IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 29, 2019, accepted September 26, 2019, date of publication October 1, 2019, date of current version October 16, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2944912

A Robotic Writing Framework–Learning Human Aesthetic Preferences via Human–Machine Interactions

XINGEN GAO[1], CHANGLE ZHOU[1], FEI CHAO[1,2], (Member, IEEE), LONGZHI YANG[3], (Senior Member, IEEE), CHIH-MIN LIN[4], (Fellow, IEEE), AND CHANGJING SHANG[2]

[1]Cognitive Science Department, School of Information Science and Engineering, Xiamen University, Xiamen 361005, China
[2]Institute of Mathematics, Physics and Computer Science, Aberystwyth University, Aberystwyth SY23 3FL, U.K.
[3]Department of Computer and Information Sciences, Northumbria University, Newcastle upon Tyne NE1 8ST, U.K.
[4]Department of Electrical Engineering, Yuan Ze University, Taoyuan City 32003, Taiwan

Corresponding author: Fei Chao (fchao@xmu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61673322, Grant 61673326, and Grant 91746103, in part by the Fundamental Research Funds for the Central Universities under Grant 20720190142, in part by the Natural Science Foundation of Fujian Province of China under Grant 2017J01128 and Grant 2017J01129, and in part by the Sêr Cymru II COFUND Fellowship, U.K.

ABSTRACT Intelligent robots are required to fully understand human intentions and operations in order to support or collaborate with humans to complete complicated tasks, which is typically implemented by employing human-machine interaction techniques. This paper proposes a new robotic learning framework to perform numeral writing tasks by investigating human-machine interactions with human preferences. In particular, the framework implements a trajectory generative module using a generative adversarial network (GAN)-based method and develops a human preference feedback system to enable the robot to learn human preferences. In addition, a convolutional neural network, acting as a discriminative network, classifies numeral images to support the development of the basic numeral writing ability, and another convolutional neural network, acting as a human preference network, learns a human user's aesthetic preference by taking the feedback on two written numerical images during the training process. The experimental results show that the written numerals based on the preferences of ten users were different from those of the training data set and that the writing models with the preferences from different users generate numerals in different styles, as evidenced by the Fréchet inception distance (FID) scores. The FID scores of the proposed framework with a preference network were noticeably greater than those of the framework without a preference network. This phenomenon indicates that the human-machine interactions effectively guided the robotic system to learn different writing styles. These results prove that the proposed approach is able to enable the calligraphy robot to successfully write numerals in accordance with the preferences of a human user.

INDEX TERMS Human-machine interaction, human preference, neural networks, robotic calligraphy, robotic writing trajectory.

I. INTRODUCTION

Intelligent machines, including robots, autonomous vehicles, and assistance systems, have been widely applied in our daily lives to support various activities, such as communication, business, transportation, and healthcare. Effective interactions between humans and machines are essential for high performance of intelligent machines. By effective information exchange and interpretation, intelligent machines are able to understand the human intentions, and thus, human-machine interactions (HMIs) are considered an important research topic to assist humans in completing tasks. A number of HMI methods have been developed to improve the capabilities of intelligent machines. For instance, human gesture and activity recognition is a useful and widely used tool for HMI [1]–[4], in addition to speech

The associate editor coordinating the review of this manuscript and approving it for publication was Luigi Biagiotti.

recognition [5], [6], emotion recognition [7]–[9], electroencephalography (EEG), electromyography (EMG), and electrocardiography (ECG) signal interpretation [10]–[13]. In addition, Alsamhi *et al.* [14] reported the current development of artificial intelligence techniques for the application areas of robotic communications, including artificial neural networks, adaptive neuro-fuzzy interference systems, machine learning, and genetic algorithms. This report inspires us to explore artificial intelligence techniques in robotics. These techniques not only allow information or instructions exchange between humans and machines but also allow machines to deepen their understanding of human intention.

The application of HMI to robotic calligraphy is, however, very limited. Most of the writing robots generate trajectories using matching and fitting methods [15]–[19]. Our previous studies [20], [21] focused on building trajectory generative models for robotic Chinese calligraphy systems by integrating AI techniques, such as convolutional autoencoder (CAE), differential evolution (DE), and generative adversarial networks (GANs). Nevertheless, some research work has indeed applied HMI to transfer human handwriting skills to robots [22]–[25]. However, these HMI-based approaches only allowed robots to directly use instructions (i.e., pen movements) given by a human to write characters or letters, rather than interpreting human preferences via HMI. As a result, the robots can use only one fixed writing style, which might not follow a human user's preferences. Furthermore, to increase the diversity of writing results, human engineers must perform a large number of demonstrations for training. Therefore, it is difficult to create new writing styles for calligraphy robots using this type of HMI method.

Note that deep reinforcement learning has been recently developed to guide robots to learn human preferences [26]. This method transforms the evaluations of robot motions to human-robot interaction processes by employing human users to evaluate each action generated by the robots. The evaluation marks are provided based on the alignment of robot actions with human expectation. Such a method has successfully guided a simulated robot to perform complex actions. However, more investigation is required for such a reinforcement learning method to write desired numbers or letters with diverse writing styles; otherwise, the method requires an unexpectedly long training time to converge.

This paper proposes a learning framework to enable calligraphy robots to learn to write numerals with human preferences by further developing the work of [26] to address the above two challenges. The proposed approach takes both human preferences and robot performance feedback simultaneously in the learning framework. In addition, to improve the learning efficiency, we incorporate our previous method [20], i.e., a GAN-based calligraphy system, into this proposed learning framework; a maximum-likelihood-like method is also designed to train the robot writing framework in order to avoid an overly lengthy training time caused by the conventional training methods in reinforcement learning. Thus, the proposed learning framework allows the robot to develop a high-performance writing skill consistent with human preferences.

The main contributions of this work are twofold: 1) a robot learning framework based on both human-robot interactions and the feedback of robotic writing results; and 2) an efficient GAN training approach for trajectory generative module optimization. The main contributions of this work and our previous two works are different. The goal of the previous work, i.e., the DE- and GAN-based calligraphy systems, was to find trajectory generative models for the Chinese strokes; in contrast, this work aims to enable the robot to learn the writing preferences of humans by using HMI. Note that a Bayesian policy learning method was proposed in [27] to introduce preferences into an agent's movement trajectories. The proposed approach shares some of the human-in-the-loop ideas, but it is implemented using different machine learning approaches, feedback systems, and training procedures.

The remainder of this paper is organized as follows: Section II briefly summarizes deep reinforcement learning from human preferences, which is a basic technology used in the work. Section III specifies the implementations of the proposed learning framework. Section IV presents the experimental setup and discusses the experimental results. Section V concludes the paper and indicates important future work.

II. PRELIMINARY

Deep reinforcement learning from human preferences as reported in [26] is able to train a preference network to reflect a human's preference, where the preference feedback is used as the reward function for a reinforcement learning (RL) system. The basic idea of the system is shown in Fig. 1. Human users are required only to mark one of four preference options (i.e., "First one", "Second one", "Both", and "Neither") to pairs of video clips, each of which shows one action of an agent. Thus, the agent uses human feedback for training, rather than a predefined reward function like other conventional reinforcement learning methods. A reward predictor is created to learn human preferences; then, the predictor's

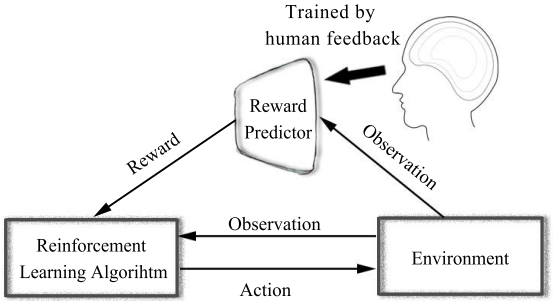


FIGURE 1. Deep reinforcement learning with human preference.

output is used as the reward signal of the RL system. The marked video data of human preferences are used to train the reward predictor. The learning process is outlined in the next two subsections.

A. HUMAN PREFERENCE SELECTION

The learning system first generates a number of samples for user selection. As a reinforcement learning system, a policy, $\pi(\boldsymbol{a}_t|\boldsymbol{o}_t)$, must interact with the agent's working environment to obtain a set of trajectory samples, $\{\boldsymbol{\tau}^0, \boldsymbol{\tau}^1, \ldots, \boldsymbol{\tau}^{N-1}\}$. In the training phase, two action segments, $(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2)$, randomly selected from the robot's actions $\{\boldsymbol{\tau}^0, \boldsymbol{\tau}^1, \ldots, \boldsymbol{\tau}^{N-1}\}$, are presented in the form of video clip to a human user for comparison. The human user evaluates the performance as presented in the two video clips and then chooses a preferable one out of four preference options, i.e., the first one, the second one, both of them, and neither of them. A Bernoulli distribution, as shown in Table 1, is adopted to define the preference. In the table, a random variable, Y, indicates which one of the two given segments, $\boldsymbol{\sigma}^1$ and $\boldsymbol{\sigma}^2$, is preferable. In this table, μ denotes the parameters of the Bernoulli distribution. Thus, the values of μ for the three options"First one", "Second one", and "Both" are set to 1, 0, and 0.5, respectively. For the option "Neither", the two segments are abandoned. In fact, the selection process effectively assigns scores to the two presented segments of the video.

TABLE 1. The Bernoulli distribution indicating the preferred segment.

Y	$\boldsymbol{\sigma}^1$	$\boldsymbol{\sigma}^2$
$P_{human}(Y)$	μ	$(1-\mu)$

B. TRAINING OF THE REWARD PREDICTOR

The predictor is implemented by a neural network, $\hat{r}_{\boldsymbol{\theta}}^{HP}(\boldsymbol{\sigma}^i)$, to predict the human user' preferences over the performance of agents. To obtain accurate predictions from the collected preference data, a probability distribution, $P_{human}(Y)$, is established for the reward predictor; thus, $P_{human}(Y)$ is defined as:

$$\hat{P}_{\boldsymbol{\theta}}(Y=\boldsymbol{\sigma}^1) = \frac{\exp \hat{r}_{\boldsymbol{\theta}}^{HP}(\boldsymbol{\sigma}^1)}{\exp \hat{r}_{\boldsymbol{\theta}}^{HP}(\boldsymbol{\sigma}^1) + \exp \hat{r}_{\boldsymbol{\theta}}^{HP}(\boldsymbol{\sigma}^2)}, \quad (1)$$

where $\boldsymbol{\theta}$ denotes the parameter of the preference network. To train the reward function, the optimal parameters of the preference network must be obtained by minimizing the cross-entropy loss between the preference network's output and the actual scores given by the user. Thus, the entropy loss is computed as:

$$loss(\boldsymbol{\theta}) = -\sum_{k=0}^{K-1} [\mu_k \log \hat{P}_{\boldsymbol{\theta}}(\boldsymbol{\sigma}_k^1) + (1-\mu_k) \log \hat{P}_{\boldsymbol{\theta}}(\boldsymbol{\sigma}_k^2)]. \quad (2)$$

In the learning system proposed in [26], any RL algorithm can be used to optimize the policy, such as asynchronous advantage actor-critic (A3C) [28], trust region policy optimization (TRPO) [29] and proximal policy optimization (PPO) [30], but none of these is very efficient for calligraphy robots. Therefore, in our application scenario, the training efficiency of these optimization methods must be improved.

III. PROPOSED FRAMEWORK

A learning framework for robotic writing to obtain the robotic writing trajectory ability with human preferences is presented in this section. The proposed framework is shown in Fig. 2; it consists of a trajectory generative module, a human preference network, a discriminative network, and a robotic arm. First, the trajectory generative module generates writing trajectories of ten numerals. The generative module consists of a generative network ($G_{\boldsymbol{\eta}}(z)$) and a multivariate normal distribution. The output of $G_{\boldsymbol{\eta}}(z)$ is the parameters of the multivariate normal distribution, which generates each numeral's trajectory. By using the generated trajectories, the robotic arm writes the corresponding numerals on a white board. Then, the writing results are captured by a camera, and both the preference network ($H_{\boldsymbol{\theta}}(\boldsymbol{x})$) and discriminative network ($D_{\boldsymbol{\omega}}(\boldsymbol{x})$)

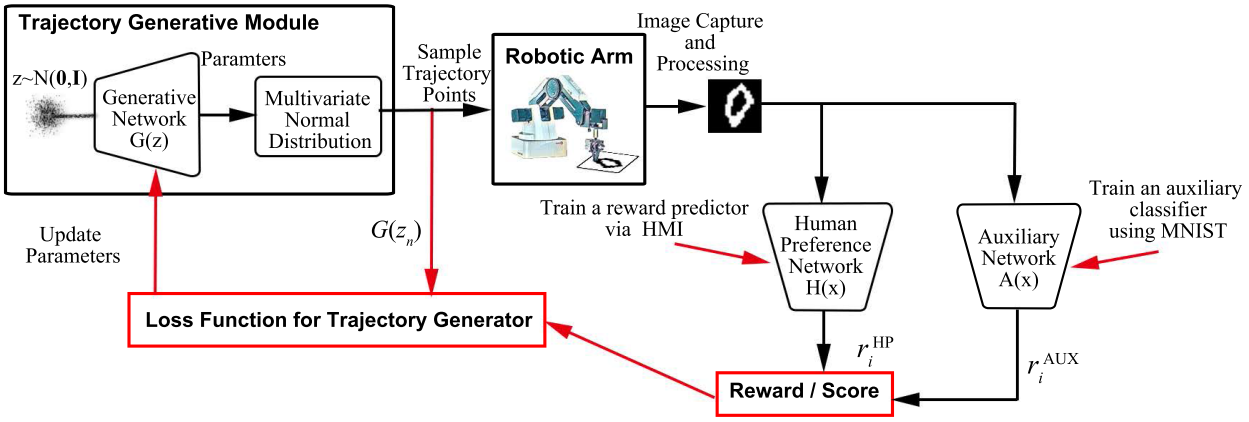


FIGURE 2. The proposed framework for robotic writing system to obtain the numeral writing ability with human preferences. The framework is a closed-loop system that is mainly composed of a trajectory generative module, a preference network, a discriminative network, and a robot system.

evaluate the quality and correctness of the writing results and generate two reward signals, r^{HP} and r^{D}, respectively. r^{HP} and r^{D} are then combined into one reward score to jointly obtain the loss values ($loss(\boldsymbol{\omega})$), which are used to update the parameters of $G_{\boldsymbol{\eta}}(z)$ in the trajectory generative module.

In the proposed framework, the human preference network mimics a human user to provide the preferences over the writing results from a writing robot; therefore, the preference network plays an important role in optimizing the trajectory generative module. The training process of the preference network is based on HMI. However, if the feedback is given by the preference network only, the calligraphy robot cannot rapidly generate a basic writing ability in the early stages of the training process. This is because a generative network with random initialization cannot possess any writing abilities for numerals, and the HMI cannot supply sufficient information to train the generative network. As a result, the optimization of the trajectory generative module consumes a large amount of training time. To overcome this limitation, a discriminative network is designed to improve the learning efficiency of the robotic writing system. In the early stage of training, the feedback of the discriminative network responds whether the robot writes the correct numerals. The implementations of the proposed framework and the training process are specified in the rest of this section.

A. TRAJECTORY GENERATIVE MODULE

The trajectory generative module aims to generate writing trajectories for the calligraphy robot. The module is composed of a generative network, $G_{\boldsymbol{\eta}}(z)$, and a multivariate normal distribution, $\mathcal{N}(G_{\boldsymbol{\eta}}(z_n), \boldsymbol{\Sigma})$. The module's input is a noise vector, z, and the output is the mean of the multivariate normal distribution, where $G_{\boldsymbol{\eta}}(z)$ is responsible for generating the parameters of the multivariate normal distribution. Note that the noise vector, z, is a random vector that follows spherical multivariate normal distribution, $\mathcal{N}(0, \boldsymbol{I})$. $G_{\boldsymbol{\eta}}(z)$ is a feed-forward neural network with four fully connected layers. The input layer contains 10 neurons; the two hidden layers contain 100 and 50 neurons, separately; and the output layer contains $3 \times N_p$ neurons; in addition, rectified linear units (ReLUs) are used as the transfer function in the hidden layers.

The multivariate normal distribution and the generative network, i.e., $\mathcal{N}(G_{\boldsymbol{\eta}}(z_n), \boldsymbol{\Sigma})$, jointly determine the numeral trajectory information. The probability density of multivariate normal distribution is defined by:

$$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{D/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m})^{T}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{m})\right], \quad (3)$$

where $\boldsymbol{x}$ denotes a D-dimensional random vector, and $\boldsymbol{m}$ and $\boldsymbol{\Sigma}$ denote the mean and covariance matrix of the distribution, respectively. Notice that in the proposed framework, $\boldsymbol{m}$ is determined by $G_{\boldsymbol{\eta}}(z)$, and the covariance matrix is a hyper-parameter.

In the proposed framework, the writing trajectory of a numeral is defined as a sequence of position values of a pen tip. The end-effector position of a calligraphy robot is considered as a $3 \times N_p$-dimensional random vector given by:

$$\boldsymbol{\xi} = (p_{x,1}, p_{y,1}, p_{z,1}, p_{x,2}, p_{y,2}, p_{z,2}, \ldots, p_{x,T}, p_{y,T}, p_{z,Np}), \quad (4)$$

where $p_{x,t}, p_{y,t}, p_{z,t}$ jointly represent the position of the end-effector in the space at a moment, t; N_p is the number of trajectory points; $p_{x,t}, p_{y,t} \in [0, 28], t = 1, 2, \ldots, N_p$ determine the two-dimensional trajectory of a numeral; and $p_{z,t} \in [0, 5], t = 1, 2, \ldots, T$ determines the pressure sequence of the writing brush.

B. PREFERENCE NETWORK

In the proposed learning framework, the input of the preference network, $H_{\boldsymbol{\theta}}(\boldsymbol{x})$, is numeral images written by the calligraphy robot, rather than the conventional set of observations and actions of the robot; the output is a value that represents the evaluation feedback of the inputs. Note that the numeral images must be processed and converted to relatively low-dimensional (28×28) grayscale images prior to being sent to the preference network. To better extract image features, a convolutional neural network is used as the preference network. The convolutional neural network consists of two convolution layers, two pooling layers, and two fully connected layers. The first convolution layer has 10 kernels with $strip = 2$ and zero-padding, and the first pooling layer has a 2×2 kernel. The second convolution layer contains 20 kernels, and the second pooling layer also has a 2×2 kernel. Then, the second pooling layer is followed by the two fully connected layers.

The training of the preference network, $H_{\boldsymbol{\theta}}(\boldsymbol{x})$, is inspired by the deep reinforcement learning method described in Section II. The training process of $H_{\boldsymbol{\theta}}(\boldsymbol{x})$ via HMI is shown in Fig. 3. The robot first writes a set of numerals using the trajectories generated by the generative module. In a certain training iteration, a pair of the written numerals are presented to the user via a graphical user interface (GUI) window, as shown in Fig. 4. Then, the user chooses one of the four preference options: "Left one", "Right one", "Both" and "Neither".

"Left one" means that the left writing result, $\boldsymbol{x}_k^1$, is preferable; in this case, μ_k is set to 1. "Right one" means that the right writing result, $\boldsymbol{x}_k^2$, is preferable; in this case, μ_k is set to 0. "Both" means that $\boldsymbol{x}_k^1$ and $\boldsymbol{x}_k^2$ are equally preferable; in this case, μ_k is set to 0.5. "Neither" means that neither of the results is preferable; in this case, the two images are abandoned. The remaining pairs of images, $(\boldsymbol{x}_k^1, \boldsymbol{x}_k^2)$, and parameters, μ_k, are retained in two data sets: $\mathcal{D} = \{(\boldsymbol{x}_0^1, \boldsymbol{x}_0^2), (\boldsymbol{x}_1^1, \boldsymbol{x}_1^2), \ldots, (\boldsymbol{x}_{K-1}^1, \boldsymbol{x}_{K-1}^2)\}$ and $\mathcal{M} = \{\mu_0, \mu_1, \ldots, \mu_{K-1}\}$. These data sets are used to optimize the preference network by using the Adam method [31] with the loss function of the preference network, which is defined in Eqs. (1) and (2). The above training process is summarized in Algorithm 1.

C. DISCRIMINATIVE NETWORK

The objective of the discriminative network, $D_{\boldsymbol{\omega}}(\boldsymbol{x})$, is to improve the learning efficiency of the robotic calligraphy

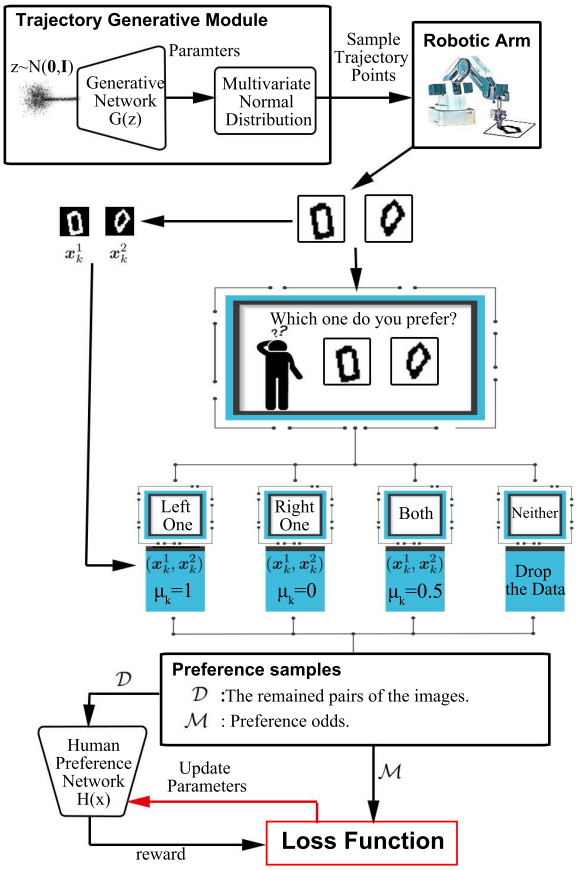


FIGURE 3. The training process of the preference network via HMI.

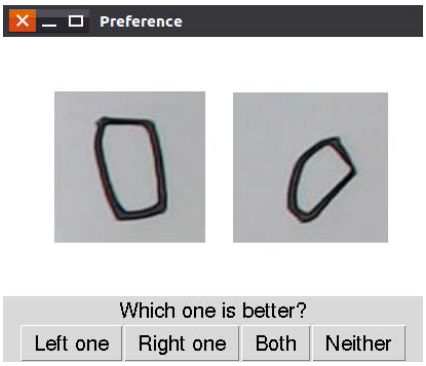


FIGURE 4. Graphical user interfaces for human preference selection, including four preference options: "Left one", "Right one", "Both" and "Neither".

system, which is done by providing an additional feedback that indicates whether the robot writes the desired numerals correctly. Similar to the preference network, the discriminative network also uses the processed numeral images (i.e., 28×28 grayscale images) as input, and its output is a value that represents the reward signal. $D_{\boldsymbol{\omega}}(\boldsymbol{x})$ is also a convolutional neural network, which consists of two convolution, two pooling, and two fully connected layers. The first convolution layer contains 10 kernels with $strip = 2$ and zero-padding,

Algorithm 1 Training Process of the Preference Network via HMI

Require: Initial parameters of human preference, $\boldsymbol{\theta}_0$; writing trajectory generator, $G_{\boldsymbol{\eta}}(z)$; the number of iterations, T; the number of human selections in each iteration, K.

Ensure: optimal $\boldsymbol{\theta}_\Delta$.

for each $i \in [0, T]$ **do**
  Input noise $z \sim \mathcal{N}(0, \boldsymbol{I})$ and obtain writing trajectories, $\{\boldsymbol{\xi}_0, \boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{N-1}\}$, from $\boldsymbol{\xi}_n \sim \mathcal{N}(G_{\boldsymbol{\eta}}(z_n), \boldsymbol{\Sigma})$.
  Robot performs writing; after image capture and processing, the images of the writing results, $\{\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{N-1}\}$, are obtained.
  for each $k \in [0, K-1]$ **do**
    Randomly choose a pair of images, $(\boldsymbol{x}_k^1, \boldsymbol{x}_k^2)$, without replacement; obtain *option* from a user.
    switch *option* **do**
      case "Left one"
        $\mu_k = 1$.
      case "Right one"
        $\mu_k = 0$.
      case "Both"
        $\mu_k = 0.5$.
      case "Neither"
        Abandon $(\boldsymbol{x}_k^1, \boldsymbol{x}_k^2)$.
  end for
  The image data, $\mathcal{D} = \{(\boldsymbol{x}_0^1, \boldsymbol{x}_0^2), (\boldsymbol{x}_1^1, \boldsymbol{x}_1^2), \ldots, (\boldsymbol{x}_{K-1}^1, \boldsymbol{x}_{K-1}^2)\}$, and the parameters of the Bernoulli distribution, $\mathcal{M} = \{\mu_0, \mu_1, \ldots, \mu_{K-1}\}$, are obtained. Take a step from $\boldsymbol{\theta}_i$ to $\boldsymbol{\theta}_{i+1}$, using Adam with the loss function: $loss(\boldsymbol{\theta}) = -\sum_{k=0}^{K-1}[\mu_k \log \hat{P}_{\boldsymbol{\theta}}(\boldsymbol{x}_k^1) + (1-\mu_k)\log \hat{P}_{\boldsymbol{\theta}}(\boldsymbol{x}_k^2)]$
end for

and the first pooling layer has a 2×2 kernel. The second convolution layer contains 20 kernels, and the second pooling layer has a 2×2 kernel as well. Then, the second pooling layer is followed by the two fully connected layers.

The training of the discriminative network is based on that of the generative adversarial network (GAN) [32]. The original objective function of GAN is represented as:

$$\min_{\boldsymbol{\omega}} \max_{\boldsymbol{\eta}} V(D, G) = E_{\boldsymbol{x}\sim p_{data}}[-\log D_{\boldsymbol{\omega}}(\boldsymbol{x})] + E_{z\sim p_z}[-\log(1 - D_{\boldsymbol{\omega}}(G_{\boldsymbol{\eta}}(z)))], \quad (5)$$

where p_{data} is the real probability distribution of images; p_z represents the probability distribution of noise; $D_{\boldsymbol{\omega}}(\cdot)$ denotes the discriminative network; and $G_{\boldsymbol{\eta}}(\cdot)$ indicates the generative module. However, in this framework, the output of $G_{\boldsymbol{\eta}}(\cdot)$ is a writing trajectory, rather than an image. Therefore, the objective function is rewritten as:

$$\min_{\boldsymbol{\omega}} \max_{\boldsymbol{\eta}} V(D, G) = E_{\boldsymbol{x}\sim p_{data}}[-\log D_{\boldsymbol{\omega}}(\boldsymbol{x})] + E_{\boldsymbol{x}\sim p_z}[-\log(1 - D_{\boldsymbol{\omega}}(\mathcal{W}(G_{\boldsymbol{\eta}}(z))))], \quad (6)$$

where $\mathcal{W}(\cdot)$ denotes the writing process of robotic arm. Thus, the loss function of the discriminative network is

represented as:

$$loss(\boldsymbol{\omega}) = E_{\boldsymbol{x}\sim p_{data}}[-\log D_{\boldsymbol{\omega}}(\boldsymbol{x})] + E_{\boldsymbol{x}\sim p_z}[-\log(1 - D_{\boldsymbol{\omega}}(\mathcal{W}(G_{\boldsymbol{\eta}}(\boldsymbol{z}))))]. \quad (7)$$

This loss function of the discriminative network is optimized by using the Adam algorithm [31]. The loss function of the generative module is specified in Section III-E. In addition, the following methods are adopted to ensure the stability of the proposed model: (1) The input noise of the generator is sampled from the latent space using a Gaussian distribution, not a uniform distribution. (2) In order to avoid overconfidence, the labels are smoothed, i.e., the real labels are replaced by random numbers between 0.7 and 1.2, and the fake labels are replaced by random numbers between 0.0 and 0.3. (3) The discriminative network is trained more frequently than the generative network.

D. ROBOTIC SYSTEM

The robotic hardware system is illustrated in Fig. 5; it consists of a 4-axis robotic arm and a camera mounted in a fixed position. A brush pen is mounted at the end-effector of the robotic arm. The working range of the arm is predefined. A writing board is placed flat in front of the robot. A coordinate conversion function, which converts the positions of the writing trajectories into the calligraphic robot coordinates, is defined as:

$$\begin{cases} p_x^{\gamma} = x_S + \gamma \cdot p_x \\ p_y^{\gamma} = y_S + \gamma \cdot p_y \\ p_z^{\gamma} = z_S + \gamma \cdot p_z, \end{cases} \quad (8)$$

where γ denotes a scale parameter controlling the size of the numerals; x_S, y_S, and z_S jointly define the initial position for each numeral; and p_x, p_y, and p_z jointly define the position of the end-effector at a certain point in time.

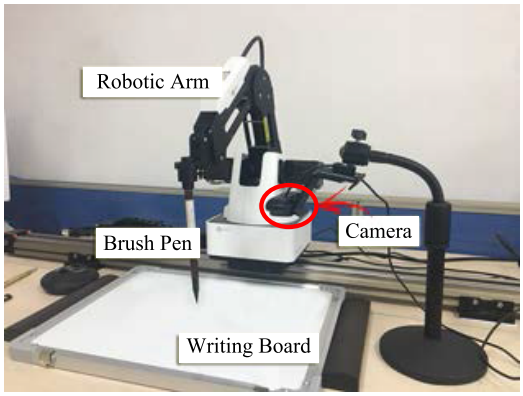


FIGURE 5. The robotic hardware.

Before writing a numeral, an inverse kinematics method as specified in the work of [25] converts every generated writing trajectory from the sequences of the end-effector positions to the robotic joint parameters. After writing a numeral, the end-effector returns to a predefined position. The written result is captured by the camera, processed by several image processing procedures, and then delivered to the preference and discriminative networks. Furthermore, to retain the consistency with the training data, the black numerals with a white background are then inverted to white lines with a black background.

E. TRAINING OF THE TRAJECTORY GENERATIVE MODULE

Obtaining the generative module with human preferences is equivalent to finding the optimal parameters of $G_{\boldsymbol{\eta}}(\boldsymbol{z})$. Two pieces of feedback, r^{HP} and r^{D}, are used for optimizing $G_{\boldsymbol{\eta}}(\boldsymbol{z})$, which are provided by the preference and discriminative networks, respectively. The reward feedback given by the preference network, $r^{HP} = H_{\boldsymbol{\theta}}(\boldsymbol{x})$, determines the estimated preference extent of the written numerals. The reward feedback given by the discriminative network, $r^{D} = D_{\boldsymbol{\omega}}(\boldsymbol{x})$, determines whether the robot writes the correct numerals.

A common solution of optimizing $G_{\boldsymbol{\eta}}(\boldsymbol{z})$ is to maximize the two rewards using the following gradient:

$$\Delta\eta^j \propto \frac{\partial r(r^{HP}, r^{D})}{\partial \eta^j}. \quad (9)$$

The solution uses the partial derivative of the combined rewards of r^{HP} and r^{D} with respect to the parameters of the generative network, $\{\eta^0, \eta^1, \ldots, \eta^{N_\eta}\}$. However, the writing is a physical process that cannot be expressed in an equation, and thus, the gradient cannot be calculated using the back-propagation algorithm. Therefore, the parameters cannot be updated using this common solution.

Alternatively, a maximum-likelihood-like method is used to optimize the trajectory generative module. By using the two feedback signals, the log likelihood of trajectories corresponding to highly rewarded written numerals can be maximized. Thus, the partial derivative of the likelihood with respect to the network's parameters and the reward signal can be used as the optimization direction and step size of the updating iterations. The network in the i-th iteration is given by:

$$\eta_{i+1}^j = \eta_i^j + r(r^{HP}, r^{D}) \cdot \frac{\partial \log p(G_{\boldsymbol{\eta}}(\boldsymbol{z}))}{\partial \eta_i^j}. \quad (10)$$

The above process is equivalent to minimize the following loss function:

$$loss(\boldsymbol{\eta}) = -\sum_{n}^{N-1}[\log p_{G_{\boldsymbol{\eta}}(\boldsymbol{z}_n)}(\boldsymbol{\xi}_n) \cdot r(r^{HP}, r^{D})], \quad (11)$$

where $\boldsymbol{\xi}_n$ is the n-th sample drawn from $\mathcal{N}(G_{\boldsymbol{\eta}_i}(\boldsymbol{z}_n), \boldsymbol{\Sigma})$. The combined reward function, $r(r^{HP}, r^{D})$, is defined as:

$$r(r^{HP}, r^{D}) = \alpha \cdot r^{HP} + (1 - \alpha) \cdot r^{D}, \quad (12)$$

where α is a variable that varies inversely with the current number of the iteration, i. Then, α is defined by:

$$\alpha = \frac{i}{T}. \quad (13)$$

Here, α is used to change the weights based on the two reward signals. r^D plays a more significant role in the early stage of training than r^{HP} does; however, r^{HP} becomes more significant in the later stages. The training process of the generative network is outlined in Algorithm 2.

Algorithm 2 Training Process of Writing Trajectory Model with Human Preferences

Require: Prepared numeral images as the samples of $p_{data}(\boldsymbol{x})$; initial parameters of discriminative, preference and generative network, $\boldsymbol{\omega}_0$, $\boldsymbol{\theta}_0$ and $\boldsymbol{\eta}_0$; covariance matrix for multivariate normal distribution, $\boldsymbol{\Sigma}$; the number of iterations, T; the number of the iteration at which to start HMI, H; the interval between human selections, S; batch size, N.

Ensure: optimal $\boldsymbol{\omega}_\Delta$, $\boldsymbol{\theta}_\Delta$ and $\boldsymbol{\eta}_\Delta$.

for each $i \in [0, T-1]$ **do**

Input noise $\boldsymbol{z} \sim \mathcal{N}(0, \boldsymbol{I})$ and get writing trajectories, $\{\boldsymbol{\xi}_0, \boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{N-1}\}$, from $\boldsymbol{\xi}_n \sim \mathcal{N}(G_{\boldsymbol{\eta}_i}(z_n), \boldsymbol{\Sigma})$.

Robot performs writing; after image capture and processing, the image samples, $\{\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{N-1}\}$, are obtained.

Training Discriminative Network:

Take a step from $\boldsymbol{\omega}_i$ to $\boldsymbol{\omega}_{i+1}$, using Adam with the loss function:

$$loss(\boldsymbol{\omega}) = E_{\boldsymbol{x}\sim p_{data}}[-\log D_{\boldsymbol{\omega}}(\boldsymbol{x})] + E_{\boldsymbol{x}\sim p_z}[-\log(1 - D_{\boldsymbol{\omega}}(\mathcal{W}(G_{\boldsymbol{\eta}}(z))))]$$

Training Preference Network:

if $i > H$ and i MOD $S = 0$ **then**

Update $\boldsymbol{\theta}_i$ using the method described in Algorithm 1.

end if

Training Generative Network:

Calculate the signal, $r_0, r_1, \ldots, r_{N-1}$, using:

$$r_n = \alpha H_{\boldsymbol{\theta}_i}(\boldsymbol{x}_n) + (1-\alpha) D_{\boldsymbol{\omega}}(\boldsymbol{x}_n),$$

where $\alpha = \frac{i}{T}$.

Take a step from $\boldsymbol{\eta}_i$ to $\boldsymbol{\eta}_{i+1}$, using Adam with the loss function:

$$loss(\boldsymbol{\eta}) = -\sum_n^{N-1}[\log p(G_{\boldsymbol{\eta}}(z_n)) \cdot r_n]$$

end for

The maximum-likelihood-like method is used here to solve the difficulty to represent the reasoning process as an equation and thus to calculate the partial derivative for the popular back-propagation solution. Note that other adaptive learning approaches may also be used here to represent the process and implement the trajectory generative module, such as [33], supported with the recent relevant development as of [34], which remains a topic for future work.

IV. EXPERIMENTATION

A. EXPERIMENTAL SETUP

The framework proposed above was applied to the tasks of writing ten Arabic numbers, i.e., the numerals from "0" to "9", by a calligraphic robot system to validate and evaluate the proposed system. For comparison, ten human users were invited to participate in these experiments to supply human preferences to the robot. The ten users selected numerals based on their owned preferences; thus, each numeral has different writing styles when the training is completed.

During the training process of the trajectory generative modules, each user made 300 selections for each numeral. In addition, the training data for the training of the discriminative network were randomly chosen from the "MNIST" [35] database of handwritten digits; each numeral has 500 images (i.e., 5,000 images in total in the database). The covariance matrix, $\boldsymbol{\Sigma}$ was an identity matrix; the numbers of trajectory points (N_p), iterations (T), and the iteration at which to start HMI (H) were set to 6, 800, and 500, respectively; the batch size (N) was set to 64; the interval between human selections (S) was set to 10; and the number of human selections in each iteration (K) was set to 10.

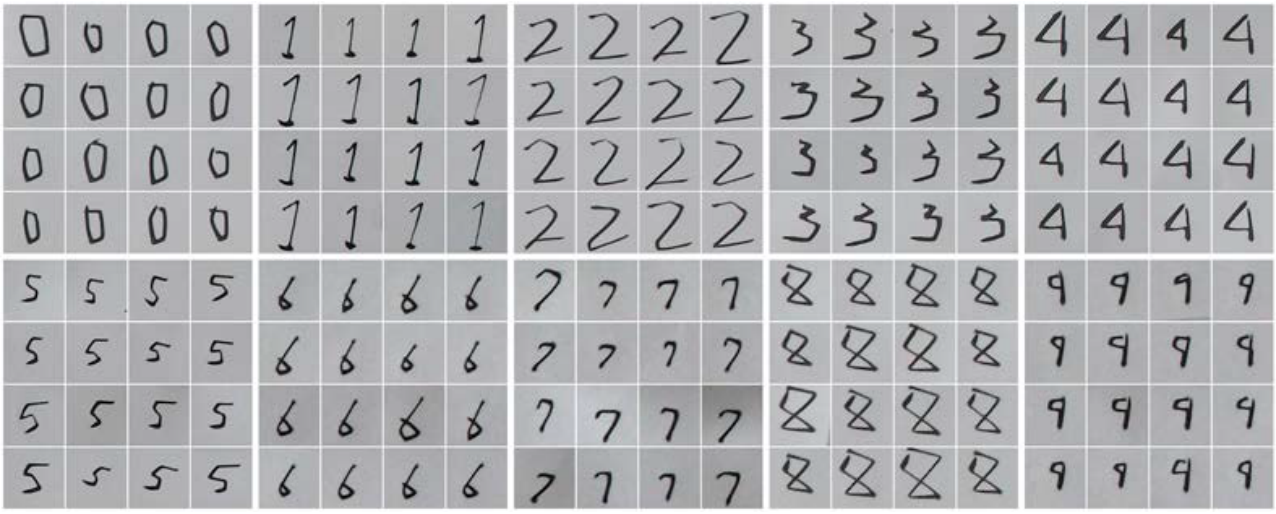
B. EXPERIMENTAL RESULTS

At the end of the experiment, the ten users were asked if the results met their expectations. As shown in Table 2, most of the users confirmed that the robotic writing results were basically consistent with their preferences. Two sets of selected writing results of the ten numerals after 800 training epochs are shown in Fig. 6. Fig. 6a demonstrates the final writing results with the preferences of selected User 1, and Fig. 6b shows the results with the preferences of selected User 2. Each numeral contains 16 images to show the writing diversity using the proposed framework. All the numerals were written by the calligraphy robot using the trajectories generated from the corresponding generative modules. This figure clearly demonstrates that all the writing results are slightly different from each other. This observation indicates that the proposed framework does not simply repeat a simple template to write the numerals; in contrast, the framework can exhibit a certain level of writing diversity.

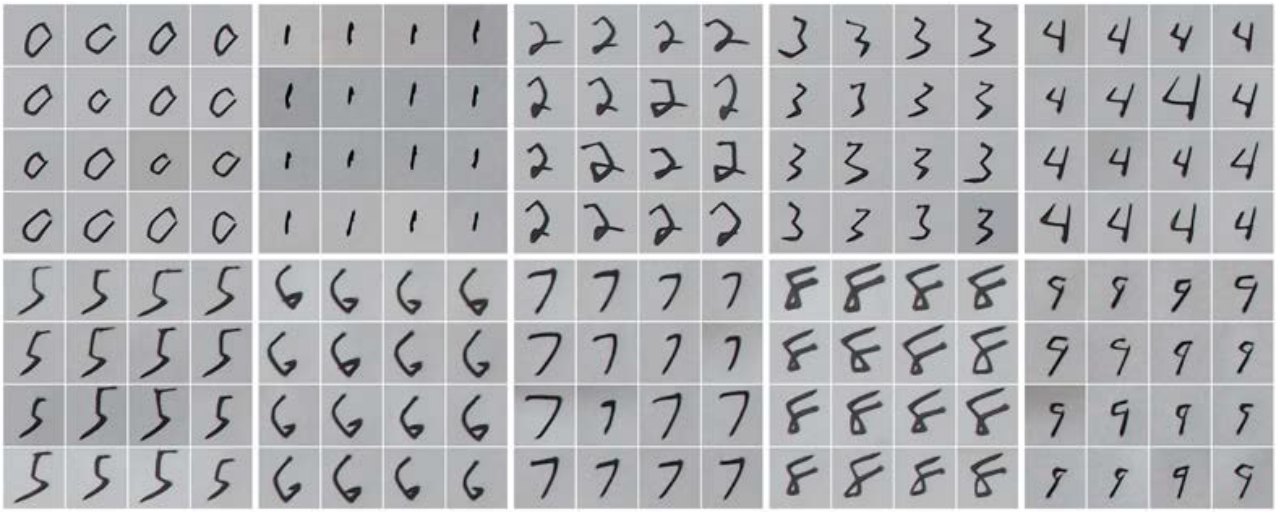
TABLE 2. The results of satisfaction survey.

The number of participant	Poor	Fair	Good
10	1	4	5

The two subfigures in Fig. 6 also illustrate the preference difference between the two selected human users. Additionally, the two subfigures indicate that the two types of writing results belong to different styles. For example, the writing results of "1" and "7" in Fig. 6a (with the preferences of user 1) were written in a form that is close to the printed character style; in particular, the writing results in the top row of "1" and the top-left corner of "7" are closer to the printed character style. However, the writing results in Fig. 6b are in a simpler style. Furthermore, several other details in the writing results also reveal the differences in writing styles between the two subfigures. The writing results of "0", "4", and "9" in Fig. 6b were written in italic type, whereas the numerals in Fig. 6a did not possess this feature. In addition, the results



(a) Writing results with the preferences of User 1



(b) Writing results with the preferences of User 2

FIGURE 6. Two sets of selected writing results of all 10 numerals. (a) shows the writing results drawn from the trajectory generative modules with the preferences of User 1, and (b) shows the writing results drawn from the trajectory generative modules with the preferences of User 2.

of "5"s in Fig. 6b were sharper than those in Fig. 6a. From these figures, it can be seen that User 1 preferred a formal writing style, whereas User 2 preferred a casual one.

C. ANALYSIS

To prove the effects of HMI on the final writing results, a comparative experiment was conducted. In this experiment, HMI and the preference network were eliminated from the learning framework of the robotic calligraphy. This means that only the discriminative network influenced the results. In addition, to analyze the differences in mathematics, the Fréchet inception distance (FID) [36] was used to compare the difference between the experimental results, which is given by:

$$d^2((\boldsymbol{m},\boldsymbol{C}),(\boldsymbol{m}_w,\boldsymbol{C}_w)) = ||m||_2^2 + Tr(\boldsymbol{C}+\boldsymbol{C}_w - 2(\boldsymbol{C}\boldsymbol{C}_w)^{1/2}), \quad (14)$$

where $\boldsymbol{m}$ and $\boldsymbol{C}$ denote the mean and covariance of the model samples, respectively, and $\boldsymbol{m}_w$ and $\boldsymbol{C}_w$ denote the mean and covariance of the samples from real world, respectively. In this work, FID is used to measure the similarity between the generated images and those included in the MNIST database.

The comparison results of FID are listed in Table 3. The table presents the FID scores of ten users, as well as FID scores from the robot framework without the preference network, i.e., GAN-based calligraphic robotic framework (labeled "Without Preference Network" in the table). Larger values indicate that the writing results are much more different from the training samples. Therefore, in each column, the smallest FID values are highlighted in bold font. "Without Preference Network" produced eight of the ten smallest FID values in the table; only the least values of the numbers "1"

TABLE 3. The FID values of the writing results.

Digit	User1	User2	User3	User4	User5	User6	User7	User8	User9	User10	Without Preference Network
1	163.7	98.93	85.66	67.79	109.16	107.54	**53.65**	99.76	139.50	90.19	63.96
2	103.45	89.97	69.24	124.9	112.1	108.26	70.22	67.72	111.53	55.48	**54.62**
3	64.46	100.65	51.40	53.02	94.22	89.73	65.366	48.29	90.52	42.52	**37.28**
4	54.26	72.56	38.85	76.37	43.46	85.76	55.01	48.34	85.70	62.68	**32.00**
5	44.87	125.14	51.17	110.87	129.86	40.43	80.34	94.05	83.81	38.96	**22.85**
6	110.54	49.96	87.31	102.55	66.17	36.29	75.44	73.88	75.35	86.92	**35.20**
7	123.34	87.90	92.77	41.80	80.77	98.47	51.71	51.81	104.47	46.92	**42.76**
8	74.52	42.12	75.73	88.10	77.20	53.83	**38.30**	82.13	97.68	68.65	49.78
9	124.86	127.35	65.14	77.03	52.53	60.32	67.40	80.99	79.97	92.49	**49.99**
10	58.56	68.77	40.95	89.51	47.89	77.42	50.67	66.28	46.32	63.24	**37.93**

and "7" were generated by User 7's preference network. This situation proved that the HMI effectively guided the robotic system to learn different writing styles. Once the preference network has truly learned each user's preferences, the whole system must generate the writing results that are different from those of the network trained by another user. Therefore, the FID values in Table 3 also demonstrated the different writing styles shown in Fig. 6.

It is interesting to investigate the causes of different writing styles. In the training process, via HMI, a human user provided data with his/her preferences for the learning system; then, the preference network used the data for training; next, the parameters of the trajectory generative module were updated by means of the reward signals from both the preference and discriminative networks. Therefore, the training data guided the trajectory generative module to fit the user's preferences. As the preference network played a significant role in this process, the robot can discern the user's preferences for the written numerals.

Based on the experimental results and the analysis, the proposed framework allows a robot to autonomously develop its writing skill, although the final writing results showed limited aesthetic effects compared to those written by a well-trained human. However, the success of the proposed framework verifies that effective interactions between humans and machines can boost the work performance of robots.

V. CONCLUSION

This paper proposes a new learning framework for robots to learn writing skills with human preferences. The proposed method incorporated human-machine interactions to obtain a trajectory generative model. The experimental results show that the proposed framework can successfully allow the robot to write numerals in accordance with human users' preferences. Moreover, the FID scores proved that the human-machine interaction effectively enables the robot to write in various writing styles.

While our proposed approach is promising, there is room for improvement. The current work ignored the writing sequence of numerals, i.e., the order of the trajectory points. The writing sequence is a difficult issue in robotic writing, especially for Chinese calligraphy [37], [38]; more research effort is required to further investigate the recurrent neural network to improve robotic writing.

ACKNOWLEDGMENT

The authors are very grateful to the anonymous reviewers for their constructive comments which have helped significantly in revising this work.

REFERENCES

[1] Z. J. Ju, X. Ji, J. Li, and H. Liu, "An integrative framework of human hand gesture segmentation for human–robot interaction," *IEEE Syst. J.*, vol. 11, no. 3, pp. 1326–1336, Sep. 2015.

[2] J. Liu, Y. Luo, and Z. Ju, "An interactive astronaut-robot system with gesture control," *Comput. Intell. Neurosci.*, vol. 2016, Mar. 2016, Art. no. 7845102.

[3] Z. Zuo, L. Yang, Y. Peng, F. Chao, and Y. Qu, "Gaze-informed egocentric action recognition for memory aid systems," *IEEE Access*, vol. 6, pp. 12894–12904, 2018.

[4] D. Zhou, M. Shi, F. Chao, C.-M. Lin, L. Yang, C. Shang, and C. Zhou, "Use of human gestures for controlling a mobile robot via adaptive cmac network and fuzzy logic controller," *Neurocomputing*, vol. 282, pp. 218–231, Mar. 2018.

[5] R. Moore, "PRESENCE: A human-inspired architecture for speech-based human-machine interaction," *IEEE Trans. Comput.*, vol. 56, no. 9, pp. 1176–1188, Sep. 2007.

[6] W. Jiang, F. Wen, and P. Liu, "Robust beamforming for speech recognition using DNN-based time-frequency masks estimation," *IEEE Access*, vol. 6, pp. 52385–52392, 2018.

[7] D. Kollias, G. Marandianos, A. Raouzaiou, and A.-G. Stafylopatis, "Interweaving deep learning and semantic techniques for emotion analysis in human-machine interaction," in *Proc. 10th Int. Workshop Semantic Social Media Adaptation Personalization*, Nov. 2015, pp. 1–6.

[8] N. Masuyama, C. K. Loo, and M. Seera, "Personality affected robotic emotional model with associative memory for human-robot interaction," *Neurocomputing*, vol. 272, pp. 213–225, Jan. 2018.

[9] I. Shahin, A. B. Nassif, and S. Hamsa, "Emotion recognition using hybrid Gaussian mixture model and deep neural network," *IEEE Access*, vol. 7, pp. 26777–26787, 2019.

[10] B. Liu, Z. Ju, and H. Liu, "A structured multi-feature representation for recognizing human action and interaction," *Neurocomputing*, vol. 318, pp. 287–296, Nov. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925231218310282

[11] J. Q. Gan, B. A. S. Hasan, and C. S. L. Tsui, "A hybrid approach to feature subset selection for brain-computer interface design," in *Proc. Int. Conf. Intell. Data Eng. Automated Learn.*, 2011, pp. 279–286.

[12] Y. Sun, C. Li, G. Li, G. Jiang, D. Jiang, H. Liu, Z. Zheng, and W. Shu, "Gesture recognition based on kinect and sEMG signal fusion," *Mobile Netw. Appl.*, vol. 23, no. 4, pp. 797–805, Aug. 2018.

[13] L. Minati, N. Yoshimura, and Y. Koike, "Hybrid control of a vision-guided robot arm by EOG, EMG, EEG biosignals and head movement acquired via a consumer-grade wearable device," *IEEE Access*, vol. 4, pp. 9528–9541, 2016.

[14] S. H. Alsamhi, O. Ma, and M. S. Ansari, "Artificial Intelligence-Based Techniques for Emerging Robotics Communication: A Survey and Future Perspectives," 2018, *arXiv:1804.09671*. [Online]. Available: https://arxiv.org/abs/1804.09671

[15] H. Zeng, Y. Huang, F. Chao, and C. Zhou, "Survey of robotic calligraphy research," *CAAI Trans. Intell. Syst.*, vol. 11, no. 1, pp. 15–26, 2016.

[16] W. Li, Y. Song, and C. Zhou, "Computationally evaluating and synthesizing Chinese calligraphy," *Neurocomputing*, vol. 135, pp. 299–305, Jul. 2014.
[17] D. Berio, S. Calinon, and F. F. Leymarie, "Learning dynamic graffiti strokes with a compliant robot," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2016, pp. 3981–3986.
[18] D. Berio, S. Calinon, and F. F. Leymarie, "Generating calligraphic trajectories with model predictive control," in *Proc. 43rd Graph. Interface Conf.*, Waterloo, Ontario, Canada: Univ. Waterloo, 2017, pp. 132–139. doi: 10.20380/GI2017.17.
[19] B. Zhao, M. Yang, H. Pan, Q. Zhu, and J. Tao, "Nonrigid point matching of Chinese characters for robot writing," in *Proc. IEEE Int. Conf. Robot. Biomimetics*, Dec. 2017, pp. 762–767.
[20] F. Chao, J. Lv, D. Zhou, L. Yang, C.-M. Lin, C. Shang, and C. Zhou, "Generative adversarial nets in robotic Chinese calligraphy," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2018, pp. 1104–1110.
[21] X. Gao, C. Zhou, F. Chao, L. Yang, C.-M. Lin, T. Xu, C. Shang, and Q. Shen, "A data-driven robotic Chinese calligraphy system using convolutional auto-encoder and differential evolution," *Knowl.-Based Syst.*, vol. 182, Oct. 2019, Art. no. 104802. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0950705119302771
[22] Y. Sun, H. Qian, and Y. Xu, "Robot learns Chinese calligraphy from demonstrations," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Sep. 2014, pp. 4408–4413.
[23] F. Chao, F. Chen, Y. Shen, W. He, Y. Sun, Z. Wang, C. Zhou, and M. Jiang, "Robotic free writing of Chinese characters via human–robot interactions," *Int. J. Humanoid Robot.*, vol. 11, no. 1, 2014, Art. no. 1450007.
[24] F. Chao, Y. Huang, C.-M. Lin, L. Yang, H. Hu, and C. Zhou, "Use of automatic Chinese character decomposition and human gestures for Chinese calligraphy robots," *IEEE Trans. Human–Mach. Syst.*, vol. 49, no. 1, pp. 47–58, Feb. 2019.
[25] F. Chao, Y. Huang, X. Zhang, C. Shang, L. Yang, C. Zhou, H. Hu, and C.-M. Lin, "A robot calligraphy system: From simple to complex writing by human gestures," *Eng. Appl. Artif. Intell.*, vol. 59, pp. 1–14, Mar. 2017.
[26] P. F. Christiano, J. Leike, T. Brown, M. Martic, S. Legg, and D. Amodei, "Deep reinforcement learning from human preferences," in *Proc. Adv. Neural Inf. Process. Syst.*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Cambridge, MA, USA: MIT Press, 2017, pp. 4299–4307. [Online]. Available: http://papers.nips.cc/paper/7017-deep-reinforcement-learning-from-human-preferences.pdf
[27] A. Wilson, A. Fern, and P. Tadepalli, "A Bayesian approach for policy learning from trajectory preference queries," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1133–1141. [Online]. Available: http://papers.nips.cc/paper/4805-a-bayesian-approach-for-policy-learning-from-trajectory-preference-queries.pdf
[28] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu, "Asynchronous methods for deep reinforcement learning," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 1928–1937. [Online]. Available: http://proceedings.mlr.press/v48/mniha16.html
[29] J. Schulman, S. Levine, P. Abbeel, M. I. Jordan, and P. Moritz, "Trust region policy optimization," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 1889–1897. [Online]. Available: http://proceedings.mlr.press/v37/schulman15.html
[30] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," 2017, *arXiv:1707.06347*. [Online]. Available: https://arxiv.org/abs/1707.06347
[31] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.*, 2015, pp. 1–15. [Online]. Available: https://arxiv.org/abs/1412.6980
[32] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.
[33] L. Yang, F. Chao, and Q. Shen, "Generalized adaptive fuzzy rule interpolation," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 4, pp. 839–853, Aug. 2017.
[34] S. Jin, R. Diao, C. Quek, and Q. Shen, "Backward fuzzy rule interpolation," *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 6, pp. 1682–1698, Dec. 2014.
[35] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
[36] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local Nash equilibrium," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 6626–6637. [Online]. Available: http://papers.nips.cc/paper/7240-gans-trained-by-a-two-time-scale-update-rule-converge-to-a-local-nash-equilibrium.pdf
[37] H.-I. Lin and Y.-C. Huang, "Visual matching of stroke order in robotic calligraphy," in *Proc. Int. Conf. Adv. Robot.*, Jul. 2015, pp. 459–464.
[38] Z. Ma and J. Su, "Stroke reasoning for robotic chinese calligraphy based on complete feature sets," *Int. J. Social Robot.*, vol. 9, no. 4, pp. 525–535, Sep. 2017. doi: 10.1007/s12369-017-0410-2.

XINGEN GAO received the B.Eng. and M.Eng. degrees from the Xiamen University of Technology, in 2009 and 2013, respectively. He is currently pursuing the Ph.D. degree with the Department of Cognitive Science, Xiamen University. His current research interests include robotics, reinforcement learning, and evolutionary algorithms.

CHANGLE ZHOU received the Ph.D. degree from Peking University, in 1990. He is currently a Professor with the Cognitive Science Department, Xiamen University, the Director of the Fujian Provincial Key Laboratory of Brain-like Intelligent Systems, and the Director of the Laboratory of Art, Mind and Computation. He is also an affiliated Professor of linguistics and applied linguistics with the Humanity College, Zhejiang University, and an affiliated Professor of the Philosophy Department, Xiamen University. His scientific contribution to AI focuses on machine consciousness and the logic of mental self-reflection. Beyond AI projects, he also carries out research on a host of other topics, including computational brain modeling, computational modeling of analogy, metaphor and creativity, computational musicology, and information processing of data regarding traditional Chinese medicine. His philosophical works address ancient oriental thoughts of the Chinese, including ZEN, TAO, YI, and so on, viewed from a scientific perspective. His research interest includes artificial intelligence.

FEI CHAO (M'11) received the B.Sc. degree in mechanical engineering from Fuzhou University, China, in 2004, the M.Sc. degree (Hons.) in computer science from the University of Wales, Aberystwyth, U.K., in 2005, and the Ph.D. degree in robotics from Aberystwyth University, Wales, U.K., in 2009. He was a Research Associate with Aberystwyth University, from 2009 to 2010, under the supervision of his Ph.D. supervisor, Prof. Mark H. Lee. He is currently an Associate Professor with the Cognitive Science Department, Xiamen University, China. He is also a Research Fellow with the Department of Computer Science, Aberystwyth University. He has published more than 50 peer-reviewed journal and conference papers. His research interests include developmental robotics, reinforcement learning, and optimization algorithms.

Dr. Chao is a member of CCF and CAAI. He is the Vice Chair of the IEEE Computer Intelligence Society Xiamen Chapter.

LONGZHI YANG (M'12–SM'17) is currently the Director of learning and teaching and an Associate Professor with the Department of Computer and Information Sciences, Northumbria University, U.K. His research interests include computational intelligence, machine learning, big data, computer vision, intelligent control systems, robotics, and the applications of such techniques in real-world uncertain environments. He received the Best Student Paper Award at the 2010 IEEE International Conference on Fuzzy Systems. He is the Founding Chair of the IEEE Special Interest Group on Big Data for Cyber Security and Privacy.

CHIH-MIN LIN (M'87–SM'99–F'10) was born in Taiwan, in 1959. He received the B.S. and M.S. degrees from the Department of Control Engineering and the Ph.D. degree from the Institute of Electronics Engineering, National Chiao Tung University, Hsinchu, Taiwan, in 1981, 1983, and 1986, respectively. He was an Honor Research Fellow with the University of Auckland, Auckland, New Zealand, from 1997 to 1998. He is currently a Chair Professor and the Vice President of Yuan Ze University, Chung-Li, Taiwan. He has published more than 170 journal articles. His current research interests include fuzzy neural networks, cerebellar model articulation controllers, intelligent control systems, and signal processing. He serves as an Associate Editor of the IEEE TRANSACTIONS ON CYBERNETICS and the IEEE TRANSACTIONS ON FUZZY SYSTEMS.

CHANGJING SHANG received the Ph.D. degree in computing and electrical engineering from Heriot-Watt University, U.K. She was with Heriot-Watt, Loughborough, and Glasgow Universities. She is currently a University Research Fellow with the Department of Computer Science, Institute of Mathematics, Physics and Computer Science, Aberystwyth University, U.K. Her research interests include pattern recognition, data mining and analysis, space robotics, and image modeling and classification.