human Computer Interaction, Kharagpur, India, pp. 1–5, 2012.

[4] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," IEEE transactions on pattern analysis and machine intelligence, vol. 29, no. 6, pp. 915–928, 2007.

[5] A.S. Mali, A.A. Kenjale, P.M. Ghatage, A.G. Deshpande, "*Mood based Music System*", International Journal of Scientific Research in Computer Science and Engineering, Vol.6, Issue.3, pp.27-30, 2018

[6] S. Ioannou, V. Gallese, and A. Merla, "*Thermal infrared imaging in psychophysiology: potentialities and limits*," Psychophysiology, vol. 51, no. 10, pp. 951–963, 2014.

[7] M. Kopaczka, K. Acar, and D. Merhof, "*Robust facial landmark detection and face tracking in thermal infrared images using active appearance models*," in International Conference on Computer Vision Theory and Applications (VISAPP), Rome, Italy, pp. 150–158, February 2016

[8] S. He, S. Wang, W. Lan, H. Fu and Q. Ji, "*Facial Expression Recognition Using Deep Boltzmann Machine from Thermal*

Infrared Images," 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, *Geneva*, pp. 239-244, 2013.

[9] 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*, Kauai, HI, USA, pp. I-I. 2001.

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

[11] H. Jung, S. Lee, J. Yim, S. Park, and J. Kim, "*Joint fine-tuning in deep neural networks for facial expression recognition*," in Computer Vision (ICCV), 2015 IEEE International Conference on. IEEE, pp. 2983–2991, 2015.

[12] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, pp. 1867-1874, 2014.

[13] P. Simard, D. Steinkraus, J. C. Platt," *Best practices for convolutional neural networks applied to visual document analysis"*, in Proceedings Seventh International Conference on Document Analysis and Recognition, pp. 958–963, 2003.

[14] A.T.Lopes, E.de Aguiar, A.F.DeSouza, and T.Oliveira-Santos,''*Facial expression recognition with convolutional neural networks: Coping with few data and the training sample order*,'' Pattern Recognition., vol. 61, pp. 610–628, Jan. 2017.

[15] D. A. Pitaloka, A. Wulandari, T. Basaruddin, and D. Y. Liliana, "*Enhancing cnn with preprocessing stage in automatic emotion recognition*," Procedia Computer Science, vol. 116, pp. 523–529, 2017.

[16] W. Li, M. Li, Z. Su, and Z. Zhu, "*A deep-learning approach to facial expression recognition with candid images*," in Machine Vision Applications (MVA), 2015 14th IAPR International Conference on. IEEE, pp. 279–282,2015.

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

[18] Maas, Andrew L.. "*Rectifier Nonlinearities Improve Neural Network Acoustic Models*." (2013).

[19] K. He, X. Zhang, S. Ren, and J. Sun, "*Delving deep into rectifiers: Surpassing human-level performance on imagenet classification*," in Proceedings of the IEEE international conference on computer vision, pp. 1026–1034, 2015.

[20] S. Ioffe and C. Szegedy, "*Batch normalization: Accelerating deep network training by reducing internal covariate shift*," arXiv preprint arXiv:1502.03167, 2015

[21] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "*Improving neural networks by preventing co-adaptation of feature detectors*," arXiv preprint arXiv:1207.0580, 2012(dropouts)

[22] K. Simonyan and A. Zisserman, "*Very deep convolutional networks for large-scale image recognition*," arXiv preprint arXiv:1409.1556, 2014.

[23] C. Szegedy *et al*., "*Going deeper with convolutions*," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, pp. 1-9, 2015.

[24] K. He, X. Zhang, S. Ren, and J. Sun, "*Deep residual learning for image recognition*," in Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778, 2016

[25] O'Shea, Keiron & Nash, Ryan. "*An Introduction to Convolutional Neural Networks*", ArXiv e-prints.2015

[26] Peng, M.; Wang, C.; Chen, T." *NIRFaceNet: A Convolutional Neural Network for Near-Infrared Face Identification*." Information vol. 7, no. 4:61, 2016

[27] Corneanu, C.A; Simon, M.O.; Cohn, J.F.; Guerrero, S.E. "*Survey on RGB, 3D, thermal, and multimodal approaches for facial expression recognition: History, trends, and affect-related applications*". IEEE Trans. Pattern Anal. Mach. Intell., 38, 1548–1568, 2016.

[28] Zhang, Wei & Zhang, Youmei & Ma, Lin & Guan, Jingwei & Gong, Shijie." *Multimodal learning for facial expression recognition* "Pattern Recognition, 48, 10.1016/j.patcog.2015.04.012, (2015).

[29] 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, pp. 94–101, 2010.

[30] M. J. Lyons, S. Akamatsu, M. Kamachi, J. Gyoba, and J. Budynek, "*The japanese female facial expression (jaffe) database*," 1998.

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

[32] S. Wang, Z. Liu, S. Lv, Y. Lv, G. Wu, P. Peng, F. Chen, and X. Wang, "*A natural visible and infrared facial expression database for expression recognition and emotion inference*," IEEE Transactions on Multimedia, vol. 12, no. 7, pp. 682–691, 2010.

[33] M. Kopaczka, R. Kolk and D. Merhof, "*A fully annotated thermal face database and its application for thermal facial expression recognition*" *2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, Houston, TX, pp. 1-6, 2018.

[34] B. Yang, J. Cao, R. Ni and Y. Zhang, "*Facial Expression Recognition Using Weighted Mixture Deep Neural Network Based on Double-Channel Facial Images*," in *IEEE Access*, vol. 6, pp. 4630-4640, 2018.

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

[36] B. Hasani and M. H. Mahoor, "*Facial expression recognition using enhanced deep 3d convolutional neural networks,*" in Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on. IEEE, pp. 2278–2288, 2017

[37] Hernández B, Olague G, Hammoud R, Trujillo L, Romero E. "*Visual learning of texture descriptors for facial expression recognition in thermal imagery*". Computer Vision and Image Understanding. 2007.

[38] S. Wang, M. He, Z. Gao, S. He, and Q. Ji, "*Emotion recognition from thermal infrared images using deep boltzmann machine*," FCS, vol. 8, no. 4, pp. 609–618, 2014.

[39] Shen, P.; Wang, S.; Liu, Z. "*Facial expression recognition from infrared thermal videos"*. Intell. Auton. Syst, 12, 323–333, 2013.

[40] Wang, S., He, S., Wu, Y. et al. "Fusion of visible and thermal images for facial expression recognition" Front. Comput. Sci. 8: 232-242, 2014.

[41] Z. Wu, T. Chen, Y. Chen, Z. Zhang, and G. Liu, "*Nirexpnet: Three stream 3d convolutional neural network for near infrared facial expression recognition*," Applied Sciences, vol. 7, no. 11, pp. 1184-1197, 2017

[42] L. Trujillo, G. Olague, R. Hammoud, and B. Hernandez, "*Automatic feature localization in thermal images for facial expression recognition*," in Computer Vision and Pattern Recognition (CVPR) Workshops, pp. 14–14, 2005.

[43] G.Sowmiya, V. Kumutha, "*Facial Expression Recognition Using Static Facial Images*", International Journal of Scientific Research in Computer Science and Engineering, Vol.6, Issue.2, pp.72-75, 2018

Authors Profile

Mrs. Reshma Nehlani pursued Bachelor of Engineering in Information Technology from Gujarat University in 2009.Currently she is pusuing Masters of Engineering from Gujarat Technological University.Her main research interest includes Image processing and Machine learning.

Dr. Devang Pandya purused his PhD programme in Computer Engineering from Sardar National Institute of Technology in 2016, Surat, India. He is serving as an Assistant Professor in Information and Communication technology department at Adani Institute of Infrastrcture Engineering, Ahmedabad. His main reaserch interest includes Image processing, Machine Learning, Video Processing and retrieval.