LEARNET: Dynamic Imaging Based Micro Expression Recognition

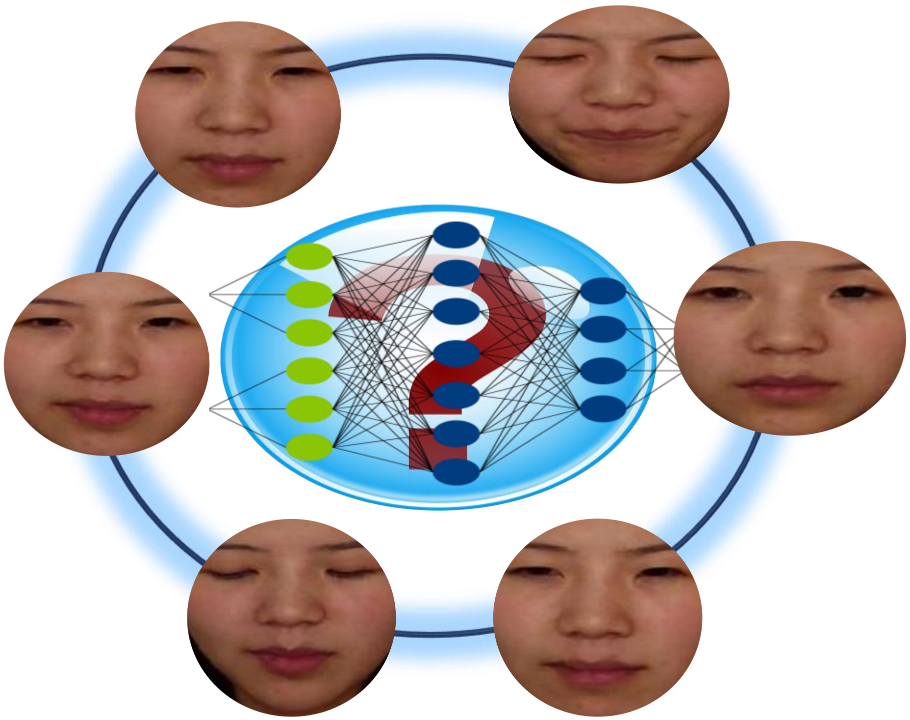
(Accepted in IEEE Transaction on Image Processing, Impact Factor: 5.071)

By: Monu Verma

Micro expressions

Dynamic Imaging

Short Falls in Existing Nets



LearNet Architecture

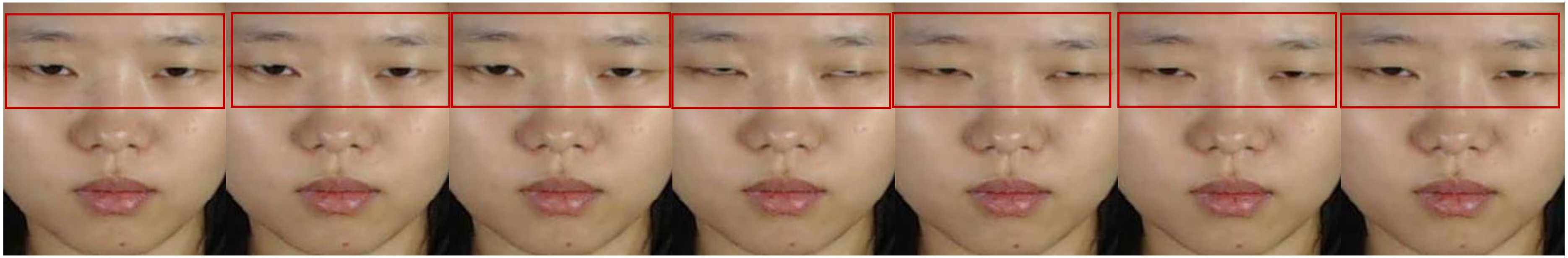
Properties of LearNet

Qualitative Analysis

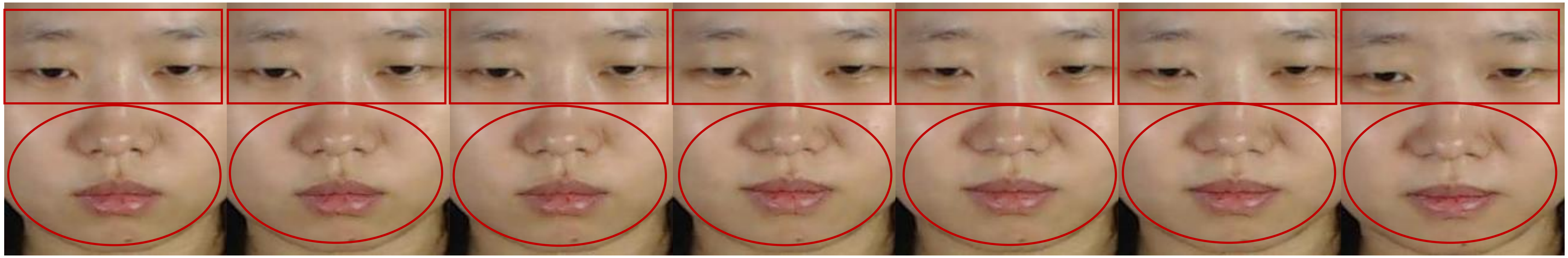
Quantitative Analysis

Conclusion

Micro Expressions



Disgust Expression



Happy expression

Dynamic Imaging

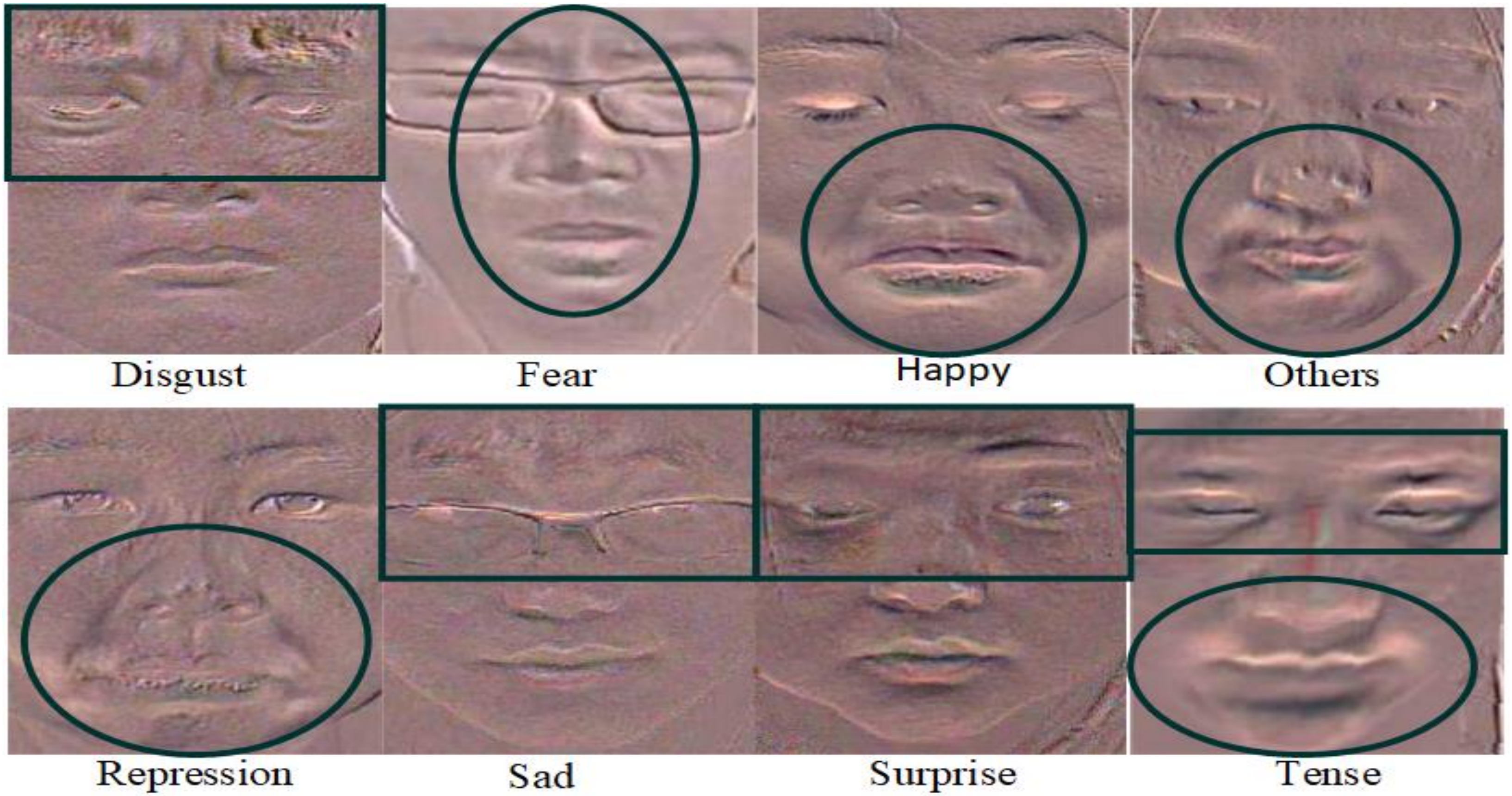


Fig.1. Dynamic responses of different micro-expressions.

Short Falls in Conventional Networks

- As dynamic images of micro expressions hold minute variations within the image sequences, existing networks like VGG-16, VGG-19 [12] and ResNet [15] fail to spot these variations.
- These networks usually follow sequential coupling mechanism with dense depth maps. Such an approach sometimes ignore the minute features favoring more visually distinguishable features.
- Conventional CNN-based networks are uses max polling to down sample the input image. Pooling layer extracts the maximum response features by the performing max operation over embedded filters. Thus, max pooling layer also neglects the micro-variation of the facial images.
- Existing networks have large computational cost as they uses large number of learning parameters.

LEARNet Architecture

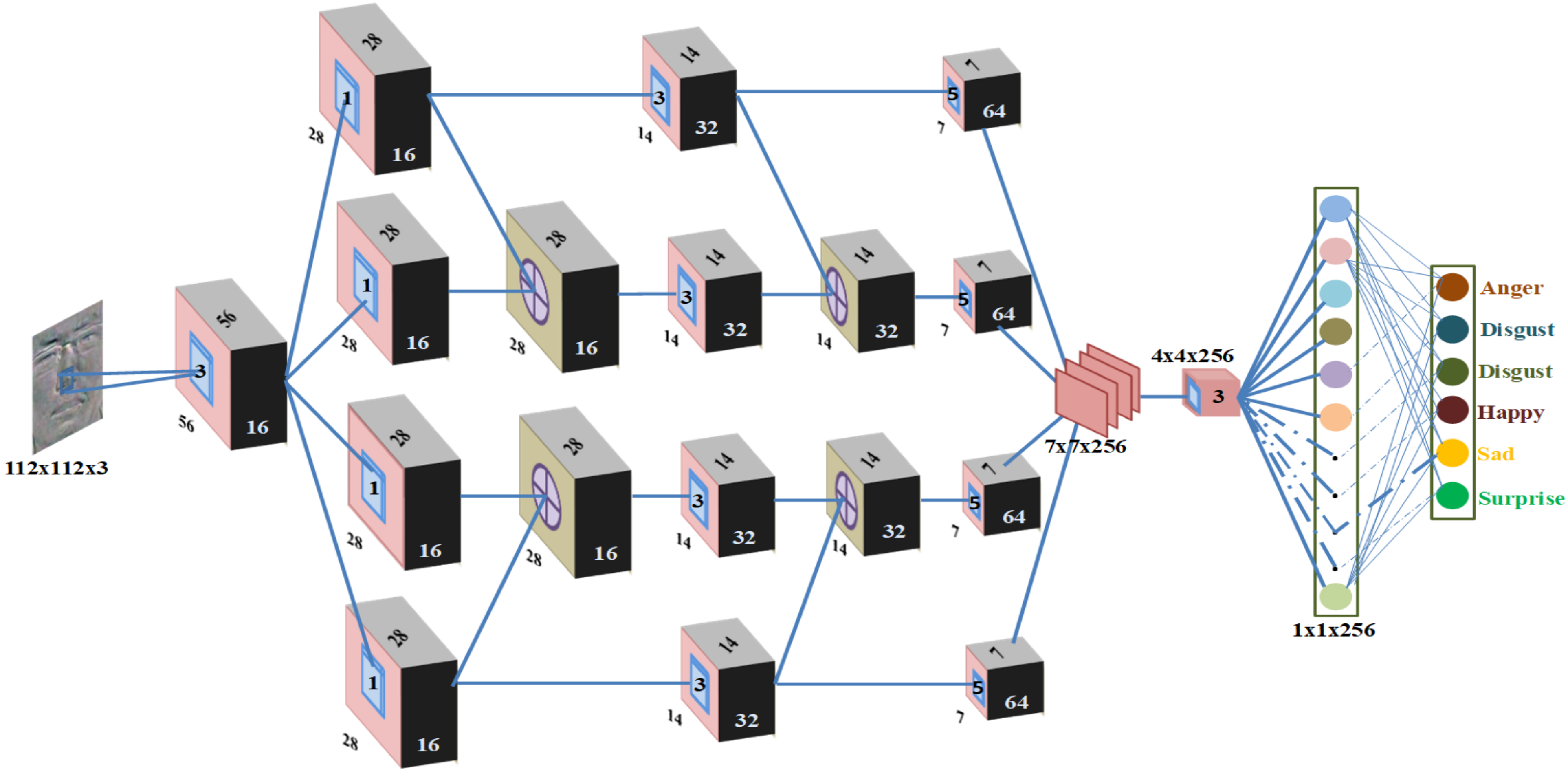


Fig. 2. Proposed LEARNet architecture

LEARNet Architecture

TABLE VII
LEARNet Detailed Configuration

Type	Sub-layer	Filter	Stride	Output	#Parameters (w+b)
Input	-	-	-	112x112x3	-
Conv- 1	-	3 3	2	56x56x16	432+16
Conv- 2	2.1	1 1	2	28x28x16	4 (256+16)
	2.2				
	2.3				
	2.4				
Add- 1	1.1	-	-	28x28x16	-
	1.2				
Conv- 3	3.1	3 3	2	14x14x32	4 (4608+32)
	3.1				
	3.3				
	3.4				
Add- 2	2.1	-	-	14x14x32	-
	2.2				
Conv- 4	4.1	5 5	2	7x7x64	4 (51200+64)
	4.2				
	4.3				
	4.4				
Concat	-	-	-	7x7x256	-
LRN	-	-	-	7x7x256	256+256
Conv- 5	-	3 3	2	4x4x256	589824+256
FC	-	-	-	1x1x256	589824+256

Properties of LearNet

- LEARNet model captures more detailed features by using the decoupled feature map mechanism, which help in preserving the minute facial muscle change information.

- LearNet utilize the hybrid feature approach by incorporating an accretion layer to extend the network in accretive way.

- Accretion layer combines the hybrid responses which are generated by previous layers. These layers enhance the learnability of the neurons for minute details and maintain the essence of the feature maps

- EXPERTNet included convolution layer with stride 2, which decrease the size of input with minimum information loss.

Qualitative Analysis

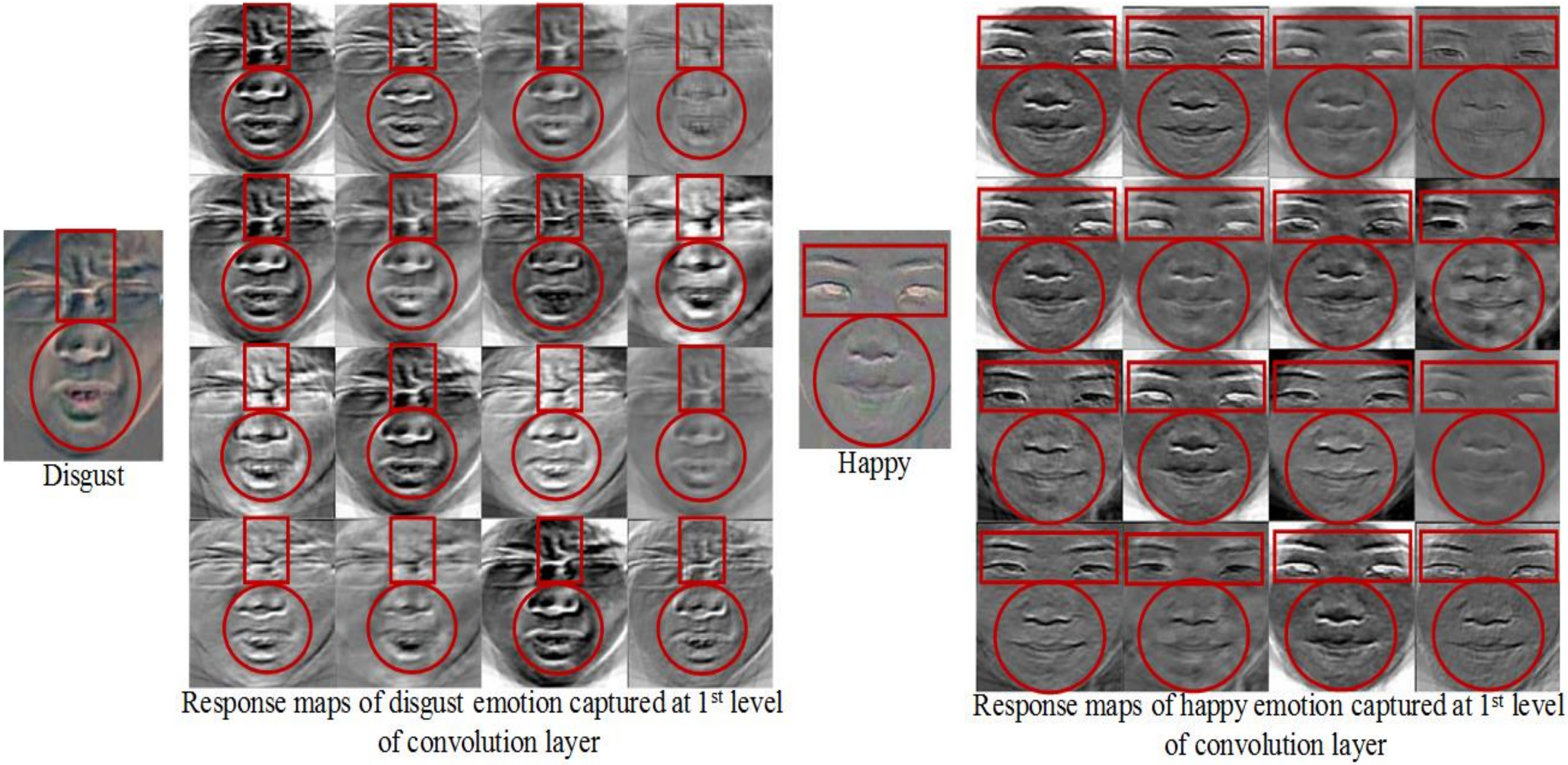


Fig. 3. Response maps of two different emotion classes a) disgust and b) happy, captured at 1st level of the convolution layer.

Qualitative Analysis

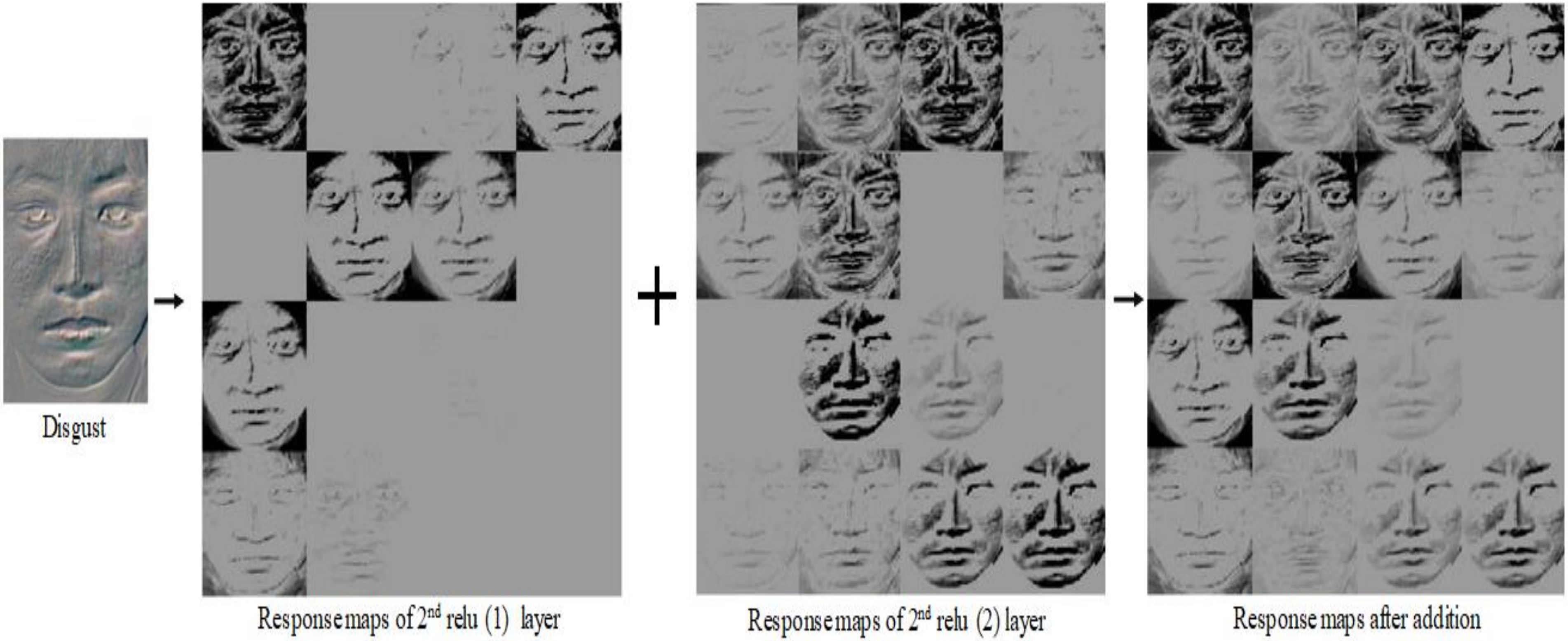


Fig. 4. Visualization of neuron responses for disgust emotion triggered by: a) Conv- 2.1 b) Conv- 2.2 and c) accretion response.

Comparative Analysis

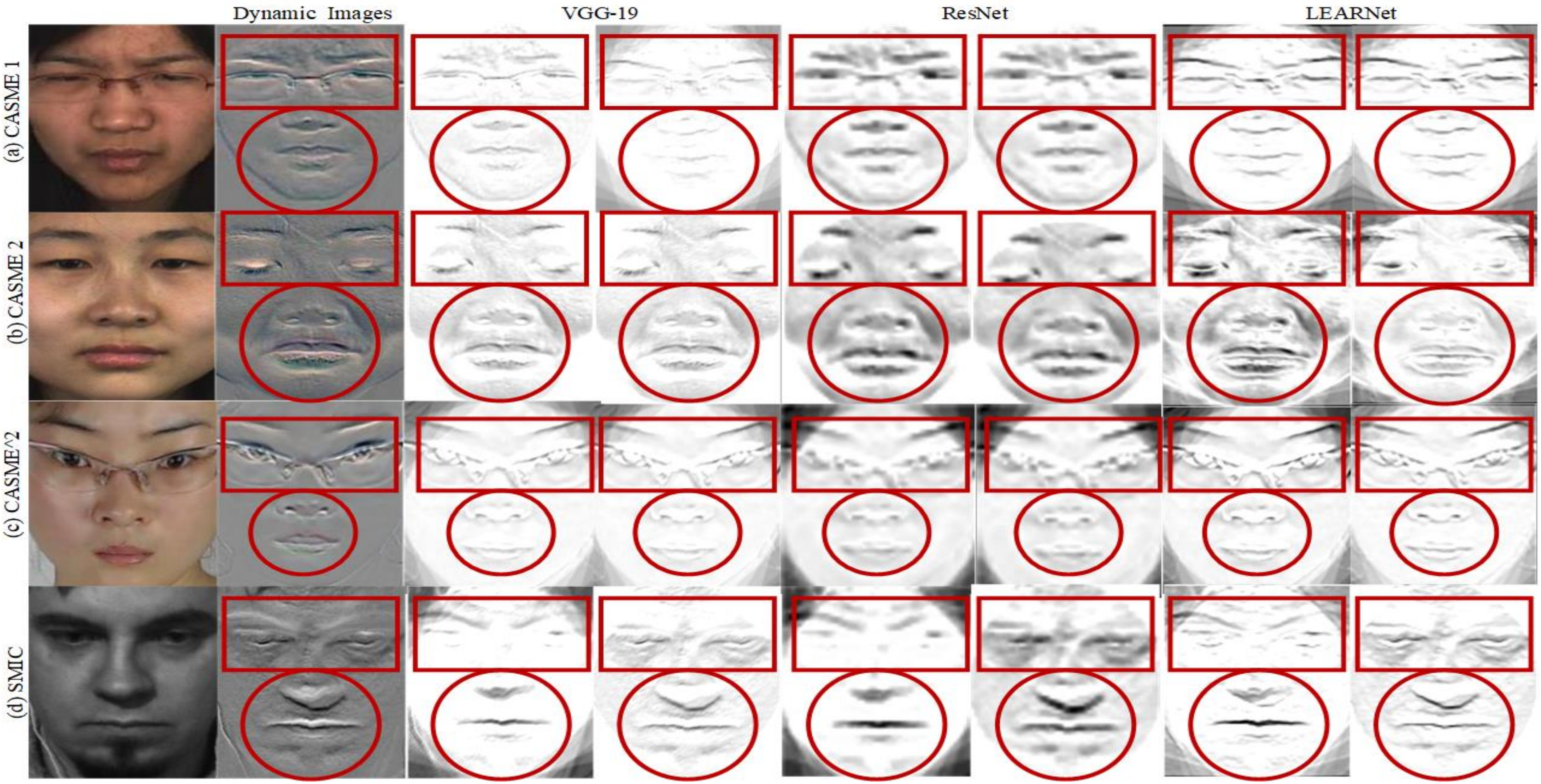


Fig. 5. Visual comparison of existing model and LEARNEet over different expression of four datasets a) CASME-I: Tension b) CASME-II: Happy c) CAS(ME)^2: Anger and d) SMIC: surprise.

Quantitative Analysis

TABLE VIII
RECOGNITION ACCURACY COMPARISON ON CASME-I, CASME-II AND CAS(ME)^2 DATASET

Method	CASME-I	CASME-II	CAS(ME)^2
LBP-TOP-SVM*	64.07	57.16	-
LBP-TOP-ELM *	73.82	-	-
MDMO-SVM*	68.86	67.37	-
CNN-LSTN*	-	60.98	-
VGG-16	36.59	44.29	44.29
VGG-19	36.59	44.29	44.28
ResNet	76.39	74.49	74.48
LEARNet	80.42	76.82	76.27

TABLE IX
Recognition Accuracy Comparison on SMIC Dataset

Method	5-Class	2-Class
LBP-TOP-SVM*	71.40	-
MDMO-SVM *	80.00	-
VGG-16	36.59	51.53
VGG-19	36.59	51.53
ResNet	71.36	88.27
LEARNet	82.66	91.09

*Results are taken from the original papers

Computational Analysis

TABLE X
Computational Complexity analysis of LEARNet and existing Networks

Network	# Layers	# Parameters (in millions)
VGG-16 [12]	16	138
VGG- 19 [12]	19	144
GoogleNet [13]	22	4
ResNet [15]	34	11
LEARNet	**14**	**1.4**

Conclusion

- We have generated dynamic images from micro expression sequence which captures the facial movements in one frame.
- The proposed architecture adopts hybrid and decoupled feature learning mechanism to learn the salient features from the expressive regions captured in the past layers.
- LEARNet uses different sized filters i.e. 1x1, 3x3 and 5x5, which enhance the capability of network by extracting micro and high-level features.
- Proposed network includes the accretion layer to merge the features of two response maps that allow to expose pertinent features robustly.

References

[1] P. Ekman, (1993). Facial expression and emotion. *American Psychologist*, *48*(4), 384.
[2] P. Ekman and W.V. Friesen, (1977). Facial action coding system.
[3] W. V. Friesen and P. Ekman, (1983). EMFACS-7: Emotional facial action coding system. *Unpublished manuscript, University of California at San Francisco*, *2*(36), 1.
[4] I. A. Essa and A. P. Pentland, (1997). Coding, analysis, interpretation, and recognition of facial expressions. *IEEE transactions on pattern analysis and machine intelligence*, *19*(7), 757-763.
[5] I. Kotsia and I. Pitas, (2007). Facial expression recognition in image sequences using geometric deformation features and support vector machines. *IEEE transactions on image processing*, *16*(1), 172-187.
[6] D. Gabor, (1946). Theory of communication. Part 1: The analysis of information. *Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering*, *93*(26), 429-441.
[7] C. Shan, S. Gong and P.W. McOwan (2009). Facial expression recognition based on local binary patterns: A comprehensive study. *Image and vision Computing*, *27*(6), 803-816.
[8] C. C. Lai and C. H. Ko, (2014). Facial expression recognition based on two-stage features extraction. *Optik-International Journal for Light and Electron Optics*, *125*(22), 6678-6680.
[9] T. Jabid, M. H. Kabir and O. Chae, (2010). Robust facial expression recognition based on local directional pattern. *ETRI journal*, *32*(5), 784-794.
[10] A. R. Rivera, J. R. Castillo and O. O. Chae, (2013). Local directional number pattern for face analysis: Face and expression recognition. *IEEE transactions on image processing*, *22*(5), 1740-1752.
[11] A. R. Rivera, J. R. Castillo and O. Chae (2015). Local directional texture pattern image descriptor. *Pattern Recognition Letters*, *51*, 94-100.
[12] B. Ryu, A. R. Rivera, J. Kim and O. Chae, (2017). Local directional ternary pattern for facial expression recognition. *IEEE Transactions on Image Processing*, *26*(12), 6006-6018.
[13] A. Krizhevsky, I. Sutskever and G. E Hinton, (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, 1097-1105.
[14] K. Simonyan and A. Zisserman (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
[15] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2818-2826.
[16] K. He, X. Zhang, S. Ren and J. Sun (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770-778.

References

[17] A. Mollahosseini, D. Chan and M. H. Mahoor, (2016). Going deeper in facial expression recognition using deep neural networks. In *Applications of Computer Vision (WACV),* 1-10.

[18] B. Hasani and M. H. Mahoor, (2017). Spatio-temporal facial expression recognition using convolutional neural networks and conditional random fields. In *12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017),* 790-795.

[19] H. Jung, S. Lee, S. Park, I. Lee, C. Ahn and J. Kim, (2015). Deep temporal appearance-geometry network for facial expression recognition. *arXiv preprint arXiv:1503.01532*.

[20] P. Khorrami, T. Paine, and T. Huang, (2015). Do deep neural networks learn facial action units when doing expression recognition?. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 19-27.

[21] H. Ding, S. K. Zhou and R. Chellappa, (2017). Facenet2expnet: Regularizing a deep face recognition net for expression recognition. In *12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017),* 118-126.

[22] Y. Kim, B. Yoo, Y. Kwak, C. Choi and J. Kim, (2017). Deep generative-contrastive networks for facial expression recognition. *arXiv preprint arXiv:1703.07140*.

[23] P. Burkert, F. Trier, M. Z. Afzal, A. Dengel and M. Liwicki, (2015). Dexpression: Deep convolutional neural network for expression recognition. *arXiv preprint arXiv:1509.05371*.

[24] K. Zhang, Y. Huang, Y. Du and L. Wang, (2017). Facial expression recognition based on deep evolutional spatial-temporal networks. *IEEE Transactions on Image Processing*, *26*(9), 4193-4203.

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

[26] M. Pantic, M. Valstar, R. Rademaker and L. Maat, (2005,). Web-based database for facial expression analysis. In *IEEE international conference on multimedia and Expo*, 5.

[27] M. Valstar and M. Pantic, (2010). Induced disgust, happiness and surprise: an addition to the mmi facial expression database. In *Proc. 3rd Intern. Workshop on EMOTION (satellite of LREC): Corpora for Research on Emotion and Affect*, 65.

[28] S. M. Mavadati, M. H. Mahoor, K. Bartlett, P. Trinh and J. F. Cohn, (2013). Disfa: A spontaneous facial action intensity database. *IEEE Transactions on Affective Computing*, *4*(2), 151-160.

[29] M. F. Valstar, B. Jiang, M. Mehu, M. Pantic and K. Scherer, (2011). The first facial expression recognition and analysis challenge. In *IEEE International Conference on Automatic Face & Gesture Recognition and Workshops (FG 2011),* 921-926.

[30] P.Viola, and M. J. Jones, (2004). Robust real-time face detection. *International journal of computer vision*, *57*(2), 137-154.

Thank You

QUERIES ?