INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 01, January 2024, pp : 408-414

e-ISSN : 2583-1062

Impact Factor : 5.725

A REVEIW ON FACIAL RECOGNITION FOR SURVIVAL

Mohammad Rafi[1], Syeda Ayesha Siddiqua[2], Yashodha Gk[3]

[1,2,3]Department of Computer Science and Engineering, UBDTCE, India.

ABSTRACT

Deep learning models currently achieve human levels of performance on real-world face recognition tasks. We review scientific progress in understanding human face processing using computational approaches based on deep learning. This review is organized around three fundamental advances. First, deep networks trained for face identification generate a representation that retains structured information about the face (e.g., identity, demographics, appearance, social traits, expression) and the input image (e.g., viewpoint, illumination).

Researchers have found that compared with other existing conditions (e.g., pleasantness), information relevant to survival produced a higher rate of retrieval; this effect is known as the survival processing advantage (SPA). Previous experiments have examined that the advantage of memory can be extended to some different types of visual pictorial material, such as pictures and short video clips, but there were some arguments for whether face stimulus could be seen as a boundary condition of SPA[1].

1. INTRODUCTION

The fields of vision science, computer vision, and neuroscience are at an unlikely point of convergence. Deep convolutional neural networks (DCNNs) now define the state of the art in computer-based face recognition and have achieved human levels of performance on real-world face recognition tasks (Jacquet & Champod 2020, Phillips et al. 2018, Taigman et al. 2014). This behavioral parity allows for meaningful comparisons of representations in two successful systems. DCNNs also emulate computational aspects of the ventral visual system (Fukushima 1988, Krizhevsky et al. 2012, LeCun et al. 2015) and support surprisingly direct, layer-to-layer comparisons with primate visual areas (Yamins et al. 2014). Nonlinear, local convolutions, executed in cascaded layers of neuron-like units, form the computational engine of both biological and artificial neural networks for human and machine-based face recognition. Enormous numbers of parameters, diverse learning mechanisms, and high-capacity storage in deep networks enable a wide variety of experiments at multiple levels of analysis, from reductionist to abstract.[12-15] This makes it possible to investigate how systems and subsystems of computations support face processing tasks.[14].

Our goal is to review scientific progress in understanding human face processing with computational approaches based on deep learning. As we proceed, we bear in mind wise words written decades ago in a paper on science and statistics: “All models are wrong, but some are useful” (Box 1979, p. 202) (see the sidebar titled Perspective: Theories and Models of Face Processing and the sidebar titled Caveat: Iteration Between Theory and Practice). Since all models are wrong, in this review, we focus on what is useful. For present purposes, computational models are useful when they give us insight into the human visual and perceptual system. This review is organized around three fundamental advances in understanding human face perception, using knowledge generated from deep learning models[4-9].

Human Versus Machine Face Recognition

Deep learning models of face identification map widely variable images of a face onto a representation that supports identification accuracy comparable to that of humans.[5] The steady progress of machines over the past 15 years can be summarized in terms of the increasingly challenging face images that they can recognize (Figure 1). By 2007, the best algorithms surpassed humans on a task of identity matching for unfamiliar faces in frontal images taken indoors (O’Toole et al. 2007). By 2012, well-established algorithms exceeded human performance on frontal images with moderate changes in illumination and appearance (Kumar et al. 2009, Phillips & O’Toole 2014). Machine ability to match identity for in-the-wild images appeared with the advent of DCNNs in 2013–2014.[11] Human face recognition was marginally more accurate than DeepFace (Taigman et al. 2014), an early DCNN, on the Labeled Faces in the Wild (LFW) data set (Huang et al. 2008). LFW contains in-the-wild images taken mostly from the front. DCNNs now fare well on in-the-wild images with significant pose variation (e.g., Maze et al. 2018, data set). Sengupta et al. (2016) found parity between humans and machines on frontal-to-frontal identity matching but human superiority on frontal-to-profile matching.[8]

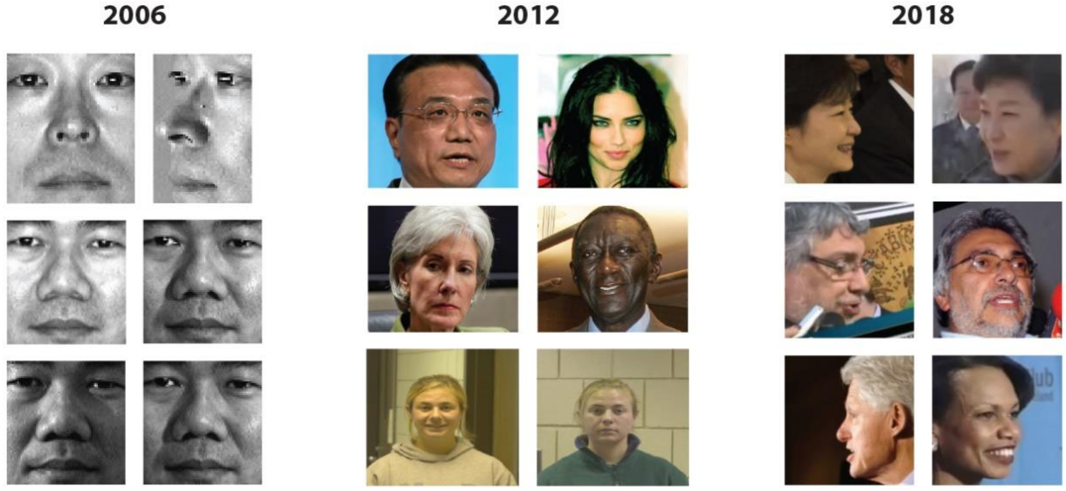

Figure.1

The progress of computer-based face recognition systems can be tracked by their ability to recognize faces with increasing levels of image and appearance variability.[7-9] In 2006, highly controlled, cropped face images with moderate variability, such as the images of the same person shown, were challenging (images adapted with permission from Sim et al. 2002). In 2012, algorithms could tackle moderate image and appearance variability (the top 4 images are extreme examples adapted with permission from Huang et al. 2012; the bottom two images adapted with permission from Phillips et al. 2011). By 2018, deep convolutional neural networks (DCNNs) began to tackle wide variation in image and appearance, (images adapted with permission from the database in Maze et al. 2018). In the 2012 and 2018 images, all side-by side images show the same person except the bottom pair of 2018 panels.

Expert Humans and State-of-the-Art Machines Work Together

DCNNs can sometimes even surpass normal human performance. Phillips et al. (2018) compared humans and machines matching the identity of faces in high-quality frontal images. Although this is generally considered an easy task, the images tested were chosen to be highly challenging based on previous human and machine studies.[9-12] Four DCNNs developed between 2015 and 2017 were compared to human participants from five groups: professional forensic face examiners, professional forensic face reviewers, super recognizers (Noyes et al. 2017, Russell et al. 2009), professional fingerprint examiners, and students. Face examiners, reviewers, and super recognizers performed more accurately than fingerprint examiners, and fingerprint examiners performed more accurately than students.[1-7] Machine performance, from 2015 to 2017, tracked human skill levels. The 2015 algorithm (Parkhi et al. 2015) performed at the level of the students; the 2016 algorithm (Chen et al. 2016) performed at the level of the fingerprint examiners (Ranjan et al. 2017c); and the two 2017 algorithms (Ranjan et al. 2017,c) performed at the level of professional face reviewers and examiners, respectively.[15]

Unimodal DCNN-DNN model for depression recognition.

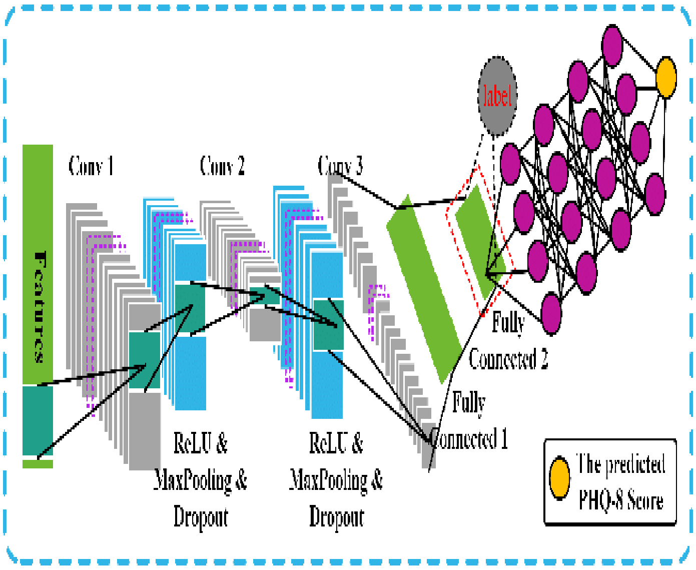
Features
Conv 1
Conv 2
Conv 3
label
ReLU & MaxPooling & Dropout
ReLU & MaxPooling & Dropout
Fully Connected 2
Fully Connected 1
The predicted PHQ-8 Score

Figure.3

RETHINKING INVERSE OPTICS AND FACE REPRESENTATIONS

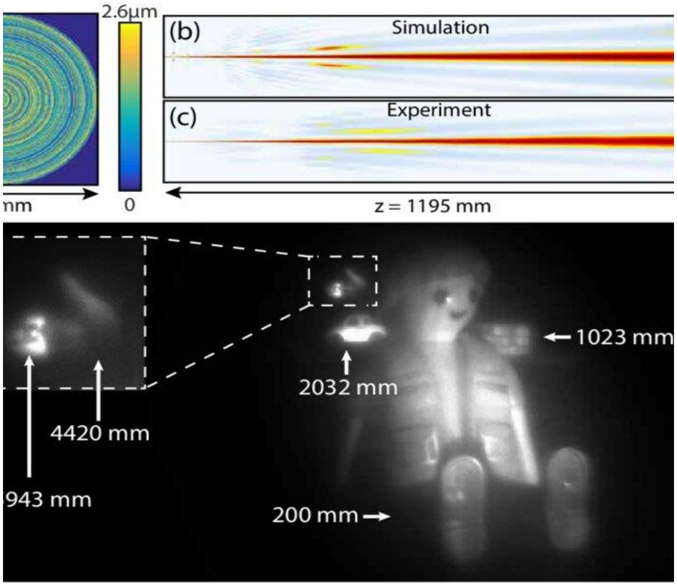

Figure. 4

Deep networks force us to rethink the universe of possible solutions to the problem of inverse optics in vision.[10] These networks operate with a degree of invariance to image and appearance that was unimaginable by researchers less than a decade ago. Invariance refers to the model's ability to consistently identify a face when image conditions (e.g., viewpoint, illumination) and appearance (e.g., glasses, facial hair) vary. The nature of the representation that accomplishes this is not well understood.[13] The inscrutability of DCNN codes is due to the enormous number of computations involved in generating a face representation from an image and the uncontrolled training data.[6] To create a face representation, millions of nonlinear, local convolutions are executed over tens (to hundreds) of layers of units. Researchers exert little or no control over the training data, but instead source face images from the web with the goal of finding as much labeled training data as possible. The number of images per identity and the types of images (e.g., viewpoint, expression, illumination, appearance, quality) are left (mostly) to what is found through web scraping.[2] Nevertheless, DCNNs produce a surprisingly structured and rich face representation that we are beginning to understand.[1-2]

Mining the Face Identity Code in Deep Networks

Figure. 5

The face representation generated by DCNNs for the purpose of identifying a face also retains detailed information about the characteristics of the input image (e.g., viewpoint, illumination) and the person pictured (e.g., gender, age). As shown below, this unified representation can solve multiple face processing tasks in addition to identification.[13-15].

Image characteristics

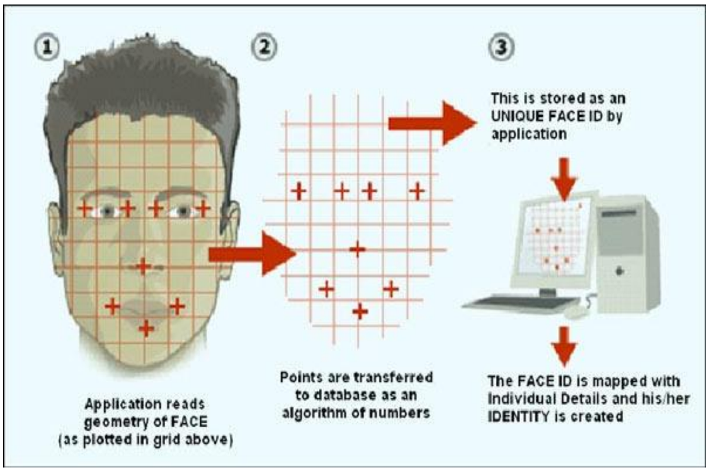
1
2
3
This is stored as an UNIQUE FACE ID by application
Application reads geometry of FACE (as plotted in grid above)
Points are transferred to database as an algorithm of numbers
The FACE ID is mapped with Individual Details and his/her IDENTITY is created

Figure. 6

Face representations generated by deep networks both are and are not invariant to image variation. These codes can identify faces invariantly over image change, but they are not themselves invariant.[3] Instead, face representations of a single identity vary systematically as a function of the characteristics of the input image. The representations generated by DCNNs are, in fact, representations of face images.[12]

Work to dissect face identity codes draws on the metaphor of a face space (Valentine 1991) adapted to representations generated by a DCNN. Visualization and simulation analyses demonstrate that identity codes for face images retain ordered information about the input image (Dhar et al. 2020, Hill et al. 2019, Parde et al. 2017).[4] Viewpoint (yaw and pitch) can be predicted accurately from the identity code, as can media source (still image or video frame) (Parde et al. 2017). Image quality (blur, usability, occlusion) is also available as the identity code norm (vector length).[1] Poor-quality images produce face representations centered in the face space, creating a DCNN garbage dump.[12] This organizational structure was replicated in two DCNNs with different architectures, one developed by Chen et al. (2016) with seven convolutional layers and three fully connected layers and another developed by Sankaranarayanan et al. (2016) with 11 convolutional layers and one fully connected layer. Image quality estimates can also be optimized directly in a DCNN using human ratings (Best-Rowden & Jain 2018).[14]

Facial expressions- Face representations generated by deep networks both are and are not invariant to image variation. These codes can identify faces invariantly over image change, but they are not themselves invariant. Instead, face representations of a single identity vary systematically as a function of the characteristics of the input image. The representations generated by DCNNs are, in fact, representations of face images.[9-12]

Figure.7

Facial expressions are also detectable in face representations produced by identity-trained deep networks.[13] Colón et al. (2021) found that expression classification was well above chance for face representations of images from the Karolinska data set (Lundqvist et al. 1998), which includes seven facial expressions (happy, sad, angry, surprised, fearful, disgusted, neutral) seen from five viewpoints (frontal and 90- and 45-degree left and right profiles). Consistent with human data, happiness was classified most accurately, followed by surprise, disgust, anger, neutral, sadness, and fear.[15] Notably, accuracy did not vary across viewpoint. Visualization of the identities in the emergent face space showed a structured ordering of similarity in which viewpoint dominated over expression.[3]

Recognition stage- Then, the participants performed an old/new recognition test in which they were shown a set of faces and asked to decide whether they had seen the faces earlier; subsequently, they were asked to evaluate the trustworthiness of these faces. The results of the two operations were combined to obtain an unbiased hit rate calculation for late-stage data analysis. [12]Each recognition task trial consisted of a fixation cross (300 ms) followed by a random face presented in the center of the screen without a time limit until a response was provided by pressing a key indicating “old” or “new” (1 or 2 on the keyboard, respectively). After this interface, the same face was presented in the center of the screen once again and remained until the participants pressed a number key to judge its facial trustworthiness (on a scale from 1 to 7), after which the screen turned black (300 ms).[11]

The recognition decisions and trustworthiness evaluations of faces were self-paced; 48 of the 96 faces had been seen previously, and the other 48 were new and served as distractors.[7] All the faces were displayed in a random order. The entire experiment lasted approximately 30 min. The specific experimental process is shown in Figure 1.[14]

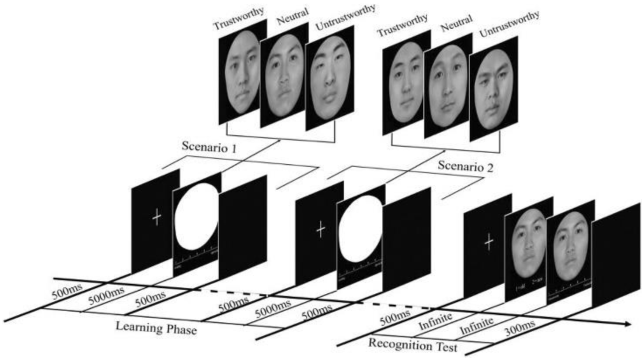

Figure 8. Graphical representation of Experimet 1 procedure.

2. EVALUATION RESULTS

Second, we investigated the response tendency to facial trustworthiness types under the two conditions during the learning stage[7].The results of the repeated measures MANOVA indicated that only the main effect of facial trustworthiness was significant, F(2, 126) ¼ 95.23, p < .001, Z2 ¼ .602, the main effect of the rating scenario was not significant (F < 1), and the interaction effect was not statistically significant, F(2, 126) ¼ 2.77, p > .05. The results of Bonferroni correction suggested that the degree of approaching

trustworthy faces was the highest (M ¼ 3.84, SD ¼ 0.122), significantly higher than the degree of approaching neutral faces (M ¼ 3.40, SD ¼ 0.109) or untrustworthy faces (M ¼ 2.66, SD ¼ 0.111, ps < .001) at a significance level of a ¼ .017. The degree of approaching neutral faces was in the middle, significantly higher than the assessment level of untrustworthy faces (p < .001) at a significance level of a ¼ .017. In general, the two polarities of facial trustworthiness can induce different responses: The participants tended to approach trustworthy faces and avoid untrustworthy faces.[4-9] This finding was relatively consistent with previous studies on facial trustworthiness (Engell et al., 2007; Todorov & Oosterhof, 2011). The scores of a scenario rating during the rating phase were also not associated with later recognition, survival, r(64) ¼ .015, ns; control, r(64) ¼ .089, ns. Furthermore, the three assessment levels of trustworthiness in the various scenarios also showed no significant correlation with the facial recognition scores (ps > .05). [13]This result is consistent with the conclusion that no significant relationship exists between rating scores and recognition scores found in previous survival processing studies (Nairne et al., 2017; Nairne & Pandeirada, 2016; Ro¨er, Bell, & Buchner, 2013; Savine et al., 2011).[11-15]

This result is consistent with the conclusion that no significant relationship exists between rating scores and recognition scores found in previous survival processing studies (Nairne et al., 2017; Nairne & Pandeirada, 2016; Ro¨er, Bell, & Buchner, 2013; Savine et al., 2011).[13]

Facial Recognition Results

In light of the “old” decision in the new/old decision task during the recognition stage, we further divided the facial trustworthiness rating results into three categories based on the evaluation results: trustworthy (more than 4), neutral (4), and untrustworthy (less than 4). The original participants’ responses across decision types are presented in Table 1. In addition, we computed the unbiased hit rate according to Wagner’s (1993) formula for the facial recognition tasks in the two scenarios, as indicated in Table 2.[4] We examined the difference in Hu’s unbiased hit rate across the various levels of facial trustworthiness under the two rating scenario. To reduce the chance of Type I errors, the degrees of freedom for all repeated measures MANOVAs were adjusted using the Greenhouse and Geisser method (Greenhouse & Geisser, 1959). The results showed that the main effect of the rating scenario condition was significant, e ¼ 1.000, F(1, 63) ¼ 12.51, p < .01, Z2 ¼ .166. The accurate recognition ratio of survival scenario (Msurvival ¼ 0.08, SD ¼ 0.006) was significantly higher than control scenario (Mmoving ¼ 0.06, SD ¼ 0.005, p< .001, Z2 ¼ .173. Except for these results, there was no significant interaction, F < 1.

3. RESULTS AND DISCUSSION

Evaluation Results

We examined the participants' tendencies to respond during the learning stage. The repeated measures MANOVA results showed that only the main effect of facial trustworthiness was significant, F(2, 122) ¼ 58.84, p < .001, Z2 ¼ .491, and the results of a Bonferroni correction indicated that the evaluation of faces as trustworthy was the highest (M ¼ 3.63, SD ¼ 0.110), significantly higher than the evaluation of faces as neutral (M ¼ 3.26, SD ¼ 0.102) or untrustworthy (M ¼ 2.59, SD ¼ 0.120; ps < 0.001) at a significance level of a ¼ .017. The evaluation of neutral faces was significantly higher than that of untrustworthy faces (p < .001) at a significance level of a ¼ .017. This finding shows that the participants tended to approach trustworthy faces and avoid untrustworthy faces. The main effect of the rating scenario and the interaction effect were not significant (Fs < 1). [8]The score of the scenario rating during the rating phase was also not associated with later recognition, survival, r(62) ¼ .008, ns; control, r(62) ¼ .037, ns Furthermore, the three assessment levels of trustworthiness in the various scenarios also showed no significant correlation with facial recognition scores (ps > .05). The evaluation results of Experiment 2 are consistent with those of Experiment 1, indicating that we should not attribute the difference in the recognition test to the deep processing of information and that other explanations must be considered.[6]

Facial Recognition Results

In trustworthiness under two rating scenario and the raw response rates of the participants in the recognition test, as shown in Table 3. The results showed that the main effect of the rating scenario was significant, e ¼ 1.000, F(1, 61) ¼ 22.12, p < .001, Z2 ¼ .267, indicating a SPA (Msurvival ¼ 0.09, SD ¼ 0.008.[15]

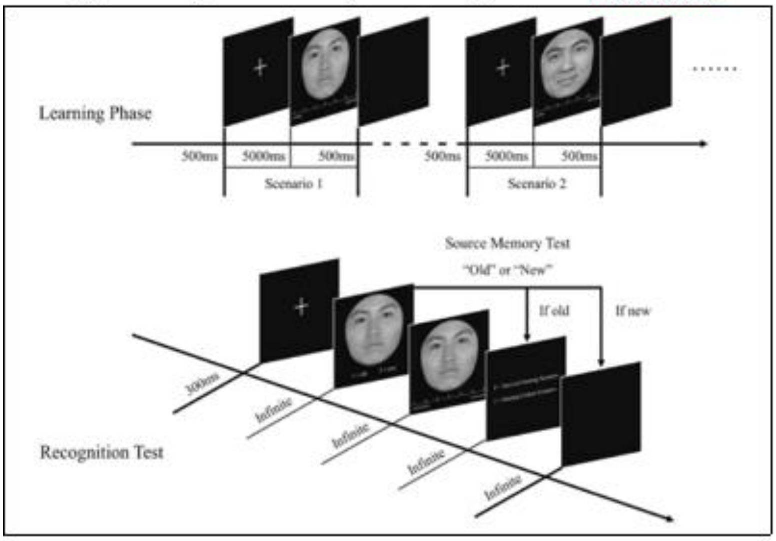

Figure. 9

4. CONCLUSION

1. Face representations generated by DCNN networks trained for identification retain information about the face (e.g., identity, demographics, attributes, traits, expression) and the image (e.g., viewpoint).
2. Deep learning face networks generatesurprisingly structured face representation from unstructured training with in-the-wild face images.
3. Individual output units from deep networks are unlikely to signal the presence of interpretable features.
4. Fundamental structural aspects of high-level visual codes for faces in deep networks replicate over a wide variety of network architectures.
5. Diverse learning mechanisms in DCNNs, applied simultaneously or in sequence, can be used to model human face perception across the lifespan

5. REFERENCES

[1] Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. Review of General Psychology, 5, 323–370. doi:10.1037111089-2680.5.4.323

[2] Bell, R., Buchner, A., Erdfelder, E., Giang, T., Schain, C., & Riether, N. (2012). How specific is source memory for faces of cheaters? Evidence for categorical emotional tagging. Journal of Experimental Psychology: Learning, Memory, and Cognition, 38, 457–472. doi:10.1037/a0026017

[3] Bell, R., Buchner, A., Kroneisen, M., & Giang, T. (2012). On the flexibility of social source memory: A test of the emotional incongruity hypothesis. Journal of Experimental Psychology Learning Memory & Cognition, 38, 1512–1529. doi:10.1037/a0028219

[4] Buchner, A., Rothermund, K., Wentura, D., & Mehl, B. (2004). Valence of distractor words increases the effects of irrelevant speech on serial recall. Memory & Cognition, 32, 722–731. doi:

[5] Buss, D. (2014). Evolutionary psychology. Boston, MA: Pearson.

[6] Chaby, L., Hupont, I., Avril, M., Luhernedu, B. V., & Chetouani, M. (2017). Gaze behavior consistency among older and younger adults when looking at emotional faces. Frontiers in Psychology, 8, 548. doi:10.3389/fpsyg.2017.00548

[7] Coren, S., & Russell, J. A. (1992). The relative dominance of different facial expressions of emotion under conditions of perceptual ambiguity. Cognition & Emotion, 6, 339–356. doi:10.1080/ 02699939208409690

[8] Engell, A., Haxby, J., & Todorov, A. (2007). Implicit trustworthiness decisions: Automatic coding of face properties in the human amygdala. Journal of Cognitive Neuroscience, 19, 1508–1519. doi:10. 1162/jocn.2007.19.9.1508

[9] Felisberti, F. M., & Pavey, L. (2010). Contextual modulation of biases in face recognition. PLoS One, 5, e12939. doi:10.1371/journal. pone.0012939

[10] Fernandes, N. L., Pandeirada, J. N. S., Soares, S. C., & Nairne, J. S. (2017). Adaptive memory: The mnemonic value of contamination. Evolution and Human Behavior, 38, 451–459. doi:10.1016/j.evolhumbehav.2017.04.003

[11] Gelin, MBonin,P,M´eot,A.,&Bugaiska,A.(2017).Doanimacyeffects persist in memory for context? Quarterly Journal of Experimental Psychology, 71, 965–974. doi:10.1080/17470218.2017.1307866 Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. Psychometrika, 24, 95–112. doi:10.1007/ BF02289823

[12] Hou, C. N. (2017). Face: The evolutionary code of intergroup trust. Beijing, China: Science Press.

[13] Holm, S. (1979). A simple sequentially rejective multiple test procedure. the Scandinavian Journal of Statistics, 6, 65–70. doi:10. 1007/BF00139637

[14] Kazanas, S. A., & Altarriba, J. (2015). The survival advantage: Underlying mechanisms and extant limitations. Evolutionary Psychology, 13, 360–396. doi:10.1177/147470491501300204

[15] Kensinger, E. A. (2007). Negative emotion enhances memory accuracy: Behavioral and neuroimaging evidence. Current Directions in Psychological Science, 16, 213–218. doi:10.1111/j.1467- 8721. 2007.00506.x