2023 IEEE International Conference on Robotics and Automation (ICRA 2023)
May 29 - June 2, 2023. London, UK

Automatic Generation of Robot Facial Expressions with Preferences

Bing Tang, Rongyun Cao, Rongya Chen, Xiaoping Chen, Bei Hua, Feng Wu*

***Abstract*— The capability of humanoid robots to generate facial expressions is crucial for enhancing interactivity and emotional resonance in human-robot interaction. However, humanoid robots vary in mechanics, manufacturing, and appearance. The lack of consistent processing techniques and the complexity of generating facial expressions pose significant challenges in the field. To acquire solutions with high confidence, it is necessary to enable robots to explore the solution space automatically based on performance feedback. To this end, we designed a physical robot with a human-like appearance and developed a general framework for automatic expression generation using the MAP-Elites algorithm. The main advantage of our framework is that it does not only generate facial expressions automatically but can also be customized according to user preferences. The experimental results demonstrate that our framework can efficiently generate realistic facial expressions without hard coding or prior knowledge of the robot kinematics. Moreover, it can guide the solution-generation process in accordance with user preferences, which is desirable in many real-world applications.**

Fig. 1. **Humanoid Robot Head.** The basic idea is to use the MAP-Elites algorithm [58] to search the high-dimensional solution space for candidates that match performance criteria in the low-dimensional feature space. The dimension of variation of interest depends on the facial expression-generating mechanism. The algorithm reveals each region's fitness potential and the trade-off between performance and preferences.

I. INTRODUCTION

Daily nonverbal communication mainly relies on facial expressions [1], [2]. Comparing with body language and voice tone, facial expressions are more effective in conveying attitudes and feelings, accurately interpreting and describing the emotions and intentions [3]–[5]. In practice, using a human face to portray humanoid robots makes them more appealing because it allows the interlocutor to learn more about their characteristics or behaviors [7], [8]. Mimicking facial expressions implies that individuals can visually encode and transfer facial expressions to facial muscle movements. However, facial mimicry [13]–[19] is merely the beginning of adaptive facial reactions. Generally, the ability of humanoid robots to generate facial expressions automatically and provide emotional information can enhance interactivity and emotional resonance [9]–[11], potentially leading to positive, long-term human-robot partnerships [9], [11], [12].

Humanoid robots usually have diverse mechanical components and appearances, resulting in different abilities to generate facial expressions. To date, few generic learning frameworks handle nonlinear mapping between motor movements and facial expressions. Among them, predetermining a set of facial expressions is the most straightforward and efficient approach [16], [18]–[27]. Other approaches try to generalize by searching for the closest match [28], [29] or using the fitness function [30], [31]. Although these methods can generate realistic facial expressions, they tend to be excessively artificial, restricting their applicability and significance in real-world human-robot interactions.

Against this background, we developed a humanoid robot head with an anthropomorphic appearance, as shown in Fig. 1. Based on this, we proposed a general learning framework that utilizes MAP-Elites [58] for automatically generating specified facial expressions. Our framework suits humanoid robots with various external appearances and internal mechanisms. It consists of an expression recognizer, a MAP-Elites module, servo drivers, and an *intermediate information processing* (IIP) module. The main functions of the IIP module are as follows. It first receives candidate solutions from the MAP-Elites module, based on the specified expression category and user preferences. Then, it converts these into actual motor commands, which the robot can recognize and execute on the physical robot. After that, it receives the facial attributes from the expression recognizer and gives feedback to the MAP-Elites module. The experiments show that our method outperforms both predefined facial expression methods and facial mimics. Most importantly, the humanoid robot effectively explores potential solutions that can generate specified facial expressions and guides the solution-generation process according to the user preferences.

Our main contributions are twofold. Firstly, we developed an anthropomorphic robot head with sophisticated internal mechanisms. Secondly, we proposed a general learning framework based on the MAP-Elites algorithm for automatic facial expression generation. In more detail, our approach

This work was supported in part by the Major Research Plan of the National Natural Science Foundation of China (92048301), Anhui Provincial Major Research and Development Plan (202004H07020008).

Bing Tang, Rongyun Cao, Xiaoping Chen, Bei Hua, and Feng Wu are with School of Computer Science and Technology, University of Science and Technology of China. Rongya Chen is with Institute of Advanced Technology, University of Science and Technology of China. (Email: {bing_tang, ryc, cry}@mail.ustc.edu.cn, {xpchen, bhua, wufeng02}@ustc.edu.cn).
*Feng Wu is the corresponding author.

does not require human intervention or prior knowledge of kinematics since it only relies on the movement range of each motor of a humanoid robot and its constraint relationships. Hence, it can be directly applied to other humanoid robot designs. To the best of our knowledge, this is the first work enabling humanoid robots to generate facial expressions with user's preferences automatically.

The rest of the paper is structured as follows. Section II briefly reviews the related work. Sections III, IV, and V describe the hardware design, the algorithm framework, and the method evaluation, respectively. Finally, Section VI concludes the paper and discusses potential future research.

II. RELATED WORKS

A. Dynamic Facial Animation

Customized dynamic avatars match the user's geometric shape, physical appearance, and dynamic facial expressions in various situations. Data-driven approaches are frequently employed to create dynamic avatars since they deliver realistic facial expressions at a low computational cost. Tracking RGB-D videos allows real-time linear modeling of blendshape models' dynamic geometry [35], [36]. High-end production utilizes special hardware configurations to create photorealistic, dynamic avatars with fine-grained skin details [37]. Jimenez *et al.* [38] calculated the appearance of the dynamic skin by combining the hemoglobin distributions acquired during various facial expressions. Saragih *et al.* [39] proposed a real-time facial puppet system in which a non-rigid tracking system captures the user's facial expressions.

Cao *et al.* [40] proposed a calibration-free technique for real-time facial tracking and animation by a single camera based on alternating regression steps to infer accurate 2D facial landmarks and 3D facial shapes for 2D video frames. Cao *et al.* [41] presented a real-time high-fidelity facial capture approach that improves the global real-time facial tracker, which produces a low-resolution facial mesh, by augmenting details such as expression wrinkles with local regressors. Yan *et al.* [42] first extracted the AUs (action units) values of the 2D video frames, which were then utilized as blendshape coefficients to animate a digital avatar. X2Face [43] and Face2Face [44] proposed to animate the target video's facial expressions by the source actor and re-render the output video with photorealism.

Prior research primarily focused on realistic video rendering instead of an application to physical robots. Complex internal architecture, constrained degrees of freedom, flexible skin, and actuation mechanisms make applying dynamic face animation to a physical robot much more challenging [30].

B. Physical Humanoid Robot

The cognitive theory [3], [32]–[34] and dimensional theory [45], [46] are the two primary, distinct perspectives on modeling emotions. The cognitive theory categorizes emotions into six basic expressions, while the dimensional theory describes emotions as multiple points in a multidimensional space [55], [56]. These viewpoints serve as theoretical guidelines for humanoid robots' structural design and facial expression generation.

Researchers have conducted numerous studies using cognitive theory. Loza D *et al.* [47] presented a realistic mechanical head that considers a number of micro-movements or AUs. Controlling and combining the matching micro-expressions is a control system that determines the force and velocity of each AU. Implementing all AUs would cause a very complicated, difficult-to-parameterize, and impractical mechanical head, which already exists in the reduced version of particular procedures. Consequently, [19], [20], [22], [25], [48] proposed systematic frameworks that identify human facial expressions or emotional categories of sentences and use the faces of humanoid robots to deliver predetermined facial expressions. Similarly, [22], [26], [27], [49]–[51] developed robot heads with varying appearances and executed a predetermined set of facial expressions or searched databases for the closest match to a particular category of expressions. Berns K *et al.* [29] used a behavior-based robot head ROMAN control system to generate fundamental facial expressions such as happiness, sadness, and surprise.

Lee *et al.* [52]–[54] presented a linear affect-expression spatial model that efficiently controls the expressions of mascot-type robots, allowing for continuous changes and diverse facial characteristics. Moreover, they provided a modeling approach for a three-dimensional linear emotional space based on the basic observation matrix of expressions (BOME), which generates successive expressions from expression data. A Markovian Emotion Model (MEM) was also presented for human-robot interaction, which uses the distance between emotions determined by self-organizing map (SOM) classification to define the initial state transition probability of the MEM [23], [48]. Kanoh M. *et al.* [24] sought to extract expression characteristics of the robot Ifbot and map them into emotional space. In addition, they proposed a method for transforming facial expressions smoothly by utilizing emotional space.

All the approaches above primarily rely on predetermined facial expressions or considerable manual labor, and have limited generalizability. Hyung *et al.* [30] presented a system to automatically generate facial expressions, which is most related to our work. However, their generation process is less efficient, and the outputs are uncontrollable. Therefore, it cannot generate facial expressions based on user preference.

III. HARDWARE DESIGN

The internal movement structure is designed with reference to FACS [3] and the characteristics of human facial muscle movements. Our hardware design is predicated on an anthropomorphic image so that people can recognize the robot's facial expressions more intuitively and consistently. Facial actuation points are driven by servo motors, with one FEETECH SM60L for mouth actuation, four KST X08H Plus for the eyeballs, and additional twelve ones for the remaining actuation points. The robot's internal connections and skull are made by 3D printing to increase availability and reduce cost. The facial expressions of the robot are captured

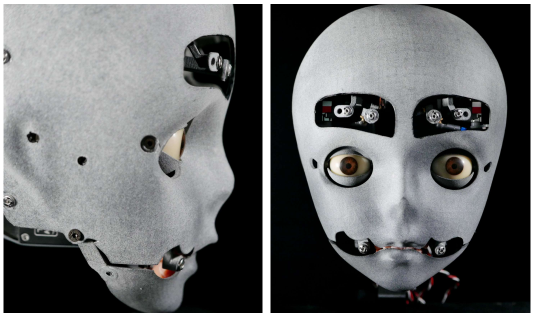

Fig. 2. **Hardware Design.** The flexible skin is attached to the robot's internal connections with snap fasteners. An RGB camera is deployed to capture pictures of the face as the robot executes the generated motor commands. Given the pictures, the current facial performance is evaluated.

TABLE I

CORRESPONDENCES BETWEEN ORIGINAL AND REDEFINED AUS

AU No.	Original Descriptions	New Descriptions
1	Inner Brow Raiser	AU1, Brow Raiser
2	Outer Brow Raiser	
4	Brow Lowerer	AU2, Brow Lowerer
5	Upper Lid Raiser	AU3, Upper Lid Raiser
6	Cheek Raiser	AU4, Cheek Raiser
12	Lip Corner Puller	AU5, Lip Corner Puller
27	Mouth Stretch/Jaw Drop	AU6, Mouth Open

using an ordinary camera. A summary of our hardware design is depicted in Fig. 2.

A. Representation of Facial Expressions

The original action units (AUs) described in FACS show the different movements of facial muscles. The collection of certain AUs provides information about which expression being displayed. For example, "Happiness" is calculated from the combination of AU6 (cheek raiser) and AU12 (lip corner puller). However, it is not necessary for all AUs to occur simultaneously in order to generate a certain facial expression. For instance, when AU1, AU2, and AU5 emerge concurrently, faces are recognized as expression "Surprise" with high likelihood. In this work, we term them *strong associated AUs*, and the rest are *weak correlated AUs*.

The amplitude of movement of each servo motor is empirically utilized to calculate the movement of each AU necessary for each facial expression. To align with the requirements of our algorithm, we have re-described and re-interpreted the description of AUs in FACS due to the specific hardware structure and limitations of servo motions. As shown in Table I, the new descriptions are the AUs that our robot implements.

B. Face Movement Module

We use Smooth-On Ecoflex-0030 to manufacture flexible facial skin attached to the robot's skull with silicone rubber adhesive (Smooth-On Sil-Poxy). The facial actuation points are connected to it with snap fasteners. The silicone is 2mm thick to ensure that all non-linear actuation point movements are reflected in the flexible skin. To simplify angular transformation and control, we standardize the movement range of servo motors to [0, 1]. All the servo motors move in different directions and intensities across a multidimensional space.

Algorithm 1 MAP-Elites [58]

$\mathcal{P} \leftarrow \emptyset$, $\mathcal{X} \leftarrow \emptyset$
for $iter = 1 \rightarrow I$ **do**
  if $iter < \mathcal{G}$ **then**
    $\mathbf{x}' \leftarrow random\_solution()$
  else
    $\mathbf{x} \leftarrow random\_selection(\mathcal{X})$
    $\mathbf{x}' \leftarrow random\_variation(\mathbf{x})$
  end if
  $\mathbf{b}' \leftarrow feature\_descriptor(\mathbf{x}')$
  $p' \leftarrow performance(\mathbf{x}')$
  if $\mathcal{P}(\mathbf{b}') = \emptyset \vee \mathcal{P}(\mathbf{b}') > p'$ **then**
    $\mathcal{P}(\mathbf{b}') \leftarrow p'$, $\mathcal{X}(\mathbf{b}') \leftarrow \mathbf{x}'$
  end if
end for
return $\mathcal{P}$ and $\mathcal{X}$

IV. MAIN FRAMEWORK

We proposed a general learning-based framework for automatically generating facial expressions based on user preferences and expression categories. Fig. 3 depicts an overview of our learning framework. Without prior knowledge of kinematics and manual pre-programming of various facial expressions, we expect the learning framework to generate facial expressions that match human perception.

To automate the generation process, utilizing the current mature and reliable expression recognizer to deliver high-performance evaluations is crucial, ensuring more accurate outcomes. According to our knowledge, no prior study has explicitly addressed the use of a mature and stable interface for effectively identifying facial expressions by utilizing affect space [45], [46]. Therefore, we use the generic expression recognition interface, which is provided by the official face++ website [66], to recognize facial expressions more precisely and acquire the corresponding facial attributes.

Considering the inefficiency of the hard-coding approach and the strong correlation between the robot's facial expressions and mechanical structure, generating facial expressions requires a good exploration of the solution space while implicitly preserving diversity. In this paper, we proposed improving the MAP-Elites algorithm to address these issues, which is detailed next.

A. MAP-Elites Improvement

Compared to conventional genetic algorithms [60]–[62], MAP-Elites [58], [59] aims to generate diverse high-performing solutions, which may be more beneficial than a single high-performing solution. MAP-Elites has found widespread application in robotics, such as robot trajectory optimization [64], maze navigation [63], and gait optimization for legged robots [58].

We use MAP-Elites because it can simultaneously find the optimal solution and preserve diversity in multiple dimensions, which is well-suited for the characteristics of robotic

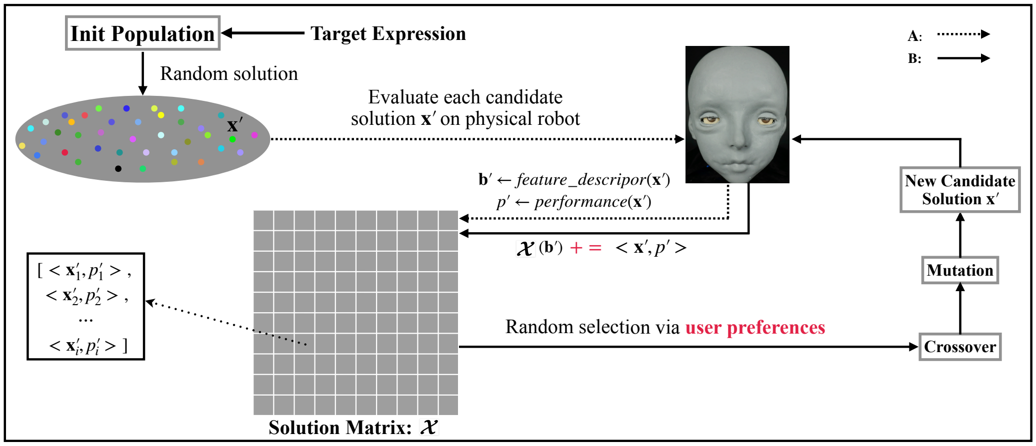

Fig. 3. **Framework Overview.** Our framework consists of two phases when generating the target expression. **Phase A**: MAP-Elites [58] creates candidate solutions randomly before evaluating their performances and features. When performance meets requirements, the solutions are placed in the feature space cells to which they belong. If multiple solutions map to the same cell, they are all preserved in that cell. **Phase B**: We randomly select several non-empty cells from the whole domain of the solution matrix or a specific region determined by user preferences. After crossover and mutation, the obtained offspring is evaluated after execution on the physical robot. Once its performance meets the criteria, the offspring will be placed into the solution matrix.

facial expressions. Algorithm 1 provides a pseudocode of the default MAP-Elites algorithm [58], [59], which can be easily conceptualized and implemented. Specifically, $\mathbf{x}$ and $\mathbf{x}'$ are candidate solutions which are n-dimentional vectors in the solution space consisting of all possible values of $\boldsymbol{x}$, where $\boldsymbol{x}$ is a description of a candidate solution; the candidate solution $\mathbf{x}'$ is mapped to the location $\mathbf{b}' = (\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_M)$ in the discrete feature space through a user-defined feature descriptor, where M is the dimension of feature space. p' denotes the performance obtained after the performance measure of the candidate solution $\mathbf{x}'$. $\boldsymbol{\mathcal{P}}$ is a performance matrix that gives the best performance $\boldsymbol{\mathcal{P}}(\mathbf{b}')$ for the corresponding location $\mathbf{b}'$ in $\boldsymbol{\mathcal{P}}$. Similarly, $\boldsymbol{\mathcal{X}}$ is a solution matrix that stores the best candidate solution $\boldsymbol{\mathcal{X}}(\mathbf{b}')$ for the matching location $\mathbf{b}'$ of $\boldsymbol{\mathcal{X}}$. After evaluating and featuring the candidate solution $\mathbf{x}'$ with better performance p', the candidate solution $\mathbf{x}'$ and performance p' are either placed at $\boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{X}}$ at location $\mathbf{b}'$ or replaced the previous element.

Since our goal is to increase the diversity of the generated facial expressions, the corresponding solutions can be placed in the solution matrix if the confidence level of the expression is the highest among the current facial attributes. Consequently, our alternative approach is to no longer require a separate performance matrix $\boldsymbol{\mathcal{P}}$, but rather to retain the solution matrix $\boldsymbol{\mathcal{X}}$. As long as the performance p' of the current candidate solution $\mathbf{x}'$ meets the criteria after evaluation, the solution and corresponding performance $\langle \mathbf{x}', p' \rangle$ will be preserved in the solution matrix $\boldsymbol{\mathcal{X}}$, i.e., $\boldsymbol{\mathcal{X}} += \langle \mathbf{x}', p' \rangle$. Here, candidate solution $\mathbf{x}' = (\boldsymbol{x}^1, \boldsymbol{x}^2, \ldots, \boldsymbol{x}^n)^T$ is an n-dimensional vector, each dimension of which is converted to actual motor angle $\boldsymbol{v}^j$ based on the movement range of the corresponding motor, resulting in a brand-new motor command $\boldsymbol{\mathcal{V}} = (\boldsymbol{v}^1, \boldsymbol{v}^2, \ldots, \boldsymbol{v}^N)^T$. p' represents the confidence level of the current target expression after the robot executes the motor command $\boldsymbol{\mathcal{V}}$.

Another novelty of this study is the strategy adopted by the $random\_selection()$ function in Algorithm 1. Instead of conducting a global random search from the solution matrix, which would result in uncontrollable offspring generation, we propose using the average offset of the motor angles corresponding to weak correlated AUs to represent user preferences. The offspring generation process would then select solutions from the user-specified region as parents.

B. Feature Descriptor

In Section III-A, we divide the collection of AUs corresponding to each basic expression into two parts: strong correlated AUs and weak correlated AUs. The former indicates a high likelihood of recognition as a specific expression if the combination occurs, while the latter increases its likelihood. Thus, the strong correlated AUs ensure the specificity of the facial categories generated, whereas the weak correlated AUs increase the diversity and originality of the generated expressions, which are our variations of interest.

For each candidate solution $\mathbf{x}'$, a feature descriptor determines the location of the solution in each cell of the feature space. In other words, $\mathbf{b}'$ is a M-dimensional vector describing the feature of $\mathbf{x}'$. In this work, $\mathbf{b}' = (\boldsymbol{b}_1, \boldsymbol{b}_2)$ has two dimensions, and $\boldsymbol{b}_1, \boldsymbol{b}_2$ represents the average angle offsets of strong correlated AUs and weak correlated AUs that are associated with the target expression, respectively.

V. EXPERIMENTS

Typically, evaluating an expression-generating system is a qualitative process. However, the solution matrix produced by the MAP-Elites algorithm allows us to quantify the system's efficiency in various ways. In this study, we demonstrate the effectiveness of our system using qualitative snapshots of the robot's face. By examining the number of solutions stored in individual cells of the solution matrix, we can visually assess the advantages of generating solutions with high performance, diversity, and originality. Importantly, our system can provide solutions that closely

TABLE II
NUMBER OF PER TARGET EXPRESSION

Expression Category	Number (Ours)	Number [30]
Sadness	711	579
Surprise	154	65
Happiness	251	154
Fear	597	416

align with user preferences by leveraging the feature space without sacrificing the quantity of generated solutions.

A. Experimental Setup

1) Target Expressions: We conducted a preliminary experiment using the expression recognizer to assess the robot's ability to generate facial expressions. The difficulty in generating "Disgust" and "Anger" is because only limited or even no degrees of freedom exist in the nose and lip regions, as shown by the pre-experimental results. Thus, they were eliminated from the target expressions. In our experiments, the target expressions for the robot are "Sadness", "Surprise", "Happiness", and "Fear".

2) Implementation Details: In MAP-Elites, we consider two features: **feature 1** represents the average offset of motor angles corresponding to the strong correlated AUs, and **feature 2** represents the average offset of motor angles corresponding to the weak correlated AUs. The performance criteria are the confidence level of the target expression, which ranges from 0 to 100. The size of solution matrix is 10×10. The size of initial population is 1000 and the number of iterations is 5000. The process takes about an hour. The user preferences we consider are the high and low offsets of **feature 2**.

To evaluate the robot's capability to generate various target expressions and verify user preferences' impact on generation results, we adopt identical experimental setups for all target expressions.

B. Experimental Results

1) Generating Ability Evaluation: Table II shows the numbers of generated solutions for each target expression under identical experimental setups that vary quantitatively. Figure 4 more vividly displays the distribution of solutions in the solution matrix, where the value ranges of **feature 1** and **feature 2** differ across various solution matrices. Despite automatically generating a certain number of solutions for each target expression, the results are entirely non-pointed. The primary objective of this paper is to verify the robot's ability to generate specified classes of facial expressions. To this end, we compared our method to a previous approach [30] that implemented expression generation based on a genetic algorithm. According to the quantitative comparison results in Table II, our approach has a remarkable advantage in generation efficiency for the same number of executions on the physical robot. In Figure 4 and Figure 6, **Min**, **Max**, **Ave** denote the minimum, maximum, and average confidence levels of all solutions in the solution matrix, respectively, illustrating that our method generates a diversity of solutions.

2) Robot Face Visualizations: To demonstrate the effect of facial expression generation, for each target expression in turn, we randomly select a certain number of solutions with confidence levels above 80 from the corresponding solution matrix in Figure 4; these solutions are converted into the corresponding motor commands and executed on the physical robot; and we intercept the robot's facial expression following the execution of each command for display. The visualization results are shown in Figure 5.

3) Expression Generation with Preferences: Certainly, the MAP-Elites algorithm can generate many reliable solutions in a short period, but its generation process is utterly random and goalless. Consequently, we consider the average offset of the motor angles corresponding to the weak correlated AUs to represent the user preference, i.e., the offset is separated into high and low offsets, and the solution matrix is divided into two identical halves accordingly. The process of offspring generation would then select solutions in the user-specified region as parents rather than conducting a global search of the solution matrix.

To validate the effects of user preferences on solutions, we choose the high and low offset regions corresponding to **feature 2** as user preferences, respectively, for each target expression. Figure 6 depicts the solution matrices generated by the MAP-Elites algorithm when the high and low offset regions are chosen as user preferences, respectively. By comparing with the solution matrices in Figure 4, solutions in **(A)** tend to shift towards the high offset region in **feature 2**, whereas those in **(B)** tend to shift towards the low offset region. This demonstrates that user preferences are instructional to a degree for generating solutions.

VI. CONCLUSIONS AND FUTURE WORK

We developed a humanoid robot head with an anthropomorphic appearance, multiple degrees of freedom, and flexible skin. Moreover, we presented a generic learning framework based on the MAP-Elites algorithm for automatically generating facial expressions that can be customized according to user preferences. Our experiments demonstrate that the framework can generate a large number of human-recognizable facial expressions without hard programming or prior knowledge of kinematics. The solution matrices offer a visualization of the diversity of the generated solutions. In addition, we have effectively guided the solution-generating process using user-specific preferences.

Although our proposed learning framework has demonstrated its powerful ability to generate facial expressions, robots with higher degrees of freedom may be able to display even richer expressions. Furthermore, while the expression generation process does not require prior knowledge of robot kinematics, it may be challenging to generate facial expressions in real-time. Those mentioned above are all possible improvements, and we leave them for future work.

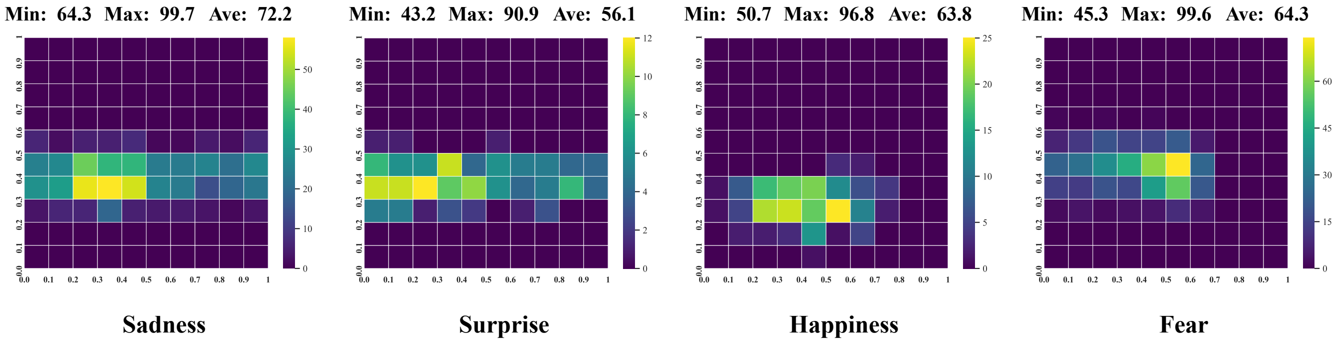

Fig. 4. **Solution Matrices of Target Expressions. feature 1** is the vertical coordinates of solution matrices, whereas **feature 2** is horizontal. Under the constraints of the target expression and AUs, the solutions are not uniformly distributed in the solution matrices.

Fig. 5. **Target Expression Visualizations.** For each target expression individually, we randomly select a certain number of solutions with confidence levels greater than 80% from the corresponding solution matrix, then convert them into motor commands, and finally display the facial expressions after executing on our physical robot head.

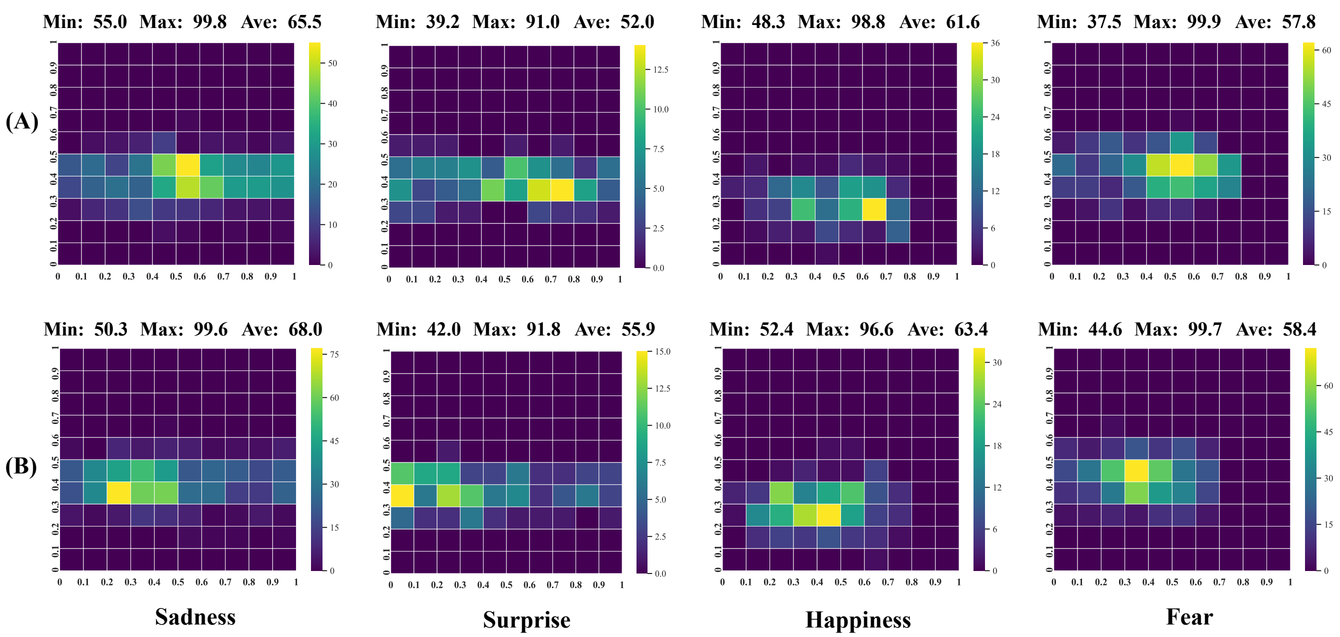

Fig. 6. **Expression Generation with Preference.** We split the **feature 2** into two equal portions: the high and low offset regions. When the high and low offsets are considered as user preferences, **(A)** and **(B)** represent the changes in the distribution of generated solutions in the solution matrix.

References

[1] A. Mehrabian, *Silent messages: Implicit communication of emotions and attitudes*, 2nd ed. Belmont, CA: Wadsworth Publishing, 1981.

[2] Phutela, Deepika, "The importance of non-verbal communication," *IUP Journal of Soft Skills* 9.4 (2015): 43.

[3] Ekman, Paul, and Wallace V. Friesen, "Unmasking the face: A guide to recognizing emotions from facial clues," *Malor Books*, 2003.

[4] C. Darwin, *The expression of the emotions in man and animals*, 200th ed. London, England: HarperPerennial, 1998.

[5] Plutchik R. "Emotions: A general psychoevolutionary theory," *Approaches to emotion*, 1984, 1984(197-219): 2-4.

[6] U. Dimberg and M. Thunberg, "Rapid facial reactions to emotional facial expressions," *Scand. J. Psychol.*, vol. 39, no. 1, pp. 39–45, 1998.

[7] T. Koda and P. Maes, "Agents with faces: The effect of personification," in *Proceedings 5th IEEE International Workshop on Robot and Human Communication. RO-MAN'96 TSUKUBA*, 2002.

[8] Takeuchi A, Naito T. "Situated facial displays: towards social interaction," in *Proceedings of the SIGCHI conference on Human factors in computing systems*. 1995: 450-455.

[9] F. Cavallo, F. Semeraro, L. Fiorini, G. Magyar, P. Sinčák, and P. Dario, "Emotion modelling for social robotics applications: a review," *J. Bionic Eng.*, vol. 15, no. 2, pp. 185–203, 2018.

[10] S. Saunderson and G. Nejat, "How robots influence humans: A survey of nonverbal communication in social human–robot interaction," *Int. J. Soc. Robot.*, vol. 11, no. 4, pp. 575–608, 2019.

[11] Nadel J, Simon M, Canet P, et al. "Human responses to an expressive robot,". *Proc. of the Sixth International Workshop on Epigenetic Robotics*. Lund University, 2006.

[12] R. Kirby, J. Forlizzi, and R. Simmons, "Affective social robots," *Rob. Auton. Syst.*, vol. 58, no. 3, pp. 322–332, 2010.

[13] B. Chen, Y. Hu, L. Li, S. Cummings, and H. Lipson, "Smile like you mean it: Driving animatronic robotic face with learned models," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 2021.

[14] Magtanong, Emarc, et al. "Inverse kinematics solver for an android face using neural network". In *29th Annual Conf. of the Robotics Society of Japan*, pp. 1Q3–1. 2011.

[15] E. Magtanong, A. Yamaguchi, K. Takemura, J. Takamatsu, and T. Ogasawara, "Inverse kinematics solver for android faces with elastic skin," in *Latest Advances in Robot Kinematics, Dordrecht: Springer Netherlands*, 2012, pp. 181–188.

[16] A. Habib, S. K. Das, I.-C. Bogdan, D. Hanson, and D. O. Popa, "Learning human-like facial expressions for android Phillip K. Dick," in *2014 IEEE International Conference on Automation Science and Engineering (CASE)*, 2014.

[17] Z. Huang, F. Ren, M. Hu, and S. Chen, "Facial expression imitation method for humanoid robot based on smooth-constraint reversed mechanical model (SRMM)," *IEEE Trans. Hum. Mach. Syst.*, vol. 50, no. 6, pp. 538–549, 2020.

[18] S. S. Ge, C. Wang, and C. C. Hang, "Facial expression imitation in human robot interaction," in *RO-MAN 2008 - The 17th IEEE International Symposium on Robot and Human Interactive Communication*, 2008.

[19] F. Hara and H. Kobayashi,"A face robot able to recognize and produce facial expression," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS '96*, 2002.

[20] D. Kim, S. Jung, K. An, H. Lee, and M. Chung, "Development of a facial expression imitation system," in *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2006.

[21] A. Esfandbod, Z. Rokhi, A. Taheri, M. Alemi, and A. Meghdari, "Human-Robot Interaction based on Facial Expression Imitation," in *2019 7th International Conference on Robotics and Mechatronics (ICRoM)*, 2019.

[22] A. Ghorbandaei Pour, A. Taheri, M. Alemi, and A. Meghdari, "Human–robot facial expression reciprocal interaction platform: case studies on children with autism," *Int. J. Soc. Robot.*, vol. 10, no. 2, pp. 179–198, 2018.

[23] Y. Maeda and S. Geshi, "Human-robot interaction using Markovian emotional model based on facial recognition," in *2018 Joint 10th International Conference on Soft Computing and Intelligent Systems (SCIS) and 19th International Symposium on Advanced Intelligent Systems (ISIS)*, 2018.

[24] M. Kanoh, S. Kato, and H. Itoh, "Facial expressions using emotional space in sensitivity communication robot 'Ifbot,"' in *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE Cat. No.04CH37566)*, 2005.

[25] Y. Chen, X. Liu, J. Li, T. Zhang, and A. Cangelosi, "Generation of Head Mirror Behavior and Facial Expression for Humanoid Robots," *Advances in Artificial Intelligence and Machine Learning*, vol. 01, no. 02, pp. 149–164, 2021.

[26] N. Endo et al., "Development of whole-body emotion expression humanoid robot," in *2008 IEEE International Conference on Robotics and Automation*, 2008.

[27] H. Miwa, T. Okuchi, H. Takanobu, and A. Takanishi, "Development of a new human-like head robot WE-4," in *IEEE/RSJ International Conference on Intelligent Robots and System*, 2003.

[28] G. Trovato, T. Kishi, N. Endo, K. Hashimoto, and A. Takanishi, "Development of facial expressions generator for emotion expressive humanoid robot," in *2012 12th IEEE-RAS International Conference on Humanoid Robots (Humanoids 2012)*, 2012.

[29] K. Berns and J. Hirth, "Control of facial expressions of the humanoid robot head ROMAN," in *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2006.

[30] H.-J. Hyung, H. U. Yoon, D. Choi, D.-Y. Lee, and D.-W. Lee, "Optimizing android facial expressions using genetic algorithms," *Appl. Sci. (Basel)*, vol. 9, no. 16, p. 3379, 2019.

[31] H.-J. Hyung, D.-W. Lee, H. U. Yoon, D. Choi, D.-Y. Lee, and M.-H. Hur, "Facial expression generation of an android robot based on probabilistic model," in *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2018.

[32] P. Ekman, "An argument for basic emotions," *Cogn. Emot.*, vol. 6, no. 3–4, pp. 169–200, 1992.

[33] Cohn J F, Ambadar Z, Ekman P. "Observer-based measurement of facial expression with the Facial Action Coding System". *The handbook of emotion elicitation and assessment*, 2007, 1(3): 203-221.

[34] S. Wolkind, "Emotion in the human face: Guidelines for research and an integration of findings. By P. ekman, W. v. friesen and P. ellsworth. Pergamon press. 1971. Pp. 191. Price £3.50," *Br. J. Psychiatry*, vol. 122, no. 566, pp. 108–108, 1973.

[35] S. Bouaziz, Y. Wang, and M. Pauly, "Online modeling for realtime facial animation," *ACM Trans. Graph.*, vol. 32, no. 4, pp. 1–10, 2013.

[36] H. Li, J. Yu, Y. Ye, and C. Bregler,"Realtime facial animation with on-the-fly correctives," *ACM Trans. Graph.*, vol. 32, no. 4, pp. 1–10, 2013.

[37] O. Alexander et al., "Digital ira: Creating a real-time photoreal digital actor," in *ACM SIGGRAPH 2013 Posters*, 2013.

[38] J. Jimenez et al., "A practical appearance model for dynamic facial color," in *ACM SIGGRAPH Asia 2010 papers on - SIGGRAPH ASIA '10*, 2010.

[39] J. M. Saragih, S. Lucey, and J. F. Cohn, "Real-time avatar animation from a single image," in *Face and Gesture 2011*, 2011.

[40] C. Cao, Q. Hou, and K. Zhou, "Displaced dynamic expression regression for real-time facial tracking and animation,". *ACM Trans. Graph.*, vol. 33, no. 4, pp. 1–10, 2014.

[41] C. Cao, D. Bradley, K. Zhou, and T. Beeler, "Real-time high-fidelity facial performance capture," *ACM Trans. Graph.*, vol. 34, no. 4, pp. 1–9, 2015.

[42] Y. Yan, K. Lu, J. Xue, P. Gao, and J. Lyu, "Feafa: A well-annotated dataset for facial expression analysis and 3d facial animation," in *2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, 2019.

[43] O. Wiles, A. S. Koepke, and A. Zisserman, "X2face: A network for controlling face generation using images, audio, and pose codes," in *Computer Vision – ECCV 2018*, Cham: Springer International Publishing, 2018, pp. 690–706.

[44] J. Thies, M. Zollhöfer, M. Stamminger, C. Theobalt, and M. Nießner, "Face2face: Real-time face capture and reenactment of rgb videos," *arXiv [cs.CV]*, 2020.

[45] J. A. Russell, "A circumplex model of affect.," *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980, doi: https://doi.org/10.1037/h0077714.

[46] J. A. Russell,"Reading emotions from and into faces: Resurrecting a dimensional-contextual perspective," in *The Psychology of Facial Expression*, J. A. Russell and J. M. Fernandez-Dols, Eds. Cambridge: Cambridge University Press, 1997, pp. 295–320.

[47] D. Loza, S. Marcos Pablos, E. Zalama Casanova, J. Gómez García-Bermejo, and J. L. González, "Application of the FACS in the Design

and Construction of a Mechatronic Head with Realistic Appearance," *J. Phys. Agent.*, vol. 7, no. 1, pp. 31–38, 2013.

[48] Y. Maeda, T. Sakai, K. Kamei, and E. W. Cooper, "Human-robot interaction based on facial expression recognition using deep learning," in *2020 Joint 11th International Conference on Soft Computing and Intelligent Systems and 21st International Symposium on Advanced Intelligent Systems (SCIS-ISIS)*, 2020.

[49] N. Liu and F. Ren, "Emotion classification using a CNN_LSTM-based model for smooth emotional synchronization of the humanoid robot REN-XIN," *PLoS One*, vol. 14, no. 5, p. e0215216, 2019.

[50] H. S. Ahn, D.-W. Lee, and B. MacDonald,"Development of a human-like narrator robot system in EXPO," in *2013 6th IEEE Conference on Robotics, Automation and Mechatronics (RAM)*, 2013.

[51] J. W. Park, H. S. Lee, and M. J. Chung, "Generation of realistic robot facial expressions for human robot interaction," *J. Intell. Robot. Syst.*, vol. 78, no. 3–4, pp. 443–462, 2015.

[52] H. Lee, J. Park, and M. Chung, "An affect-expression space model of the face in a mascot-type robot," in *2006 6th IEEE-RAS International Conference on Humanoid Robots*, 2006.

[53] H. S. Lee, J. W. Park, and M. J. Chung, "A linear affect–expression space model and control points for mascot-type facial robots," *IEEE Trans. Robot.*, vol. 23, no. 5, pp. 863–873, 2007.

[54] H. S. Lee, J. W. Park, S. H. Jo, and M. J. Chung, g, "A linear dynamic affect-expression model: facial expressions according to perceived emotions in mascot-type facial robots," in *RO-MAN 2007 - The 16th IEEE International Symposium on Robot and Human Interactive Communication*, 2007.

[55] Bartneck, Christoph. "How convincing is Mr. Data's smile: Affective expressions of machines". *User Modeling and User-Adapted Interaction*. 11 (2001): 279-295.

[56] A. Ortony and T. J. Turner, "What's basic about basic emotions?," *Psychol. Rev.*, vol. 97, no. 3, pp. 315–331, 1990.

[57] O. Palinko, F. Rea, G. Sandini, and A. Sciutti, "Robot reading human gaze: Why eye tracking is better than head tracking for human-robot collaboration," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016.

[58] A. Cully, J. Clune, D. Tarapore, and J.-B. Mouret, "Robots that can adapt like animals," *Nature*, vol. 521, no. 7553, pp. 503–507, 2015.

[59] J.-B. Mouret and J. Clune, "Illuminating search spaces by mapping elites," *arXiv [cs.AI]*, 2015.

[60] H. Shahrzad, D. Fink, and R. Miikkulainen, "Enhanced optimization with composite objectives and novelty selection,". *arXiv [cs.NE]*, 2018.

[61] J.-B. Mouret, "Novelty-Based Multiobjectivization," in *New Horizons in Evolutionary Robotics*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 139–154.

[62] A. Cully and Y. Demiris, "Quality and diversity optimization: A unifying modular framework," *IEEE Trans. Evol. Comput.*, vol. 22, no. 2, pp. 245–259, 2018.

[63] V. Vassiliades, K. Chatzilygeroudis, and J.-B. Mouret, "A comparison of illumination algorithms in unbounded spaces," in *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, 2017.

[64] E. Samuelsen and K. Glette, "Multi-objective Analysis of MAP-Elites Performance," *arXiv [cs.NE]*, 2018.

[65] S. Fioravanzo and G. Iacca, "Evaluating MAP-Elites on constrained optimization problems," in *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, 2019.

[66] "Expression Recognizer." *Face++ Facial Emotion Recognition*, http://www.faceplusplus.com.cn/emotion-recognition.