Multimedia Systems (2021) 27:615–625
https://doi.org/10.1007/s00530-020-00705-1

SPECIAL ISSUE PAPER

Image super-resolution model using an improved deep learning-based facial expression analysis

Pyoung Won Kim[1]

Received: 15 July 2020 / Accepted: 4 October 2020 / Published online: 13 October 2020

Abstract

Image upsampling and noise removal are important tasks in digital image processing. Single-image upsampling and denoising influence the quality of the resulting images. Image upsampling is known as super-resolution (SR) and referred to as the restoration of a higher-resolution image from a given low-resolution image. In facial expression analysis, the resolution of the original image directly affects the reliability and validity of the emotional analysis. Hence, optimization of the resolution of the extracted original image during emotion analysis is important. In this study, a model is proposed, which applies an image super-resolution method to an algorithm that classifies emotions from facial expressions.

Keywords Face · Facial expression analysis · Image super-resolution · Emotions

1 Introduction

The human face provides many essential signals for interpersonal communication. The face can demonstrate spoken languages and controls conversations by gazing or nodding. Thus, someone's emotional state and intentions can be determined through facial expressions [1]. Classification of emotions through facial expression is based on the anatomy-based principle that emotions are revealed in facial expressions. The method of classifying facial expressions into action units (AU) and analyzing emotions using a combination of AU has been widely verified for reliability and validity [2].

Since human coder analysis of AU and classification of emotions take a tremendous amount of time, using a facial expression engine like FaceReaderTM is preferred. In facial expression analysis, the resolution of the original image directly affects the reliability and validity of the emotional analysis. Single-image super-resolution (SR) is an old research domain with remarkable results in recent years due to its deep learning (DL) adoption.

✉ Pyoung Won Kim
pwkim@inu.ac.kr

[1] Department of Korean Language Education, College of Education, Incheon National University, 12 Gaetbeol-ro, Yeonsu-gu, Incheon, Republic of Korea

This study proposes a model, which applies an image SR method to an algorithm that classifies emotions from facial expressions and an alternative methodology for assessing emotional states. The rest of this paper is organized as follows. An overview of the facial expression and emotion classification is presented in Sect. 2. Section 3 shows the face reading and emotion classification technology. In Sect. 4, a facial expression analysis model using DL method is described. Finally, Sect. 5 contains some conclusive remarks.

2 Facial expression and emotion classification

2.1 Facial behavior coding system

The interest in facial expressions has a long history and the universality hypothesis debate about human emotions has continued to this day [3]. The universality hypothesis explains that all humans communicate six basic emotional states (i.e., happy, surprise, fear, disgust, anger, and sad state) using the same facial movements across all cultures [4].

Classification of emotions through facial expression is based on the anatomy-based principle that emotions are revealed in facial expressions [1]. The function of the amygdala in the limbic system is to control physiological arousal, which is embodied in facial expressions or responses from the autonomic nervous system. Facial expressions and emotions are closely intertwined [5]. Almost all facial muscles are innervated by a single nerve. Facial expressions are caused by the movement of the numerous muscles which are supplied by the facial nerve that are attached to and move the facial skin [6] (Fig. 1).

Research on facial expression usually monitors three groups of facial muscles: the corrugator supercilii muscle that pulls the inner half of each eyebrow downward, the zygomaticus major muscle that controls the movement of the mouth to create a smile-like expression, and the frontalis muscle that pulls the eyebrows upwards [6] (Fig. 2).

The facial expressions are not only external manifestations of emotions, but they can also trigger or modulate emotional experiences. This is commonly known as the facial feedback hypothesis (FFH) [7]. The standard method of measuring emotion through facial expressions is the facial behavior coding system (FACS). FACS allows the human coder to evaluate emotions based on observable AU, which are facial movements that describe facial and emotional expressions [1]. All facial expressions are described through a combination of AU with associated intensities (Table 1).

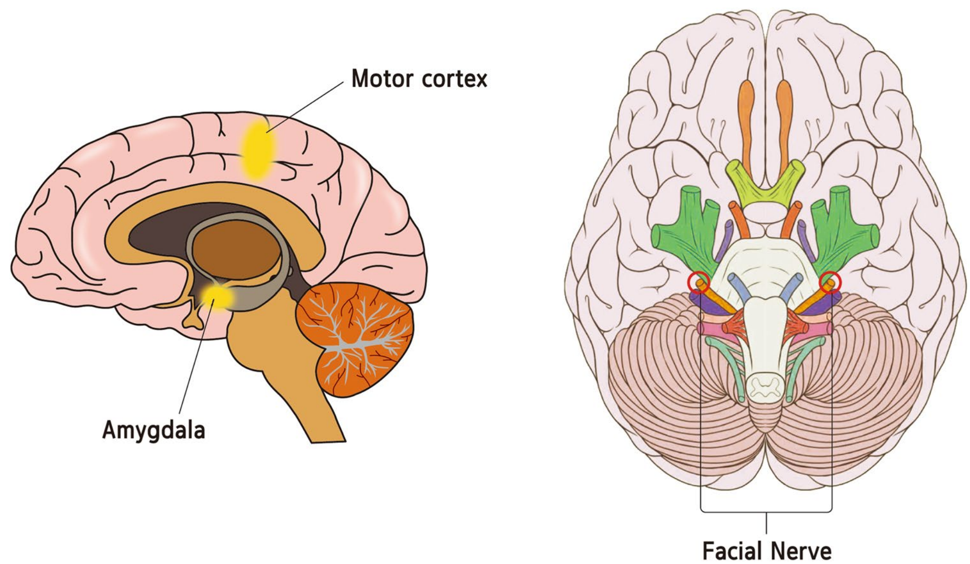


Fig. 1 Amygdala and facial nerve

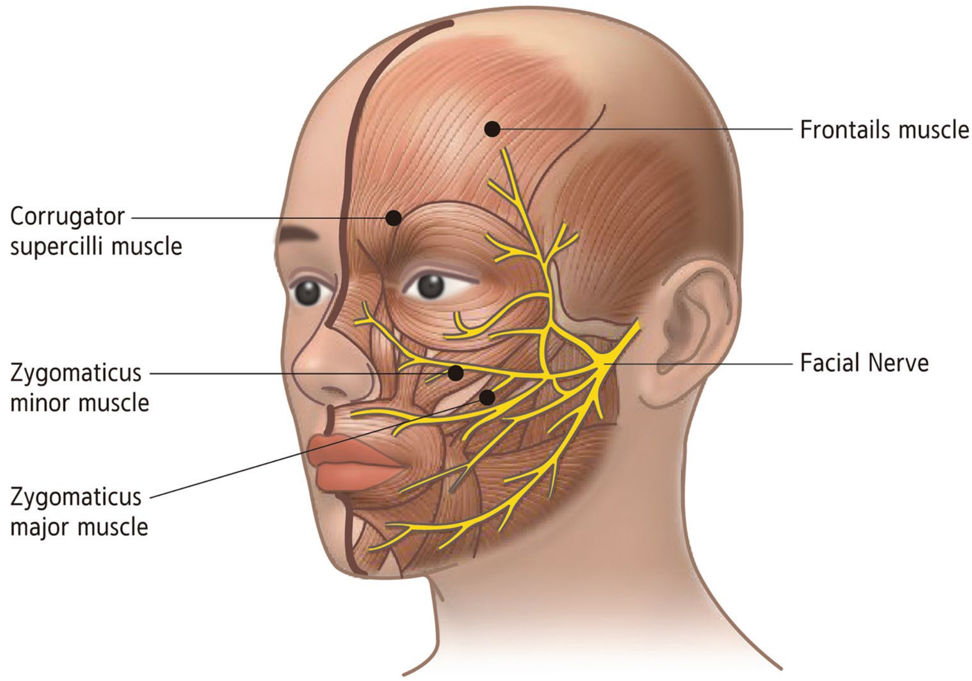
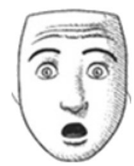


Fig. 2 Three groups of facial muscles

Table 1 FACS action units (AU) for basic emotions

Basic emotions	McCloud's expression	Action units	Facial expression
Happiness		AU 1	Inner corner of eyebrow raised
		AU 6	Cheek raiser and lid compressed
		AU 12	Lip corner pulled
		AU 14	Dimpled
Surprise		AU 1	Inner corner of eyebrow raised
		AU 2	Outer corner of eyebrow raised
		AU 5	Upper lid raised
		AU 15	Lip corner depressed
		AU 16	Lower lip depressed
		AU 20	Lip stretched
		AU 26	Jaw dropped
Fear		AU 1	Inner corner of eyebrow raised
		AU 2	Outer corner of eyebrow raised
		AU 4	Brow lowered
		AU 5	Upper lid raised
		AU 15	Lip corner depressed
		AU 20	Lip stretched
		AU 26	Jaw dropped
Disgust		AU 2	Outer corner of eyebrow raised
		AU 4	Brow lowered
		AU 9	Nose wrinkled
		AU 15	Lip corner depressed
		AU 17	Chin raised
Anger		AU 2	Outer corner of eyebrow raised
		AU 4	Brow lowered
		AU 7	Lid tightened
		AU 9	Nose wrinkled
		AU 10	Upper lip raised
		AU 20	Lip stretched
		AU 26	Jaw dropped
Sadness		AU 1	Inner corner of eyebrow raised
		AU 4	Brow lowered
		AU 15	Lip corner depressed
		AU 23	Lip tightened

The method of classifying facial expressions into AU and analyzing emotions using a combination of AU has been widely verified for reliability and validity.

2.2 Classification of emotions

Psychologists have attempted to determine the diverse range of human emotions using few independent factors. Early theorists charted emotions in two or more dimensions [8, 9]. The two-dimensional circumplex model of emotion is particularly popular among emotional models. Russell's circumplex model of affect can be defined as arousal-valence space model, where emotions arise from the interaction between two-dimensional factors: one representing the degree of pleasantness, ranging from positive to negative valence, and the other representing arousal [8] (Fig. 3).

Arousal measures the level or amount of physical response intensity, while valence determines whether the emotional direction is good or bad. Plutchik's wheel of emotions uses color to express the level of arousal.

Plutchik's model shows the visual tool for understanding eight primary emotions in polar opposite pairs, as shown in Fig. 4. In the emotion model of Plutchik, the high and low

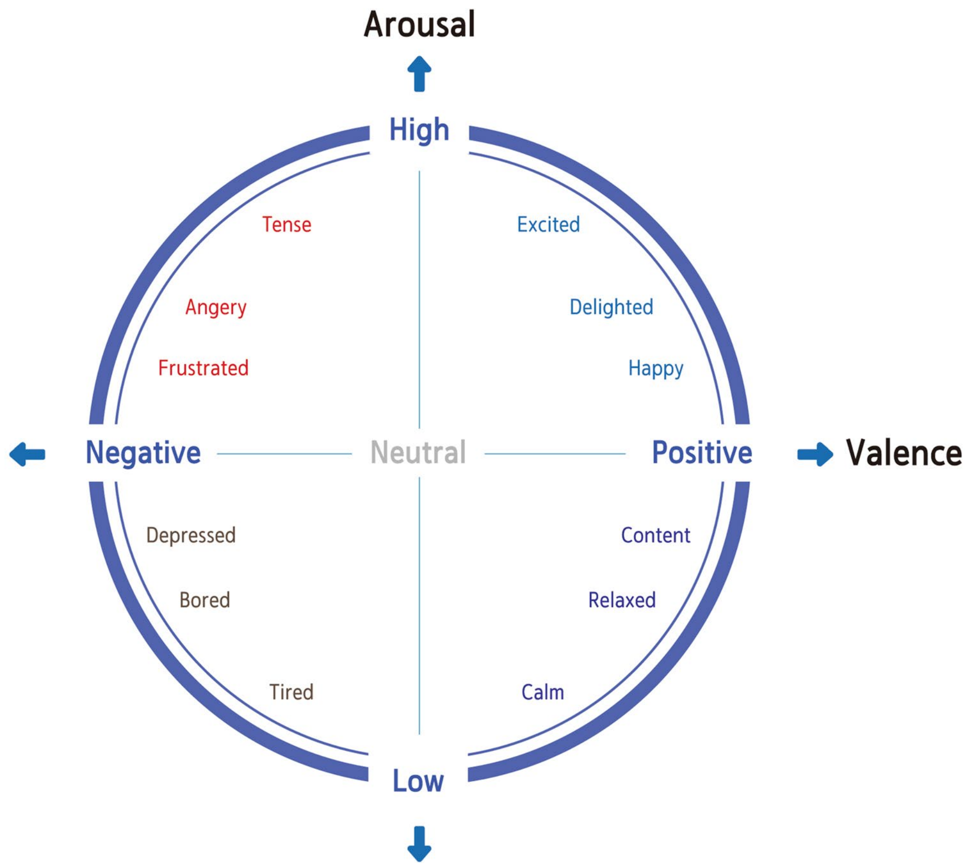


Fig. 3 Russell's circumplex model of affect

levels of arousal are explained by changes in saturation, as shown in Fig. 5.

Several methods are used to assess the levels of arousal: observation of viewer behavior, self-reporting, biofeedback measurement, and subject interviews [10, 11]. Bio-signals provide objective measurements via electroencephalograms, electrocardiograms, blood pressure monitoring, or skin conductance. A range of bio-signals has been extensively investigated in the fields of physiology and engineering.

Arousal influences the autonomic nervous system, as shown in Fig. 6, and is associated with physiological changes, such as increases in heart rate, perspiration, muscle tension, and rapid respiration. As arousal increases, a sympathetic discharge in the nervous system is activated [12]. As the blood supply to the heart increases, there is a corresponding decrease in blood flow in the digestive system, which can result in diarrhea or constipation [6]. These responses enable arousal to be expressed as specific bio-signals and also reflect the degree of arousal. Electrodermal activity (EDA) or the galvanic skin response measures the sweat gland response on the palm of the hand, which is closely associated with emotional arousal [13]. Skin conductance is often used as an indicator of effective processes and emotional arousal, because this indicator is inexpensive, unobtrusive, and reliable to measure [13]. As stress or the immersion level increases, more sweat is produced, which is reflected by an increase in the EDA. A decrease in stress or immersion level also reflects a decrease in both sweat activity and EDA.

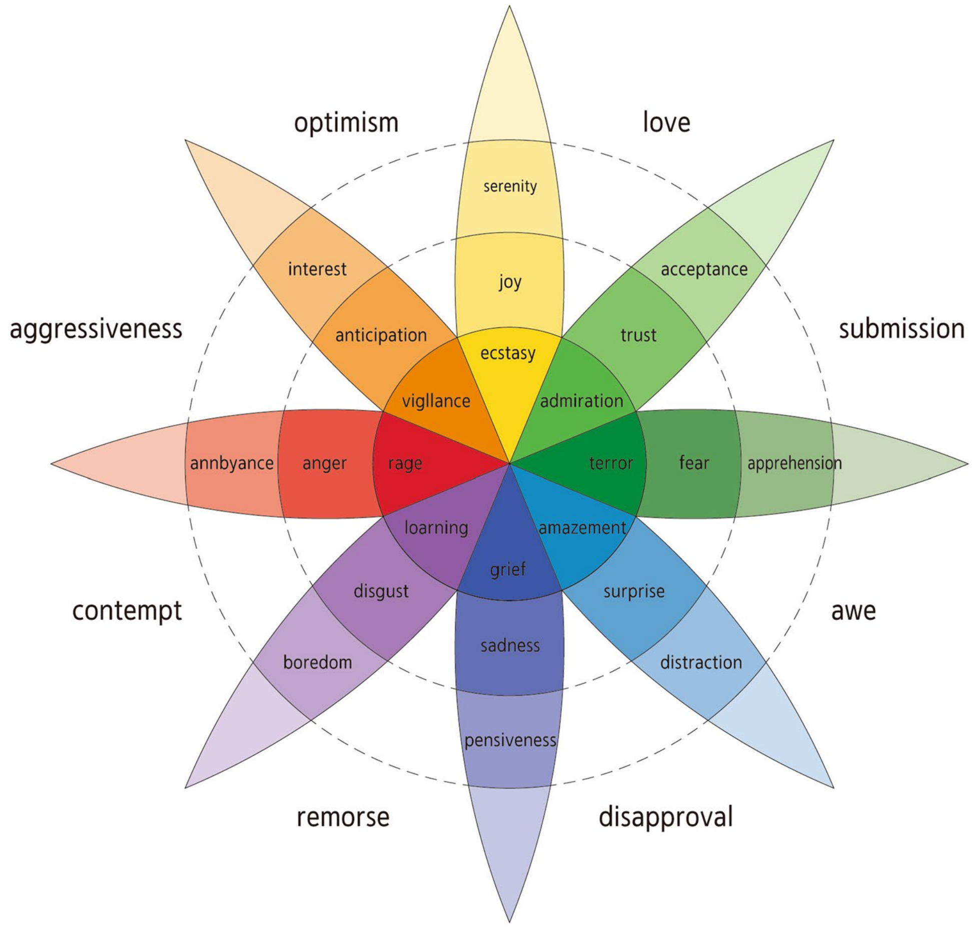
3 Face reading and emotion classification technology

Since human coder analyzes AU and the classification of emotions takes a tremendous amount of time, use of a facial expression engine becomes imperative. FaceReader, AFFDEX, and FACET are representative facial expression engines that are three algorithms used for classifying emotions from facial expressions [14]. Among these algorithms, FaceReader is the most widely used system and can recognize facial expressions with an accuracy of 90% [15].

3.1 Face recognition

FaceReader architecture consists of three consecutive steps: face finding, face modeling, and face classification. Face

Fig. 4 Plutchik's emotion wheel



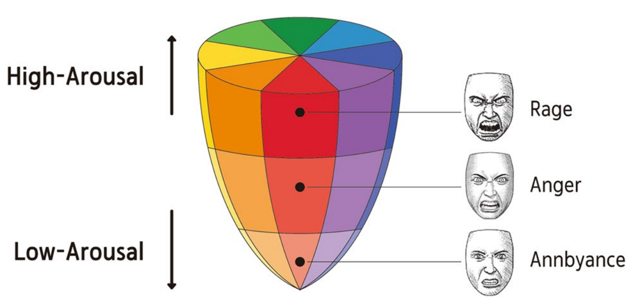


Fig. 5 Plutchik's emotion model

finding uses a method called the Viola–Jones cascade classifier, which minimizes computation time while achieving high detection accuracy [16] (Fig. 7).

The Viola–Jones algorithm was the most innovative and effective way to recognize human faces until DL technologies such as convolutional neural networks became popular.

3.2 Face modeling

The second step of face recognition is to synthesize a newly recognized face image with the active appearance model (AAM), which is an artificial face model that describes the location of 500 major points on the face [17]. The AAM is a technology used to find the shape of a specific face using shape and appearance models. The AAM uses a set of annotated images to calculate the main sources of variation found in face images and uses principal component analysis to reduce the model dimensionality. New faces can then be described as deviations from the mean face using a vector (Fig. 8).

3.3 Face classification

Humans can quickly read other person's facial expressions and recognize emotions, but transferring this skill to an algorithm is very complex. Machines are rather taught to recognize emotions instead of explicitly programing them to do so. The final stage of the FaceReader architecture is emotion classification, which is performed by training an artificial neural network that uses the vectors produced in the second step as input [18].

Over 10,000 manually annotated images have been used as training materials for the FaceReader's classification algorithm. The network was trained to classify the seven basic or

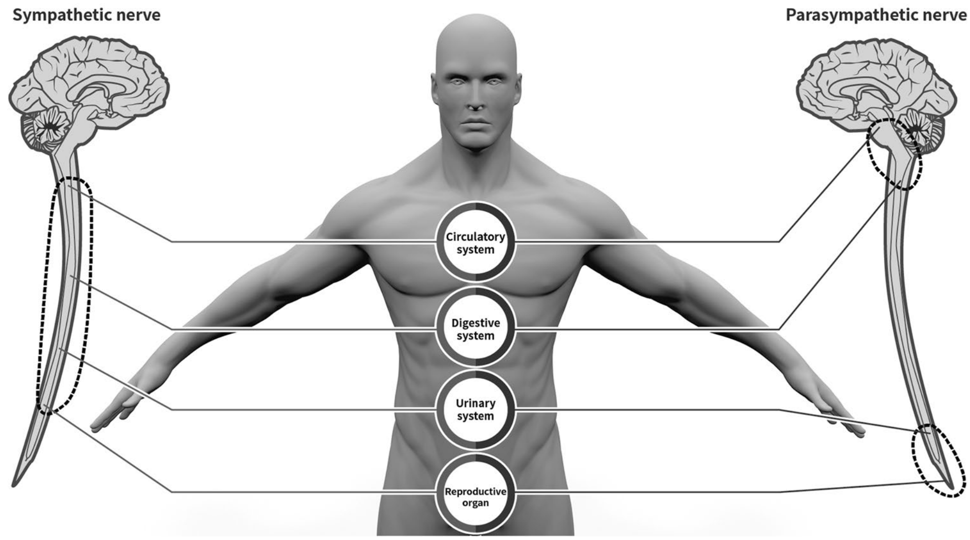


Fig. 6 The autonomic nervous system

universal emotions described by Ekman, such as happy, surprise, fear, disgust, anger, sad, and neutral state [1] (Fig. 9).

4 Facial expression analysis model using deep learning

4.1 Measurement of emotional valence using image super-resolution

Considering the mechanism of analyzing AU and classifying emotions, the accuracy of valence can be increased during preprocessing using the image SR method and DL and then supplying the result to an existing facial expression engine as input.

Single-image upsampling and denoising are essential in image processing [19–22]. The principal purpose of single-image upsampling is to reconstruct a high-resolution image from an existing low-resolution image [23]. Upsampling can be used in image applications such as medical image analysis, video surveillance, video processing and streaming, and in television [24–26]. Many single-image upsampling methods have been proposed with each method having a different level of complexity and performance based on the application goal. The nearest neighbor, bilinear (BI), and bicubic (BC) methods are well-known interpolation methods among them. Single-image SR is a research domain with remarkable results in recent years due to its DL adoption.

Both the speed and emotion classification of image SR have limitations in processing data in real time, so the model needs to be improved to realize the processing speed. Ultimately, DL model for image SR must to be integrated with emotion classification.

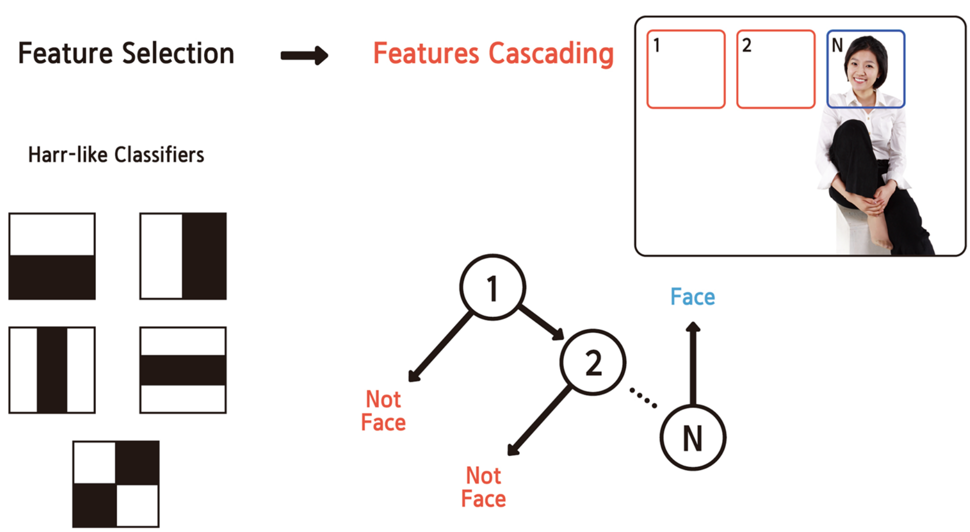


Fig. 7 Viola–Jones cascade classifier algorithm

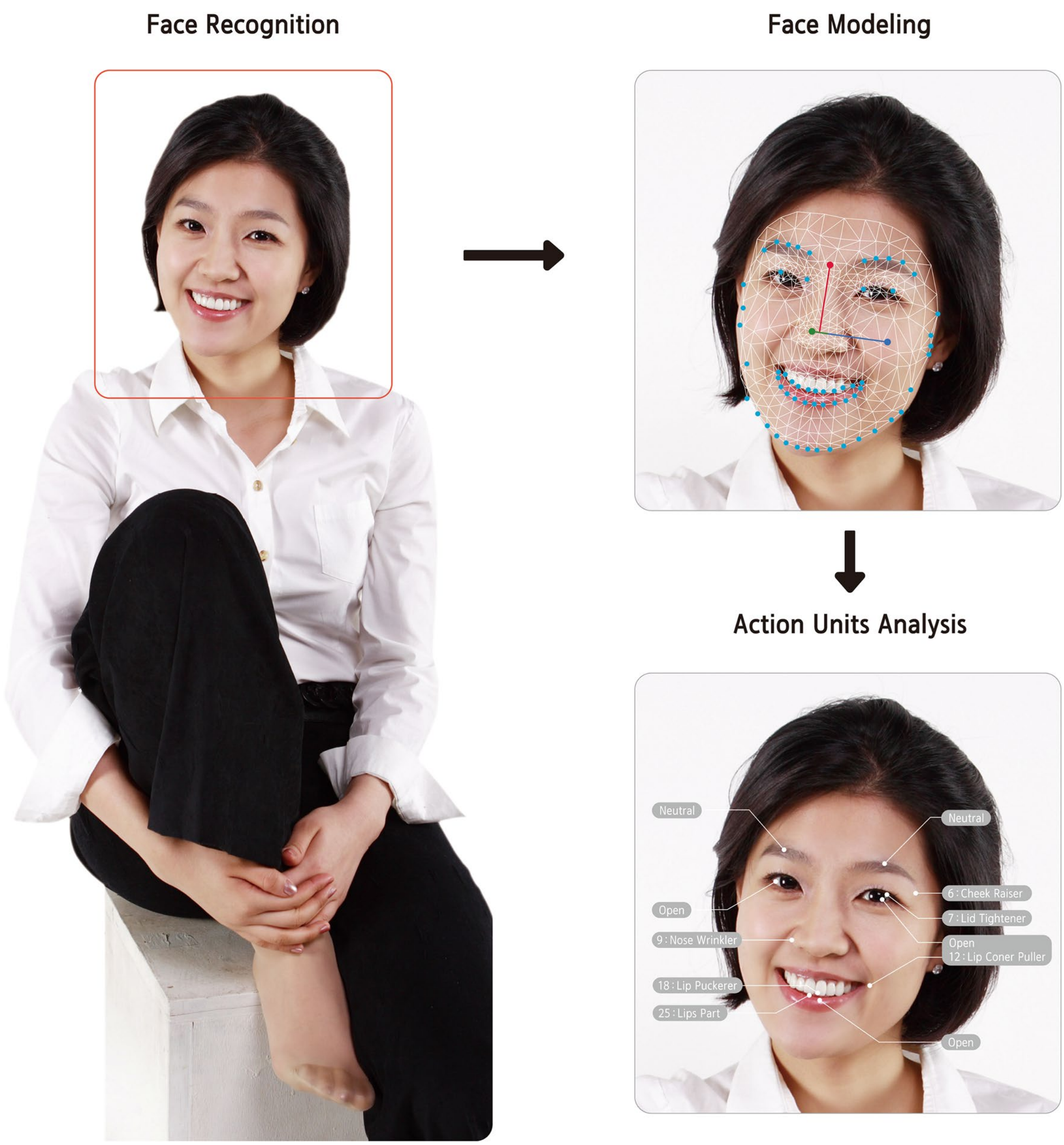


Fig. 8 Face modeling using active appearance model

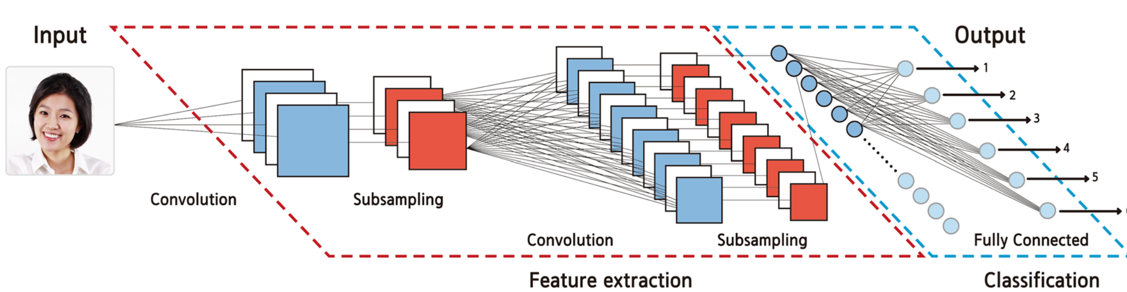


Fig. 9 Face classification using an artificial neural network

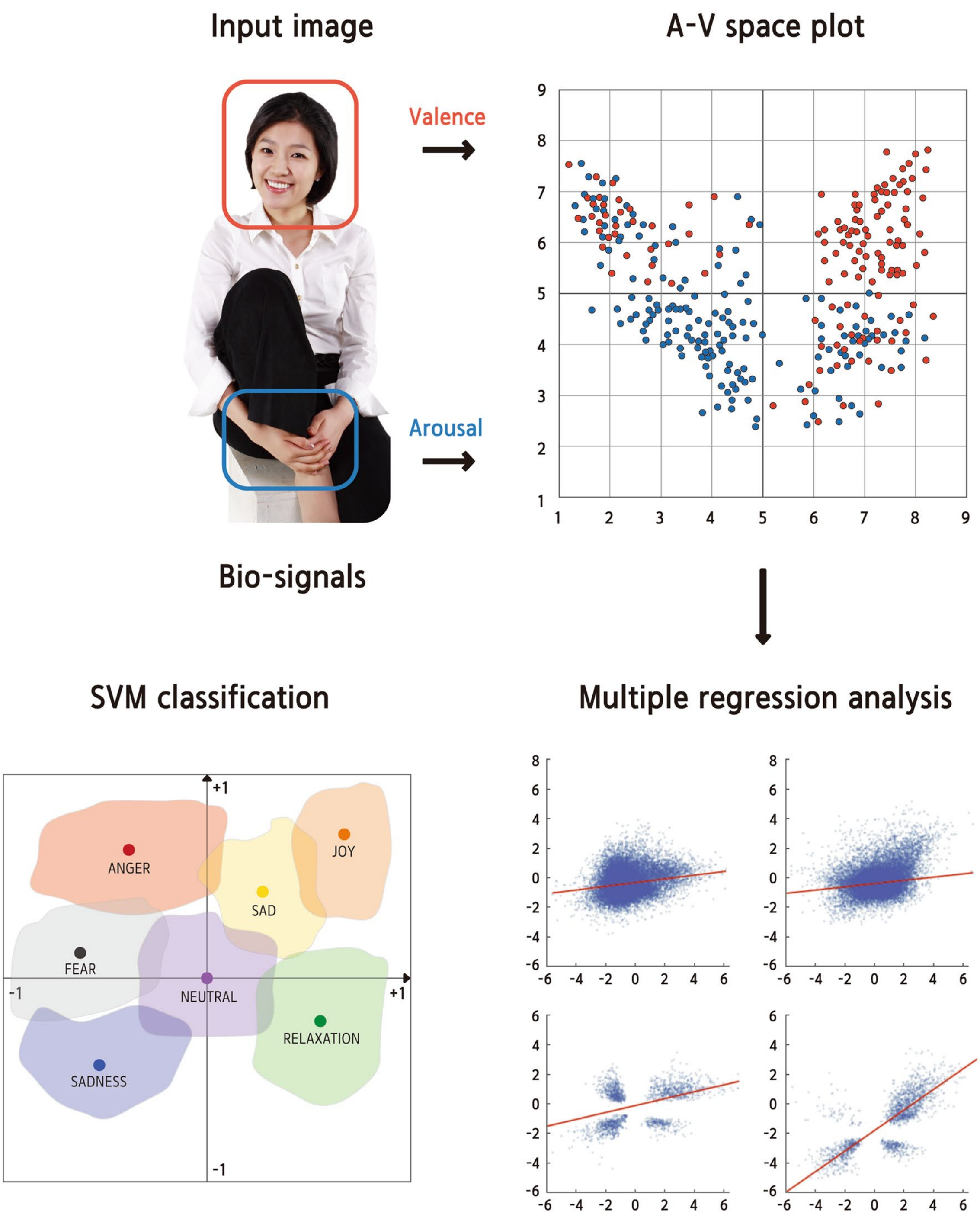


Fig. 10 Face classification using support vector machines

4.2 Proposed deep learning model for measuring bio-signals

By integrating the arousal level obtained by measuring bio-signals with existing facial expression engines, building an accurate emotional model as shown in Fig. 10 is possible.

Figure 11 shows the emotion classification map for support vector machines (SVM) using McCloud's facial expression images [27]. Using the SVM, automatic classification of emotions associated with new images is possible by mapping

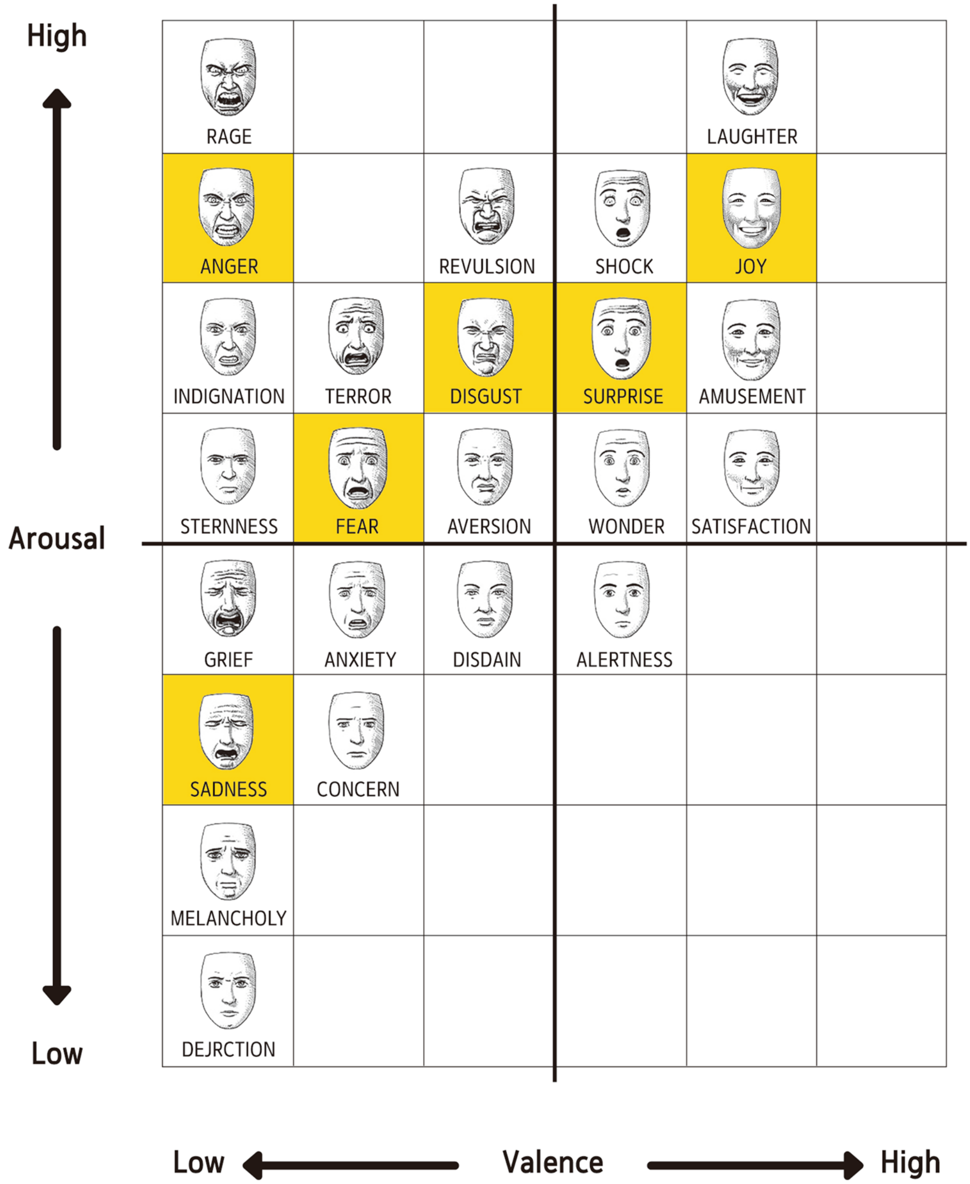


Fig. 11 Different facial expressions on the arousal-valence space for SVM classification

the results obtained via multiple regression analysis to an emotion classification map.

Since SVM is a discrete model that classifies emotions, increasing the accuracy of emotion classification using a neural network model to build a more sophisticated emotion estimation algorithm is possible. In particular, building a model that evaluates the intensity of emotion classification in real time is possible by utilizing various bio-signal time-series data including brain waves with excellent time resolution, as shown in Fig. 12.

5 Conclusions

Various DL models for image preprocessing continue to evolve. In this work, two models based on an existing strategy to improve the quality of emotions classification from facial expressions using an image SR method based on DL is proposed. The first suggestion is to separate the variables of the model that analyze emotions into two dimensions: facial expressions and bio-signals. The second strategy is the addition of a DL process that analyzes

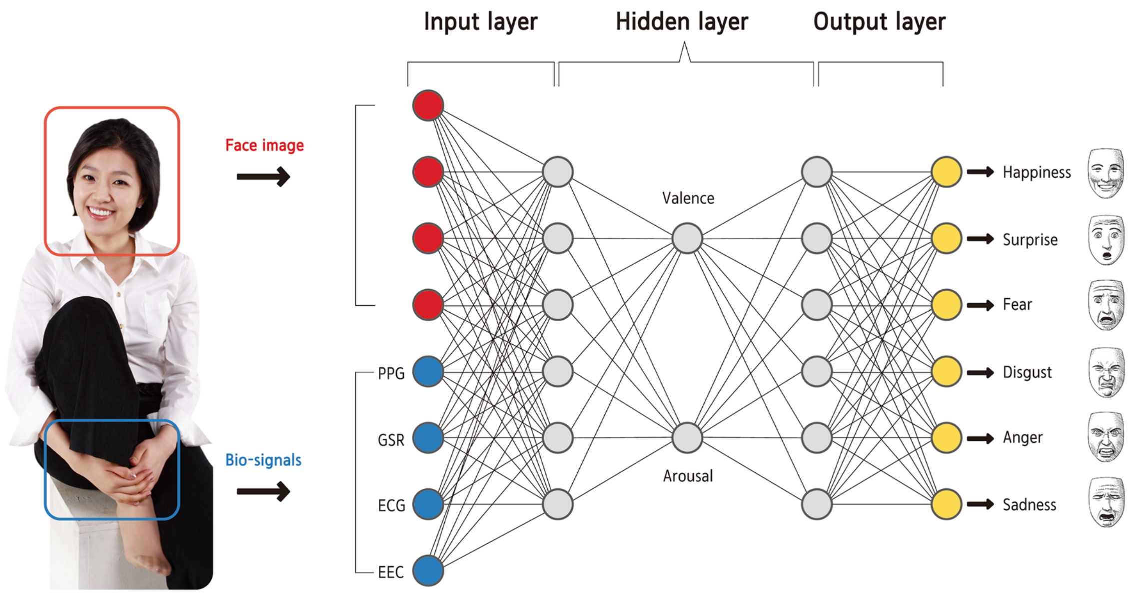


Fig. 12 Proposed deep learning model for facial expression engines

facial expressions and another process that analyzes bio-signals. The most efficient way is to create a DL model that integrates the process of image SR and the process of classifying emotions. In future, the proposed model will be developed into an integrated emotional analysis model based on neural network.

Acknowledgements This work was supported by the Incheon National University Research Grant in 2020.

References

1. Ekman, P., Keltner, D.: Facial expression of emotion. In: Lewis, M. and Haviland-Jones, J. (eds). Handbook of Emotions, pp. 236–249. New York, (2000)
2. Navaz, A., Adel, S., Mathew, S.: Facial image pre-processing and emotion classification: a deep learning approach, IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA). Abu Dhabi, United Arab Emirates, pp.1–8 (2019)
3. Darwin, C.: The Expression of the Emotions in Man and Animals. John Murray, London (1872)
4. Jack, R., Garrod, O., Yu, H., Caldara, R., Schyns, P.: Facial expressions of emotion are not culturally universal. PNAS **109**(19), 7241–7244 (2012)
5. Namburi, P., Al-Hasani, R., Calhoon, G., Bruchas, M., Tye, K.: Architectural representation of valence in the limbic system. Neuropsychopharmacology **41**(7), 1697–1715 (2016)
6. Fox, S.: Human Physiology. McGraw-Hill Higher Education, New York (2018)
7. Tomkins, S.: Affect, Imagery, and Consciousness. Springer, New York (1962)
8. Russell, J.: A circumplex model of affect. J. Pers. Soc. Psychol. **39**(6), 1161–1178 (1980)
9. Plutchik, R.: The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. Am. Sci. **89**(4), 344–350 (2001)
10. Dillard, J., Wilson, B.: Communication and effect. Commun Res **20**, 637–646 (1993)
11. van der Walma Molen, J., Klijn, M.: Recall of television versus print news: retesting the semantic overlap hypothesis. J. Broadcast Electron. Media **48**, 89–107 (2010)
12. Lang, A., Dhillon, K., Dong, Q.: Effects of emotional arousal and valence on television viewers' cognitive capacity and memory. J. Broadcast Electron. Media **39**(3), 313–327 (1995)
13. Figner, B., Murphy, R.: Using skin conductance in judgment and decision making research. In: Schulte-Mecklenbeck, M., Kuehberger, A., Ranyard, R. (eds.) A Handbook of Process Tracing Methods for Decision Research, pp. 163–184. Psychology Press, New York (2011)
14. Stöckli, S., Schulte-Mecklenbeck, M., Borer, S.: Facial expression analysis with AFFDEX and FACET: a validation study. Behav. Res **50**, 1446–1460 (2018)
15. Noldus Information Technology: Reference Manual, Wageningen, 2015
16. Viola, P., Jones, M.: Robust real-time face detection. Int. J. Comput. Vision **57**(2), 137–154 (2004)
17. Cootes T., Taylor, C.: Statistical models of appearance for computer vision. Technical report, University of Manchester, Wolfson Image Analysis Unit, Imaging Science and Biomedical Engineering (2000)
18. Bishop, C.: Neural Networks for Pattern Recognition. Clarendon Press, Oxford (1995)
19. Hardie, R., Barnard, K., Armstrong, E.: Joint map registration and high-resolution image estimation using a sequence of undersampled images. IEEE Trans. Image Process **6**(12), 1621–1633 (1997)

20. Tipping, M., Bishop, C.: Bayesian image super-resolution. Conf. Proc. Adv. Neural Inf. Syst **16**, 1303–1310 (2003)
21. Sun, J., Xu, Z., Shum, H.: Image super-resolution using gradient profile prior. Conf Proc IEEE Comput. Vis. Pattern Recognit. pp, 1–8 (2008)
22. Chang, H., Yeung, D., Xiong, Y.: Super-resolution through neighbor embedding. Conf Proc IEEE Comput. Vis. Pattern Class. 1, pp. 275–282 (2004)
23. Baker, S., Kanade, T.: Hallucinating faces, Conf Proc IEEE Int. Autom. Face Gesture Recognit. pp. 83–88 (2000)
24. Wu, J., Anisetti, M., Wu, W., Damiani, E., Jeon, G.: Bayer demosaicking with polynomial interpolation. IEEE Trans. Image Process. **25**(11), 5369–5382 (2016)
25. Paul, A.: Real-time power management for embedded M2M using intelligent learning methods. ACM Trans. Embedded Comput. Syst. **13**(5), 148 (2014)
26. Paul, A., Rho, S., Bharanitharan, K.: Interactive scheduling for mobile multimedia service in M2M environment. Multimed. Tools Appl. **71**(1), 235–246 (2014)
27. McCloud, S.: Making Comics: Storytelling Secrets of Comics, Manga and Graphic Novels. HarperCollins, New York (2006)