IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 23, 2020, accepted April 28, 2020, date of publication May 6, 2020, date of current version May 20, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2991767

Deep Learning-Based Hypothesis Generation Model and Its Application on Virtual Chinese Calligraphy-Writing Robot

WEI-YEN WANG, (Fellow, IEEE), MIN-JIE HSU, LI-AN YU, YI-HSING CHIEN, AND CHEN-CHIEN HSU, (Senior Member, IEEE)
Department of Electrical Engineering, National Taiwan Normal University, Taipei 106, Taiwan

Corresponding author: Chen-Chien Hsu (jhsu@ntnu.edu.tw)

This work was supported in part by the Chinese Language and Technology Center of the National Taiwan Normal University (NTNU) through the Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE), Taiwan, and in part by the Ministry of Science and Technology, Taiwan, through the Pervasive Artificial Intelligence Research (PAIR) Labs, under Grant MOST 109-2634-F-003-006 and Grant MOST 109-2634-F-003-007.

ABSTRACT In recent years, a tremendous amount of effort has been devoted to modeling the cognition of human brain, particularly hypothesis generation process. Most research of the hypothesis generation model is probability-based. However, computation of human brains is still neuron-based instead of calculating the probability. As an attempt to solve this problem in this paper, we propose a novel neuron-based hypothesis generation model, called hypothesis generation net, to model human cognition, including how to make decisions and how to do actions. Basically, the proposed hypothesis generation model consists of two parts, i.e., a hypothesis model and an evaluation model. When these two models interact, the system is able to generate hypotheses to solve complex tasks based on historical experiences. To validate the feasibility of the proposed hypothesis generation model, we show a virtual robot with its cognition system can learn how to write Chinese calligraphy in a simulation environment, where an image-to-action translation via a cognitive framework is proposed to learn the pattern of Chinese characters. Based on the proposed deep thinking and learning mechanism, the virtual robot is able to write Chinese calligraphy well, which is a difficult task requiring extremely complicated motions, through thinking and practicing according to a human writing sample.

INDEX TERMS Hypothesis generation model, deep neural networks, Chinese calligraphy, image-to-action translation.

I. INTRODUCTION

In recent years, artificial intelligence (AI) has dramatically affected human's life in many areas such as security, domotics, automatic system, face recognition, object recognition, market analysis, to name a few of them. Most of these research studies concern artificial narrow intelligence (ANI). However, devices, machines, and robots which need to adapt to a changeable environment require deep thinking and complex perception to handle uncertainties and make correct decisions. As a result, artificial general intelligence (AGI) is becoming an important topic for investigation by many researchers. AGI is a kind of strong AI which attempts to model human cognition and human mind. One of the key elements of AGI's kernel is the cognition system. Traditionally, cognitive psychology includes several parts, e.g., reasoning, memory, and perception. Among them, hypothesis generation model [1] is an important research topic for reasoning as to how a human makes decisions by generating possible states based on historical experiences to solve a problem. In a hypothesis generation structure, the decision maker requires the actual state of the world in order to rectify the behavior if the current state is wrong. To the knowledge of the authors, most research investigating hypothesis generation model is probability-based. That is, the posterior distribution is calculated to make new inferences based on historical experiences [2]. However, computation of human brains is nevertheless neuron-based instead of calculating the

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

probability. As an attempt to solve this problem in this paper, we propose a novel neuron-based hypothesis generation model, called hypothesis generation net, to model human cognition, including how to make decisions and how to do actions.

In the last few years, deep neural networks have made a series of breakthroughs. They are widely utilized in images classification [3]–[7], objects detection [8], as well as voice synthesis or image translation [9], [10]. Autoencoder (AE) [11], [12] is a kind of unsupervised learning neural network which learns and extracts features automatically. The hidden layer of AE consists of two parts, i.e., the encoder and the decoder. The aim of the encoder is to compress an input into a set of latent vectors. Then, these latent vectors can be processed by a decoder to reconstruct the input. Traditional AE is usually utilized for dimensionality reduction or feature extraction. In recent years, AE has been widely applied in generating images, including converting picture colors, removing watermarks, denoising images, etc. As a result, there have been various types of research on autoencoder, such as variational autoencoder [13], denoising autoencoder [14], sparse autoencoder [15], etc. Another related method in unsupervised learning is generative adversarial networks (GANs) [16]–[18], which utilize a discriminator model to classify output images into 'real' or 'fake' and utilize a generator model to produce 'fake' images which the discriminator model cannot distinguish from 'real' images. The GANs model has inspired many subsequent works for image synthesis, such as DCGAN [19] and Deepfake algorithm [20], [21], which can swap one person's face with another in a video or an image. Motivated by AE and GAN, a neuron-based hypothesis generation model is established in this paper. Through deep learning realization, the proposed hypothesis generation model has the ability to learn and generate hypotheses by practicing based on historical experiences, addressing the problem of image to action translation.

To validate the feasibility of the proposed hypothesis generation model, we show a virtual robot with its cognition system can learn how to write Chinese calligraphy in a simulation environment through thinking and practicing from a human writing sample. Chinese calligraphy writing, which is regarded a difficult task requiring extremely complicated motions [22]–[25], focuses on changing the speed, press, strength, orientation, and angle [26] of a writing brush to write aesthetic calligraphy. It is complicated for designers to analyze the strokes of characters in different styles. Therefore, profound skills are needed to write Chinese characters well. Pressing the brush heavily or lightly causes the stroke of the Chinese characters to become thick or thin, respectively. Moreover, the turning angle and timing for manipulating the brush are also important. Given the challenges, there have been researches focusing on the development of Chinese calligraphy-writing robots [22]–[31]. To simplify the tasks required, most of image-based researches utilized 3-axis vector $[x, y, z]$ to control the robot to write Chinese calligraphy [24]–[29] because 6-axis $[x, y, z, roll, pitch, yaw]$ motion planning for Chinese calligraphy writing is a complex task for robots. It is intuitive to extract the position component $[x, y, z]$ from a Chinese calligraphy character by skeletonization and thickness of the calligraphy characters. However, the orientation and tilt of the writing brush are much more complicated to calculate because Chinese calligraphy characters can be written with many different motions. That is, different motions can achieve the same writing result. The relationship between motion and writing result is not a one-to-one, but a many-to-one mapping function. While the generation of position vector sequences for the writing brush is straightforward through machine vision operations, the combinations of orientation and tilt sequences, however, are extremely numerous for the writing brush. Therefore, it is difficult to generate coordinates of roll, pitch and yaw of the writing brush from a human writing sample by directly using computer vision methods. In light of the above difficulties, it is therefore our objectives to apply the proposed neuron-based hypothesis generation model to a virtual robotic system through a simulation environment where the virtual robot with its cognition system can learn and think how to write Chinese characters well by practicing.

The rest of this paper is organized as follows. Section 2 discusses the related works. Section 3 presents the proposed hypothesis generation model. In Section 4, we apply the hypothesis generation model to a virtual robotic calligraphy writing system. Simulation results are shown in Section 5 to verify the performances of the proposed method. The conclusions are drawn in Section 6.

II. RELATED WORKS

To build an artificial cognitive system to model the hypothesis generation process, every single neuron of deep neural networks is important. By connecting multiple neurons, we can construct a system to simulate the structure of a human brain to fulfill the function of reasoning and judgement. Without hypothesis generation processes, the system is not able to understand the surroundings and learn by itself. Therefore, deep neural networks are utilized to realize the hypothesis generation process to model the psychological learning process of human beings to accomplish different types of tasks.

In a hypothesis generation model, most investigations indicate that the hypotheses made by humans come close to the Bayesian model [1], [21], [32], [33], where inference comes from hypothesis generation and evaluation as:

$$\mathrm{P}\left(h|d\right) = \frac{P\left(d|h\right)P\left(h\right)}{\sum\limits_{h' \in H} P\left(d|h'\right)P\left(h'\right)}. \tag{1}$$

where H is a complete set of hypotheses, $h, \ h' \in H$, d is the sensor input, $P\left(h|d\right)$ is a posterior probability to hypothesis h, $P(h)$ denotes its prior probability, and $P\left(d|h\right)$ represents the likelihood of the sensory input data under hypothesis h. Because H is a complete set of hypotheses, it is impossible to generate the whole space of hypotheses in many cases.

To solve the approximation of posterior probability with less biases coming from the incomplete hypotheses, Markov chain Monte Carlo (MCMC) method can help approximate the posterior probability by (2) as:

$$\mathrm{P}_N(h|d) = \frac{1}{N}\sum_{n=1}^{N} f(h_n = h) \tag{2}$$

where $f(\cdot)$ is 1 if the statement is true, otherwise is 0. h_n is a random sample hypothesis from the Markov chain. If N goes to infinity, we obtain a non-bias approximation of the posterior probability.

However, the computing units in human brains are neurons. That is, the decision, memory, and perception come from a central nervous system. Even though much research supports that MCMC can also be explained with neuroscience as cortical circuits, the hypothesis generation from humans can be regarded as a complicated neural network. Actually, all of the hypotheses are from neural computing in human brains. It is therefore possible for us to design deep neural networks to simulate the hypothesis generation process. Referencing to the concept of AE and GANs, we thus propose a neural network architecture to model the hypothesis generation process.

AE is a type of unsupervised learning, which was first introduced by Zhu *et al.* [10]. The method is utilized to compress an input into a latent vector via an encoder. This latent vector usually presents the important part of the data. The encoder also has the function of nonlinear dimension reduction. After that, the decoder utilizes the latent vector to reconstruct the input data. Comparing the inputs with the outputs, we learn the weights of the encoder and decoder according to the loss function $\sum_{i=1}^{n}(x_i - \hat{x}_i)^2/n$. Fig. 1 shows the schematic diagram of the autoencoder.

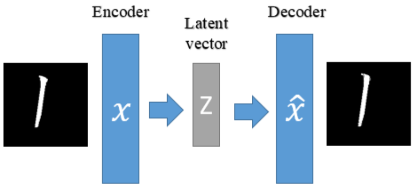


FIGURE 1. Schematic diagram of autoencoder (AE).

Ng [15] introduced GANs, which are deep neural net architectures for training a generative model via an adversarial process. GANs consist of two nets, i.e., a generator net G and a discriminator net D. The generator G generates samples from a prior noise distribution and the discriminator D is trained to distinguish whether the samples come from the real data distribution or the generator's distribution. The generator is then trained to compete with the discriminator D by minimizing $\log(1-D(G(z)))$, so that the discriminator is unable to distinguish whether the samples are real data or generator's data.

III. HYPOTHESIS GENERATION MODEL

To describe the proposed neuron-based hypothesis generation model, we utilize a virtual robotic system as an application to better illustrate the derivation process. Through the realization of the hypothesis generation process by neural networks, we can construct a cognition model for the robotic system to make hypotheses from historical experiences. Based on the hypothesis generation model, the virtual robot can learn how to write Chinese calligraphy. Instead of using top-down strategy to learn writing Chinese calligraphy, we utilize bottom-up strategy to build the cognition architecture to learn with the neural networks.

Fig. 2 illustrates the architecture of the proposed hypothesis generation model for application to a robotic system, which consists of two parts, i.e., a hypothesis model and an evaluation model. The hypothesis model makes hypotheses to solve the problem according to the past experiences stored in DNN1. The function of the evaluation model is to judge the hypotheses. The result which the virtual robot observes is stored in DNN2 so that the virtual robot can recall the result and the historical experiences in the future to help DNN1 make a new hypothesis by judging the previous hypothesis from DNN1.

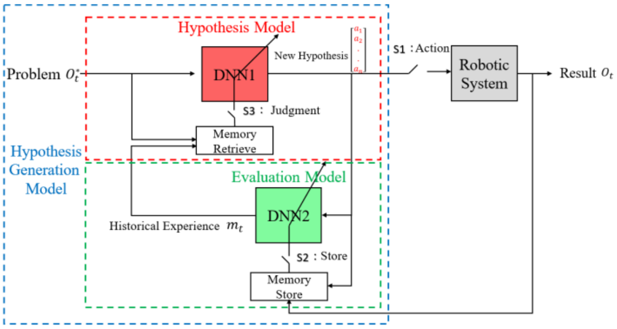


FIGURE 2. Architecture of the proposed hypothesis generation model.

For instance, when we need the virtual robot to pick a bottle, the hypothesis model produces an action vector as the angles for controlling the motors. Then, we close switch s1 so that the virtual robot can execute the action vector which is received from DNN1. Then, the evaluation model stores the result and the hypothesis in DNN2 by closing switch s2. If the observed vector O_t is not "pick a bottle", the hypothesis model needs to make a new hypothesis according to historical experiences. To make a new hypothesis, we connect DNN1 by closing switch s3, which makes the next hypothesis. DNN2, which stores historical experience, helps compute the gradient of the error with the vector m_t and the expected observed vector O_t^* to update only DNN1. This update law is similar to the generator's update of GANs, but this architecture represents a general form for various robotic systems. Through several iterations, we store the best hypothesis according to the optimization criterion $\min(\|O_t^* - m_t\|)$. Note that we do not need to know the relationship between the action vector and the task ''pick the bottle'' because the virtual robot will think and learn the concept by itself.

IV. HYPOTHESIS GENERATION MODEL-BASED CONTROL FOR VIRTUAL ROBOTIC CALLIGRAPHY-WRITING SYSTEM

A. CALLIGRAPHY-WRITING ROBOT WITH HYPOTHESIS GENERATION NET

Chinese calligraphy-writing represents a big challenge for a robot if the coordinates are not prescheduled. Even with computer vision, it is still difficult to calculate 6-axis coordinates [x, y, z, *roll*, *pitch*, *yaw*] for the robot to write Chinese calligraphy. We know the relationship between 2D coordinates [x, y] and the Chinese calligraphy image by image processing, but the other coordinates [z, *roll*, *pitch*, *yaw*] are still difficult to design. It is therefore of significance to implement the proposed hypothesis generation model, so that a virtual robot can think and learn how to figure out the method of writing Chinese calligraphy. To prevent the time-consuming process in learning to write Chinese calligraphy in a real environment, we utilize a virtual robotic system [34] shown in Fig. 3 that we developed to simulate the process of brush writing.

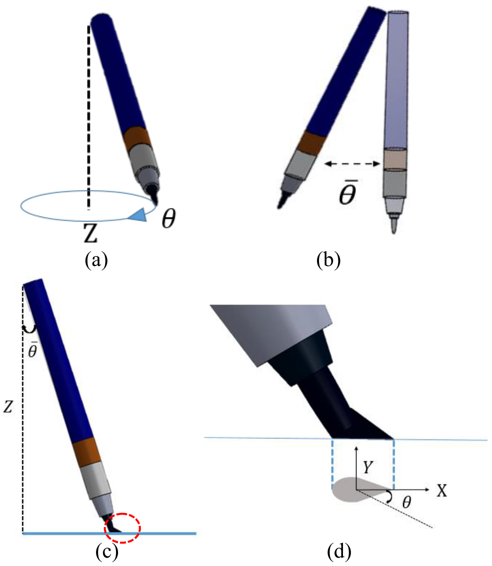


FIGURE 3. Virtual robotic system [34] showing the pose of the writing brush. (a) The rotation angle θ. (b) The tilt angle $\bar{\theta}$. (c) Schematic diagram illustrating the brush writing a character. (d) Partial enlargement of the brush.

To write Chinese calligraphy characters in this paper, 5-axis reduced form $\left[X, Y, Z, \theta, \bar{\theta}\right]$ without the spin axis, instead of using 6-axis [x, y, z, *roll*, *pitch*, *yaw*] form, is utilized to describe the Cartesian coordinate [34], the rotation angle, and the tilt angle of the writing brush. This is because the writing brush seldom spins when writing Chinese calligraphy. The vector [X, Y] represents the Cartesian coordinate, and [Z] represents the vertical axis coordinate which renders Chinese characters being thick or thin. The vector $\left[\theta, \bar{\theta}\right]$ controls the rotation and tilt of the brush which significantly influences the Chinese calligraphy being aesthetic or not. Figs. 3(a) and 3(b) show a schematic diagram of the rotation θ and tilt $\bar{\theta}$, respectively, of a writing brush. Figs. 3(c) and 3(d) show a schematic diagram which illustrates the brush writing a character according to the coordinate $\left[X, Y, Z, \theta, \bar{\theta}\right]$ in the simulation environment.

B. CALLIGRAPHY NET MODEL

The architecture of the hypothesis generation model for a robotic calligraphy-writing system is shown in Fig. 4. Firstly, we utilize the fast thinning algorithm [35] to extract data from the strokes of Chinese characters from a human writing sample. Next, we split the original image into several region of interest (ROI) images in accordance with the trajectory of the stroke. The number of ROI images is chosen to be the number of skeleton points. Every ROI during the writing process corresponding to [X, Y] is given by coordinates of a stroke. On the other hand, every coordinate $\left[Z, \theta, \bar{\theta}\right]$ corresponding to ROI image can be obtained by training the *Writer Net.* By using the coordinates [X, Y] and $\left[Z, \theta, \bar{\theta}\right]$, the writing results can be observed through the virtual robotic system. Then, we train *Estimator Net* by a simulative image written by the virtual robot to memorize and recognize the result from the virtual robotic system. Following that, we connect the *Writer Net* and *Estimator Net* as hypothesis generation net and lock *Estimator Net* to train *Writer Net* to minimize the loss between the original image and the image memorized by the *Estimator Net*. The learning process continues by alternating between *k1* iterations for optimizing *Writer Net* and *k2* iterations for optimizing *Estimator Net*. Keeping optimizing this training pattern until the simulative image becomes very close to the original image indicates that the robotic system has the ability to do better actions. Through the interaction between *Estimator Net* and *Writer Net*, they simultaneously progress to accomplish the hypothesis generation process. Therefore, we can obtain more accurate coordinates to write Chinese calligraphy. The loss functions of *Writer Net* and *Estimator Net* are respectively shown as:

$$loss_{\theta_E} = \sum_{i=0}^{l-1} \frac{1}{wh} \sum_{y=0}^{h-1} \sum_{x=0}^{w-1} (E(W(R(C(I)_{x,y}))) \\ -S(W(R(C(I)_{x,y}))))^2 \tag{3}$$

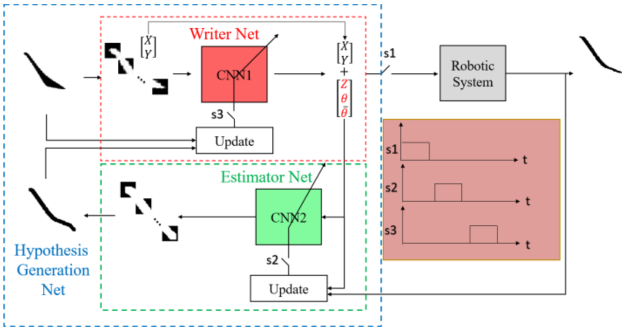

Writer Net
$\begin{bmatrix}X\\Y\end{bmatrix}$
CNN1
$\begin{bmatrix}X\\Y\end{bmatrix}$
+
$\begin{bmatrix}Z\\\theta\\\bar{\theta}\end{bmatrix}$
s1
Robotic System
s3
Update
Estimator Net
CNN2
s2
Update
Hypothesis Generation Net
s1
t
s2
t
s3
t

FIGURE 4. Scheme of hypothesis generation model for a robotic calligraphy-writing system.

$$loss_{\theta_W} = \sum_{i=0}^{l-1} \frac{1}{wh} \sum_{y=0}^{h-1} \sum_{x=0}^{w-1} \left(E(W(R(C(I)_{x,y}))) - R(C(I)_{x,y}) \right)^2 \quad (4)$$

$$R(C(I)_{x,y}) = I_{C(I)_{k-10} \sim C(I)_{k+10}}, \quad k = 1, 2, \ldots, 20 \quad (5)$$

where R is defined as ROI, and l is the length of the trajectory of the strokes. $C(\cdot)$ is defined as a function which sorts skeleton data according to the writing direction. The function $W(\cdot)$ is the proposed *Writer Net* that outputs a 3-dimension coordinates $[Z, \theta, \overline{\theta}]$ according to the ROI image. Function $S(\cdot)$ is the virtual robotic system [34] which outputs the writing result according to the coordinates $[X, Y, Z, \theta, \overline{\theta}]$. $E(\cdot)$ is the proposed *Estimator Net* that outputs an image according to the coordinates $[X, Y, Z, \theta, \overline{\theta}]$. We utilize mean square error (MSE) to measure the performance of the writing result. Fitting *Estimator Net* $E(\cdot)$ to the virtual robotic system $S(\cdot)$, we have $loss_{\theta_E}$ as the mean square error between $E(\cdot)$ and $S(\cdot)$. Note that *Writer Net* needs to write calligraphy results as close as possible to the human writing sample. Thus, we have $loss_{\theta_W}$ as the mean square error between $E(\cdot)$ and $R(\cdot)$. Thus, the *Estimator Net* and the *Writer Net* can be updated by minimizing the loss functions.

To help readers better understand the implementation, a pseudo-code is included below to illustrate the overall processes of the proposed hypothesis generation model.

Algorithm 1 Training Process of Hypothesis Generation Net

```
Input: Input of human writing sample, $R$
Output: A set of coordinate points, $d$
    Initialisation: Writer Net and Estimator Net with random weights
repeat
    Step 1:
    Input $R$ to Writer Net to produce a set of coordinates $d$
    Step 2:
    Virtual robot writes calligraphy according to $d$ in the simulation environment
    Step 3:
    for $i = 0$ to $kl$ do
        Update Estimator Net to approximate the network to the writing result by the virtual robot
    end for
    Step 4:
    for $i = 0$ to $k2$ do
        Update Writer Net so that output of Estimator Net approximates $R$
    end for
until stop button is pressed
```

C. WRITER NET AND ESTIMATOR NET

The detailed architecture of the proposed *Writer Net* is shown in Table 1, which consists of eleven layers with weights.

TABLE 1. Detailed architecture of the proposed writer net.

Layer	Type	Filter/Stride	Output Size
1	Image Input		20×20×1
2	Convolution and ReLU	3×3/1	20×20×128
3	Convolution and ReLU	3×3/1	20×20×128
4	Convolution and ReLU	3×3/1	20×20×128
5	Max Pooling	2×2/2	10×10×128
6	Convolution and ReLU	3×3/1	10×10×128
7	Convolution and ReLU	3×3/1	10×10×256
8	Convolution and ReLU	3×3/1	10×10×256
9	Max Pooling	2×2/2	5×5×256
10	Convolution and ReLU	3×3/1	5×5×512
11	Convolution and ReLU	3×3/1	5×5×512
12	Convolution and ReLU	3×3/1	5×5×512
13	Dropout (50%)		
14	LSTM		1×1×1024
15	RNN		1×1×3

Writing samples as input to the *Writer Net* are 20 × 20 grey scale images. All the convolutional layers have 3 × 3 filters and ReLu activation. Downsampling is utilized after the convolution layers by a max pooling layer with a stride of 2. When the previous layer is a max pooling layer, the number of the feature map is doubled to extract the feature from the higher dimensional data input. The dropout layer is set to fifty percent. LSTM and RNN in Table 1 are performed because our input writing samples are the ROI images of the stroke image. These ROI images are related to each other since the writing process is continuous. Figs. 5(a) and 5(b) show the right-falling stroke and its trajectory, respectively. Fig. 5(c) shows the ROI blocks along with the trajectory of the strokes, which demonstrates that every image has strong relation and causality with the adjacent images. Furthermore, every coordinate $[Z, \theta, \overline{\theta}]$ which maps to each image should be smooth and soft changing. The angles of the brush cannot change drastically if the states are close. Then, LSTM and RNN are utilized to suppress the variation of $[Z, \theta, \overline{\theta}]$.

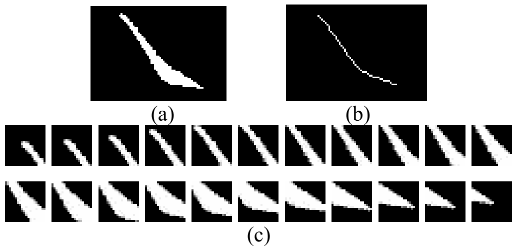


FIGURE 5. Input images of the Writer Net. (a) Right-falling stroke. (b) The trajectory of stroke. (c) ROI images of stroke.

The architecture of the proposed *Estimator Net* is shown in Table 2, which consists of fourteen layers with weights. The input vectors are 3-dimensional coordinates $[Z, \theta, \overline{\theta}]$. The convolutional layers also have 3 × 3 filters and ReLu activation. The transpose convolutional layers are utilized to upscale with a stride of 2. The dropout layer is also set to fifty percent. Then two fully-connected layers are utilized

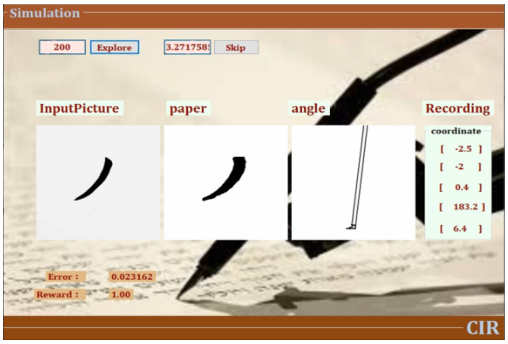
TABLE 2. Detailed architecture of the proposed estimator net.

Layer	Type	Filter/Stride	Output Size
1	Input		1×1×3
2	Fully Connected		1×1×512
3	LSTM		1×1×1024
4	Fully Connected		1×1×12800
5	Convolution and ReLU	3×3/1	5×5×512
6	Convolution and ReLU	3×3/1	5×5×512
7	Convolution and ReLU	3×3/2	10×10×512
8	Convolution and ReLU	3×3/1	10×10×256
9	Convolution and ReLU	3×3/1	10×10×256
10	Convolution and ReLU	3×3/2	20×20×256
11	Convolution and ReLU	3×3/1	20×20×512
12	Convolution and ReLU	3×3/1	20×20×512
13	Transpose Convolution and ReLU	3×3/2	40×40×512
14	Dropout (50%)		
15	Fully Connected		1×1×1024
16	Fully Connected		1×1×400

to extract features into final output of 400 nodes to obtain 20 × 20 images by reshaping the output.

V. SIMULATION RESULTS

We conduct our experiments on Intel Xeon CPU E3-1246 v6 of 3.70 GHz and NVIDIA GeForce GTX 1080 Ti with 32GB memory. To avoid spending too much time training the robotic arm to write Chinese calligraphy, we build a robotic simulation environment shown in Fig. 6 for our virtual robot to simulate the process of Chinese calligraphy writing [34], [36]. As shown in Fig. 6, the left picture box ''InputPicture'' shows the stroke of Chinese character written by a human. The middle picture box ''paper'' reveals the writing result of the virtual Chinese calligraphy-writing robot. The picture box ''angle'' shows the current state of the brush. The current 5-axis coordinates are also shown on the right side of Fig. 6. Except for InputPicture, all the other boxes update the status simultaneously when the simulation environment receives output of the *Writer Net*. The image of the Chinese calligraphy stroke captured by a webcam has a size of 200 × 200. We then convert the image into a grey-scale image as the input.



coordinate
[-2.5]
[-2]
[0.4]
[183.2]
[6.4]

Error : 0.023162
Reward : 1.00

FIGURE 6. The robotic simulation environment.

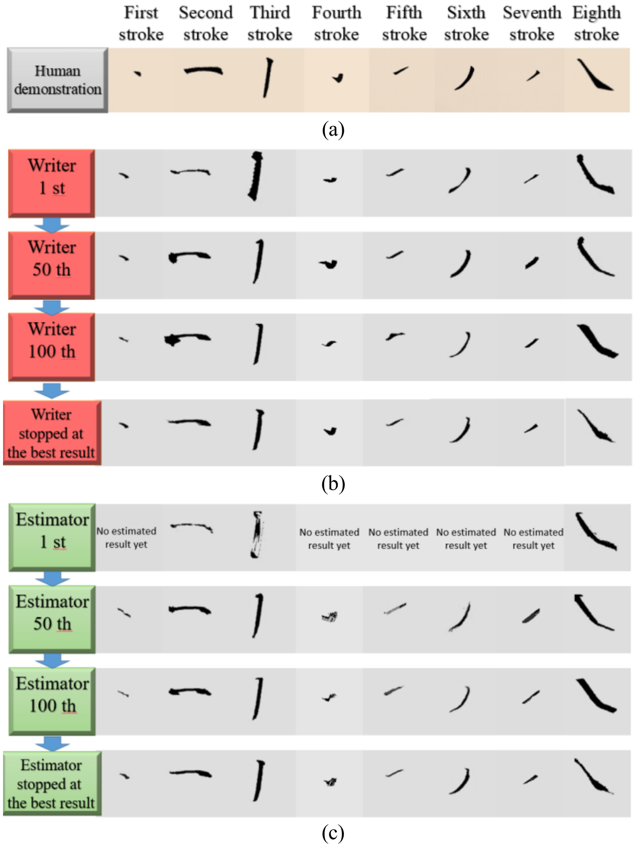


FIGURE 7. Diagram of the training process of writing eight strokes. (a) Ideal images written by human. (b) Writing results from Writer Net. (c) Images created from estimator.

The experiment is conducted under Python 3.6 that utilizes Tensorflow backend with Keras library and NVIDIA CUDA 9.0 library for parallel computation. Mean square error (MSE) is utilized to measure the performance of the hypothesis generation net. We utilize root mean square prop (RMSProp) [37] to be the optimizer. Fig. 7(a) shows the eight ideal Chinese strokes of Chinese character 'yong' (永). Figs. 7(b) and 7(c) show the training process of the eight strokes by the *Writer Net* and the *Estimator Net*, respectively. The images shown in Fig. 7(b) are drawn by the *Writer Net* which predicts the coordinates. Through the simulation system, the *Writer Net* emulates a similar image which the robotic arm could draw. Fig. 7(c) shows the images from the *Estimator Net* according to the coordinates provided from *Writer Net*. In the beginning, the *Estimator Net* generates images according to the coordinates far different from the *Writer Net*. Gradually, the results of the *Estimator Net* become more and more similar to the Chinese character written by the *Writer Net*. Therefore, the coordinates produced by the *Writer Net* become more and more similar to the ideal target, and this process simulates human's learning process. Firstly, a human generates a behavior based on the learning task as what the *Writer Net* does. Secondly, the *Writer Net* tries to imitate the human's writing. Next, the result is stored in the memory. Then, the *Writer Net* interacts with the

Estimator Net to analyze this behavior similar to what the hypothesis generation net does. For the next time, while doing the same action, the human will retrieve the experience from the previous time and do a better action.

Combining some strokes, we are able to form a complete Chinese character. The writing results of Chinese character 'yong' (永) are shown in Figs. 8(b)-8(e). Based on these strokes, four Chinese characters, i.e., yong (永, means permanence), tsun (寸, means inch), da (大, means big), and jiang (江, means river) shown in Figs. 9(a)-9(d) are the images of human writing samples. Figs. 9(e)-9(h) show the simulation results written by the virtual robot.

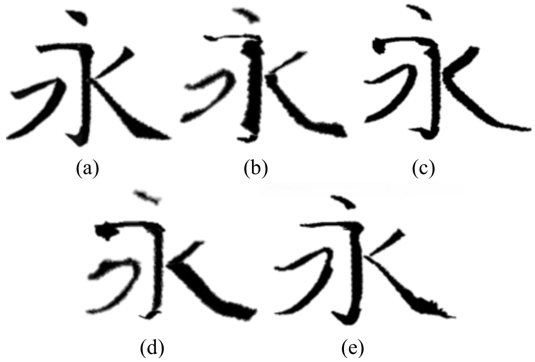


FIGURE 8. Writing of Chinese characters 'yong' (永). (a) Image of human writing sample of 'yong' (永). (b) First generation of writing 'yong.' (c) 50th generation of writing 'yong.' (d) 100th generation of writing 'yong.' (e) The best result of writing 'yong.'

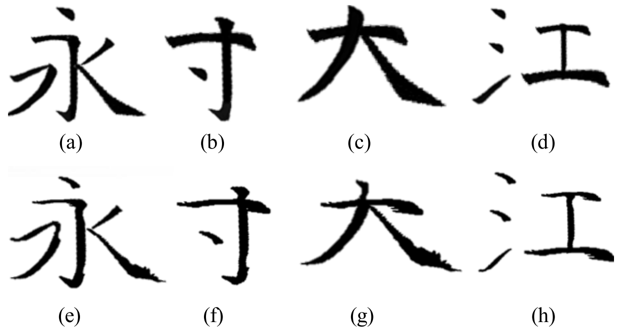


FIGURE 9. Four Chinese characters. (a) Image of human writing sample of 'yong' (永). (b) Image of human writing sample of 'tsun' (永). (c) Image of human writing sample of 'da' (大). (d) Image of human writing sample of 'jiang' (江). (e) Simulation result of 'yong' (永). (f) Simulation result of 'tsun' (永). (g) Simulation result of 'da' (大). (h) Simulation result of 'jiang' (江).

Remark 1: In the training process shown in Figs. 7 and 8, we found that there exist some variations in writing 'yong' at the 100th generation. Training process may overcorrect the writing because the *Estimator Net* probably forgets some of the past information if the model does not recall the past experiences for a long time. Then, the *Writer Net* makes new hypothesis according to the *Estimator Net* which forgets the past behaviors. That is, the *Writer Net* sometimes makes the same mistake because of forgotten memories. For example, the loss curve of the eighth stroke in the training process in Fig. 10 shows an undesired oscillating phenomenon. In the future, we plan to investigate memory systems to allow the hypothesis generation net with deeper impression about significant experiences, so that the performance of writing Chinese calligraphy can be improved.

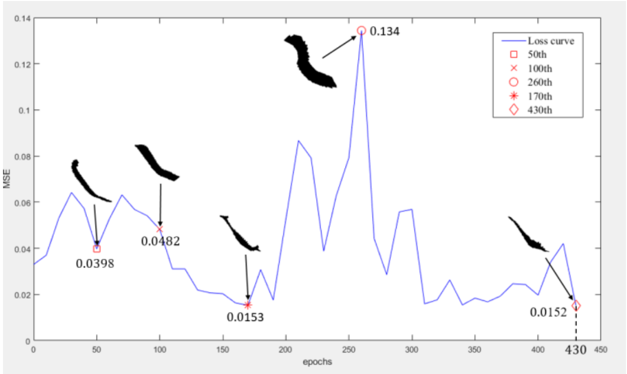


FIGURE 10. Loss curve of the eighth stroke.

VI. CONCLUSION

This paper presents a novel hypothesis generation model for a virtual robotic system to learn to write Chinese calligraphy through thinking and practicing according to a human writing sample. The proposed model has three main parts, i.e., the *Writer Net*, *Estimator Net*, and hypothesis generation net. These three models represent human's action, memory storage, and judgment. Through simulating human's psychological process, the hypothesis generation net has the ability to think and achieve some actions by itself. In our experiments, we design a *Writer Net* with 11 layers and an *Estimator Net* with 14 layers. The outputs of the hypothesis generation net are a series of coordinates for the virtual robot. Consequently, this cognitive framework is able to accomplish unsupervised image-to-action translation. Simulation results demonstrated in this paper confirmed the effectiveness and feasibility of the proposed hypothesis generation model to learn how to write Chinese calligraphy.

ACKNOWLEDGMENT

The authors would like to thank the National Center for High-Performance Computing for computer time and facilities to conduct this research.

REFERENCES

[1] I. Dasgupta, E. Schulz, and S. J. Gershman, "Where do hypotheses come from," *Cognit. Psychol.*, vol. 96, pp. 1–25, Aug. 2017.

[2] T. L. Griffiths and J. B. Tenenbaum, "Optimal predictions in everyday cognition," *Psychol. Sci.*, vol. 17, no. 9, pp. 767–773, Sep. 2006.

[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. 26th Annu. Conf. Neural Inf. Process. Syst. (NIPS)*, Dec. 2012, pp. 1–9.

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[5] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[6] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Representations*, May 2015, pp. 1–14.

[7] M.-J. Hsu, Y.-H. Chien, W.-Y. Wang, and C.-C. Hsu, "A convolutional fuzzy neural network architecture for object classification with small training database," *Int. J. Fuzzy Syst.*, vol. 22, no. 1, pp. 1–10, Feb. 2020.

[8] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[9] M.-Y. Liu, T. Breuel, and J. Kautz, "Unsupervised image-to-image translation networks," in *Proc. 26th Annu. Conf. Neural Inf. Process. Syst. (NIPS)*, Dec. 2017, pp. 1–9.

[10] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2223–2232.

[11] D. H. Ballard, "Modular learning in neural networks," in *Proc. 6th Nat. Conf. Artif. Intell. (AAAI)*, Jul. 1987, pp. 279–284.

[12] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.

[13] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," in *Proc. Int. Conf. Learn. Representations (ICLR)*, May 2014, pp. 1–14.

[14] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local," *J. Mach. Learn. Res.*, vol. 11, pp. 3371–3408, Dec. 2010.

[15] A. Ng, "Sparse autoencoder," *CS294A Lect. Notes*, vol. 72, pp. 1–19, Jan. 2011.

[16] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. 26th Annu. Conf. Neural Inf. Process. Syst. (NIPS)*, Dec. 2014, pp. 2672–2680.

[17] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," in *Proc. Int. Conf. Learn. Representations (ICLR)*, May 2016, pp. 1–16.

[18] J. Zhao, M. Mathieu, and Y. LeCun, "Energy-based generative adversarial network," in *Proc. Int. Conf. Learn. Representations (ICLR)*, Apr. 2017, pp. 1–17.

[19] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, *arXiv:1511.06434*. [Online]. Available: http://arxiv.org/abs/1511.06434

[20] P. Korshunov and M. Sébastien, "DeepFakes: A new threat to face recognition? Assessment and detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Dec. 2018, pp. 1–5.

[21] Y. Li and S. Lyu, "Exposing deepfake videos by detecting face warping artifacts," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Dec. 2018, pp. 46–52.

[22] K. Lo, K. Kwok, S. Wong, and Y. Yam, "Brush footprint acquisition and preliminary analysis for chinese calligraphy using a robot drawing platform," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2006, pp. 5183–5188.

[23] S.-K. Kim, J. Jo, Y. Oh, S.-R. Oh, S. Srinivasa, and M. Likhachev, "Robotic handwriting: Multi-contact manipulation based on reactional internal contact hypothesis," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Sep. 2014, pp. 877–884.

[24] Z. Ma and J. Su, "Aesthetics evaluation for robotic chinese calligraphy," *IEEE Trans. Cognit. Develop. Syst.*, vol. 9, no. 1, pp. 80–90, Mar. 2017.

[25] 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.

[26] 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.

[27] 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.

[28] M.-J. Hsu, Y.-H. Chien, Y.-T. Wu, W.-Y. Wang, and C.-C. Hsu, "Mechanical design and kinematics analysis of robotic calligraphy system using a delta-like robot manipulator," in *Proc. Int. Conf. Fuzzy Theory Appl. (iFUZZY)*, New Taipei City, Taiwan, Nov. 2019, pp. 105–107.

[29] M.-J. Hsu, Y.-H. Chien, W.-Y. Wang, and C.-C. Hsu, "Design and implementation of robotic calligraphy system," in *Proc. Int. Conf. Adv. Robot. Intell. Syst. (ARIS)*, Taipei, Taiwan, Sep. 2017.

[30] 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.

[31] X. Zhang, Y. Li, Z. Zhang, K. Konno, and S. Hu, "Intelligent chinese calligraphy beautification from handwritten characters for robotic writing," *Vis. Comput.*, vol. 35, nos. 6–8, pp. 1193–1205, Jun. 2019.

[32] M. C. Frank and N. D. Goodman, "Predicting pragmatic reasoning in language games," *Science*, vol. 336, no. 6084, p. 998, May 2012.

[33] F. H. Petzschner, S. Glasauer, and K. E. Stephan, "A Bayesian perspective on magnitude estimation," *Trends Cognit. Sci.*, vol. 19, no. 5, pp. 285–293, May 2015.

[34] Y.-H. Chien, M.-J. Hsu, L.-A. Yu, W.-Y. Wang, and C.-C. Hsu, "Robotic calligraphy system using delta-like robot manipulator and virtual brush model," *iRobotics*, vol. 2, no. 4, pp. 1–6, Dec. 2019.

[35] T. Y. Zhang and C. Y. Suen, "A fast parallel algorithm for thinning digital patterns," *Commun. ACM*, vol. 27, no. 3, pp. 236–239, Mar. 1984.

[36] C.-Y. Tzou, M.-J. Hsu, J.-Z. Jian, Y.-H. Chien, W.-Y. Wang, and C.-C. Hsu, "Mathematical analysis and practical applications of a serial-parallel robot with delta-like architecture," *Int. J. Eng. Res. Sci.*, vol. 2, no. 5, pp. 80–91, May 2016.

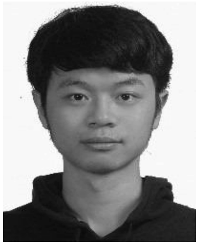
[37] T. Tieleman and G. Hinton, "Lecture 6.5-RmsProp: Divide the gradient by a running average of its recent magnitude," *COURSERA, Neural Netw. Mach. Learn.*, vol. 4, no. 2, pp. 26–31, Oct. 2012.

WEI-YEN WANG (Fellow, IEEE) received the Diploma degree in electrical engineering from the National Taipei Institute of Technology, in 1984, and the M.S. and Ph.D. degrees in electrical engineering from the National Taiwan University of Science and Technology, Taipei, Taiwan, in 1990 and 1994, respectively.

From 1990 to 2006, he worked concurrently as a Patent Screening Member of the National Intellectual Property Office, Ministry of Economic Affairs, Taiwan. Since 2003, he has been certified as a Patent Attorney, in Taiwan. In 1994, he was appointed as an Associate Professor with the Department of Electronic Engineering, St. John's and St. Mary's Institute of Technology, Taiwan. From 1998 to 2000, he worked with the Department of Business Mathematics, Soochow University, Taiwan. From 2000 to 2004, he was with the Department of Electronic Engineering, Fu Jen Catholic University, Taiwan, where he became a Full Professor, in 2004. In 2006, he was a Professor and the Director of the Computer Center, National Taipei University of Technology, Taiwan. From 2007 to 2014, he was a Professor with the Department of Applied Electronics Technology, National Taiwan Normal University, Taiwan. From 2011 to 2013, he was the Director of the Information Technology Center, National Taiwan Normal University, where he is currently a Professor with the Department of Electrical Engineering. His current research interests and publications are in the areas of fuzzy logic control, robust adaptive control, neural networks, computer-aided design, digital control, and CCD camera-based sensors. He has authored or coauthored over 200 refereed conference and journal papers in the above areas.

Dr. Wang is a Fellow of IET. He was a recipient of the Best Associate Editor Award of the IEEE Transactions on Cybernetics. He is also serving as an Associate Editor for the IEEE Transactions on Cybernetics and the *International Journal of Fuzzy Systems*.

MIN-JIE HSU was born in Taipei, Taiwan, in 1993. He received the B.S. degree in electrical engineering from National Taiwan Normal University, Taipei, in 2015, where he is currently pursuing the Ph.D. degree with the Department of Electrical Engineering. His research interests include artificial intelligence, fuzzy logic systems, neural networks, and reinforcement learning.



LI-AN YU received the B.S. and M.S. degrees in electrical engineering from National Taiwan Normal University, Taipei, Taiwan, in 2015 and 2017, respectively. He is currently a Postgraduate Research Assistant with the Computational Intelligence Laboratory, National Taiwan Normal University, where he is involved in artificial intelligence design. His research interests include neural networks and computer vision.

YI-HSING CHIEN was born in Taipei, Taiwan, in 1978. He received the M.S. degree in electrical engineering from Fu Jen Catholic University, Taipei, in 2007, and the Ph.D. degree in electrical engineering from the National Taipei University of Technology, Taipei, in 2012. He is currently a Postdoctoral Researcher with the Department of Electrical Engineering, National Taiwan Normal University, Taipei. His current research interests and publications are in the areas of fuzzy logic control, robust adaptive control, machine learning, and neural networks.

CHEN-CHIEN HSU (Senior Member, IEEE) was born in Hsinchu, Taiwan. He received the B.S. degree in electronic engineering from the National Taiwan University of Science and Technology, Taipei, Taiwan, in 1987, the M.S. degree in control engineering from National Chiao Tung University, Hsinchu, in 1989, and the Ph.D. degree from the School of Microelectronic Engineering, Griffith University, Brisbane, QLD, Australia, in 1997.

He was a Systems Engineer with IBM Corporation, Taipei, for three years, where he was responsible for information systems planning and application development, before commencing his Ph.D. studies. He joined the Department of Electronic Engineering, St. John's University, Taipei, as an Assistant Professor, in 1997, where he was appointed as an Associate Professor, in 2004. From 2006 to 2009, he was with the Department of Electrical Engineering, Tamkang University, Taipei. He is currently a Professor with the Department of Electrical Engineering, National Taiwan Normal University, Taipei. He is the author or coauthor of more than 200 refereed journal and conference papers. His current research interests include digital control systems, evolutionary computation, vision-based measuring systems, sensor applications, and mobile robot navigation. He is a Fellow of IET.