An Intelligent Error Correction Algorithm for Elderly Care Robots

Xin Zhang 1,2, Zhiquan Feng 1,2,* , Xiaohui Yang 1,2, Tao Xu 1,2, Xiaoyu Qiu 1,2 and Ya Hou 1,2

Abstract: With the development of deep learning, gesture recognition systems based on the neural network have become quite advanced, but the application effect in the elderly is not ideal. Due to the change of the palm shape of the elderly, the gesture recognition rate of most elderly people is only about 70%. Therefore, in this paper, an intelligent gesture error correction algorithm based on game rules is proposed on the basis of the AlexNet. Firstly, this paper studies the differences between the palms of the elderly and young people. It also analyzes the misread gesture by using the probability statistics method and establishes a misread-gesture database. Then, based on the misreading-gesture library, the maximum channel number of different gestures in the fifth layer is studied by using the similar curve algorithm and the Pearson algorithm. Finally, error correction is completed under the game rule. The experimental results show that the gesture recognition rate of the elderly can be improved to more than 90% by using the proposed intelligent error correction algorithm. The elderly-accompanying robot can understand people’s intentions more accurately, which is well received by users.

Keywords: human-computer interaction; gesture recognition; intelligent error correction; artificial intelligence; AlexNet network

1. Introduction

With the rapid development of vision technology and artificial intelligence, people have higher requirements for human-computer interaction. Compared with the traditional human-computer interaction, the interaction based on human biological characteristics is simpler and more flexible. Gesture, as the second language of human beings, has rich meaning. Gesture recognition has become a research hotspot in the field of human-computer interaction [1]. It is understood that there are three main categories of static gesture recognition technologies based on monocular vision [2]: the first is template matching technology, which matches the feature parameters of the gesture to be recognized with the template feature parameters stored in advance. It then completes the recognition task by measuring the similarity between them. The second is statistical analysis technology, a classification method based on probability and statistics theory that determines the classifier by a statistical sample feature vector. This technology requires people to extract specific feature vectors from the original data and classify these feature vectors instead of identifying the original data directly. The third type is a neural network technology, which has the ability of self-organization and self-learning, the characteristics of distribution, the ability of pattern generalization, and can effectively resist noise and deal with incomplete patterns. In the preparatory work, this paper used a Caffe [3,4] framework to train the AlexNet [5–7] network and carried out 180K iterative training by optimizing solver parameters. The convolution full connection process is shown in Figure 1. In the process of training, this
paper used the hold-out method to divide the data set into a training set and testing set. The size of the training set was 18,000 pictures (each type of gesture). The size of the testing set was 4000 pictures (each type of gesture) and the input size of each picture was $227 \times 227 \times 3$. A total of 7 types of gestures were trained. However, the recognition results are not satisfactory. The gesture recognition rate of many elderly people is below 80%. If this algorithm is applied to the elderly-accompanying robot, the robot will not be able to provide services for the elderly.

Therefore, this paper proposes a gesture error correction algorithm for the elderly-care robot. The purpose is to improve the recognition rate of the robot, improve the service quality of the elderly-care robot, and improve the elderly-care experience. Firstly, based on the behavior of the elderly, the recognition rate of hand gestures in natural interaction is calculated, a misread-gesture database is established, and the causes of recognition errors are analyzed. Then, based on the misread-gesture database, the reason for the low recognition rate of gestures based on the robot cognitive layer is explored. With the help of curve fitting and the Pearson correlation coefficient algorithm, we calculate the channel layer number with the biggest difference in the fifth layer feature map of convolution layer among different gesture categories in the gesture library. Finally, on this basis, a game rule correction algorithm is established.

2. Related Work

2.1. Gesture Recognition

In 1959, American scholar B. Shackel designed the computer console from the perspective of ergonomics [8], which was considered to be the first document of human-computer interface. At this time, humans had the idea of interacting with machines. The main ways of human-computer interaction were voice and gesture. Early gesture recognition was based on a data glove [9]. In 1983, Grime et al. first used gloves with node markers to recognize palm bone gestures and completed a simple cognition [10]. In the 1990s, due to the advantage of accurate positioning of peripheral devices, References [9,11] used data gloves to realize the recognition of 46 specific gestures. Later, a vision-based gesture recognition method appeared. In the beginning, it was a simple gesture recognition based on geometric features of gesture: References [9,12] used an information-entropy algorithm to segment the background image and successfully applied it in video data stream through a parallel computing algorithm. The recognition rate of the target image reached 95%. Ref [13] proposed to extract geometric features from depth images and histograms of orientation gradients from color images to realize gesture classification. Ref [14] proposed a method to detect the human body by combining Kinect multi-scale depth information and gradient information. Based on random images, the idea of positive and negative sample mutual limitation is adopted to identify each part of the human body. According to the distance of each part, the human posture vector was constructed to identify the skeleton. According to the body classification, the optimal classification hyperplane and kernel function are constructed, and the improved support vector machine was used for body classification.
With the advent of the era of artificial intelligence, deep learning became popular again. Because a convolutional neural network can directly process two-dimensional images, a gesture-recognition algorithm based on deep learning arose at the historic moment. A parallel convolutional neural network was designed to improve the recognition rate of static gestures in the complex background and changing lighting conditions [15], with the help of CNN (Convolutional Neural Networks) attitude prediction to assist the long-term and short-term memory (LSTM) network to estimate the probability of five kinds of look-down gestures observed by the camera in the car [16]. With the development of camera and sensor technology, the depth information of gestures was easier to capture, and the research of dynamic gesture recognition was becoming more and more popular. In this paper [17], a three-dimensional separable convolutional neural network is proposed to solve the problem of gradient discretization caused by separation operation by a jumping and hierarchical learning rate, which achieved high accuracy recognition of the dynamic gesture in a low complexity model. Single image input mode was easily affected, so multi-channel information fusion was developed, which makes up for the limitation of single-mode by fusing multiple modes. A biologically-inspired data fusion architecture is proposed, which realizes the fusion of visual data and somatosensory data of skin-like strain sensors made of single-walled carbon nanotubes in the feature layer, and the recognition accuracy can reach 100% [18].

2.2. Intelligent Error Correction

With the rapid development of the neural network, the recognition accuracy was still unsatisfactory in practical application, so an intelligent error correction algorithm was also applied to the gesture recognition algorithm based on deep learning. In 2006, the Dynamic Bayesian classifier was proposed to correct similar gesture recognition errors by combining motion-based and pose-based features [19]. Ref [20] proposed a real-time gesture recognition algorithm. This system includes 36 kinds of gestures, and mainly improves the accuracy of similar gestures by establishing a combination of features, including position, direction, and speed. In reference [21], HMM (Hidden Markov Model) was used to improve the recognition rate of similar gestures in the database by capturing the jumping trajectory information and quantifying the motion features in 3D space. In reference [22], the mechanism of gesture error recognition based on convolution neural network was explored, and an intelligent error correction algorithm based on a probability statistical model and convolution features was proposed. Through the intelligent detection and error correction of gesture recognition, intelligent interactive teaching was realized.

With the continuous development of human-computer interaction, robots no longer only accepted human ideas, but began to have their own “ideas”. At first, the robot could correct its steady-state motion error to achieve the purpose of self-correction. Reference [23] obtained the motion error information from the internal encoder of the robot, fed the obtained information back to the iterative learning controller (ILC) to calculate the compensation variable, added the compensation variable to the original position reference command of the robot, and finally corrected the steady-state motion error of the robot. Furthermore, the robot could judge the error by learning some physiological characteristics of humans when facing the error. In reference [24], the evaluative feedback obtained from human brain signals measured by scalp EEG was used to accelerate the repetitive learning of robot agents in a sparse reward environment. The robot decoded it into a noise error feedback signal and used the feedback signal to sense the impending error and successfully avoid it. Finally, the obstacle-avoidance function was tested. Later, robots were able to correct some human irregularities. In reference [25], the robot was used to capture and display humans’ wrong operation, and then the standardized operation was demonstrated. The goal of correcting humans’ irregular operation was achieved, and the efficiency of human-computer cooperation was improved. Literature [26,27] proposed a gesture-correction method using implicit feedback in reinforcement learning (RL). This method used the event-related activity of human electroencephalogram (EEG)-error-related
potential (ErrP) as the implicit feedback generated by RL. The NAO (a humanoid robot developed by SoftBankRobotics) robot would judge whether its recognition result was correct according to the feedback of the human EEG, and finally executed the corresponding instructions according to the feedback.

To conclude, first of all, most of the gesture recognition algorithms are applied to young people who can make standard gestures, which is not suitable for the object of this paper: the elderly. Therefore, the recognition results applied to the elderly are not satisfactory. Secondly, most of the error correction systems only make error reminders by judging whether there are errors with the set template. In this paper, based on the elderly-care robot, an intelligent error correction algorithm is studied. The mechanism of error correction is explored from the behavior level of the elderly and the robot cognition level, and the error correction mechanism is established. In the aspect of human behavior in the coming year, this paper counts the gestures with a low recognition rate, establishes a misread-gesture library and summarizes the reasons for the low recognition rate of some gestures combined with the analysis of the palm characteristics of the elderly in the nursing home. In the aspect of robot cognition, this paper explores the features based on the gesture-recognition library by mistake. The difference of the characteristic matrix of each channel in the map is used to correct the error recognition by game rules, and the error correction algorithm based on game rules is formed.

3. Intelligent Error Correction Algorithm

In human-computer interaction, naturalness is an important influencing factor of interaction comfort, but naturalness will affect the recognition rate of gestures, so gesture recognition in a daily life environment will inevitably make mistakes, especially for the special group of the elderly who need to be accompanied. Therefore, in order to improve the gesture recognition rate of the elderly escort robot and improve the escort effect of the elderly escort robot, this study analyzes gesture recognition from two levels and establishes an error correction mechanism. On the one hand, we should count the recognition probability of each gesture based on the behavior of the elderly and establish a misread-gesture database and probability matrix. Finally, it is concluded that the recognition error is caused by the non-standard gestures made by the elderly. On the other hand, we deeply discuss the differences between channels in the fifth convolution layer and use these differences to correct errors based on the convolution layer under the rules of the game, as shown in Figure 2.

![Figure 2. General framework of intelligent error correction algorithm.](image-url)
3.1. Reasons for Low Recognition Rate

The main goal of this section is to explore the reasons for the low recognition rate of gestures at the behavior level. The main method is to select the gesture with a low recognition rate through the probability statistics method and record the corresponding gesture error recognition and the probability of error recognition. Before that, to better understand the behavioral characteristics of gestures of the elderly, our research group paid a special visit to the local nursing home, studied the characteristics of gestures of the elderly, and listed several significant characteristics different from young people, as shown in Table 1.

| Gestural Characteristics of the Elderly | Gesture Characteristics of Young People |
|----------------------------------------|----------------------------------------|
| 1: Fingers bend naturally              | 1: Fingers are naturally straight       |
| 2: Constant shaking of the palm        | 2: The palm of a hand is steady         |
| 3: Fingers have no power               | 3: Fingers have high power              |

Based on the characteristics of the palms of the elderly summarized in Table 1, this study conducted experiments on several commonly used gestures in our life and created recognition probability statistics. Seven gestures of 30 elderly people were photographed and recorded, and 3000 photos were recorded for each gesture. Finally, we reached the probability of correct gesture recognition in practical use through probability statistics, as shown in Table 2.

| Signal Types (Number) | Fist (00) | Thumbs Out (01) | Index Finger Out (02) | Little Finger Out (03) | Five Fingers Merge (04) | Five Fingers Spread (05) | “Ok” Gesture (06) |
|-----------------------|-----------|-----------------|-----------------------|-----------------------|------------------------|-------------------------|-------------------|
| Elderly (%)           | 98.1      | 76.4            | 82.4                  | 75.4                  | 97.2                   | 72.7                    | 98.1              |
| Young (%)             | 98.4      | 90.3            | 91.2                  | 96.7                  | 97.3                   | 96.7                    | 98.9              |

It can be seen from Table 2 that some features of the palms of the elderly reduce the recognition rate of the gesture recognition algorithm, which makes the robot think that the elderly have made wrong gestures. According to Table 2, we need to select the gestures with a recognition rate of lower than 85%, select the gestures that recognize errors in the statistical process, establish the misread-gestures database, and then analyze the misread gestures. Because the network training is completed under supervised learning, even if there are recognition errors, the recognition result is in the range of 00 to 06. Therefore, we found that one gesture is often incorrectly recognized as another gesture, as shown in Table 3.

| Actual Number | 00 | 01 | 02 | 03 | 04 | 05 |
|---------------|----|----|----|----|----|----|
| Predicted Number | 00 | 016 | 0.108 | 0.003 | 0.781 | 0.219 |
| 01 | 0.002 | 0.012 | 0.074 | 0.872 | 0.0 |
| 02 | 0.008 | 0.048 | 0.114 | 0.037 | 0.962 |
| 03 | 0.016 | 0.076 | 0.105 | 0.038 | 0.781 |
| 04 | 0 | 0 | 0 | 0 | 0.219 |
| 05 | 0 | 0 | 0 | 0 | 0.962 |

Line 00 and column 01 represent the probability that the recognition gesture category number is 00, but the actual input gesture category number is 01. Due to the non-standard
gestures of the elderly, the robot mistakenly recognizes gesture 01 as 00. From Table 2, we can see that the recognition rate of 00 gestures is 98.1%. However, it still needs to be listed in Table 3, because 01, 02, and 03 gestures may be misrecognized as 00 gestures. This indicates that when the recognition result is 00, the recognition may be wrong because the actual input gesture number may be 01 or 02 or 03. The 04 gestures are listed in Table 3 for the same reason. Recognition is correct only when the input gesture number is consistent with the output gesture number. It can be further confirmed from Table 3 that gesture-recognition errors are caused by non-standard gestures made by the elderly. The degree of finger bending and fist-clenching is different from the standard, so in the recognition process, one gesture may be recognized as another gesture.

3.2. Error Correction Algorithm Based on Convolution Layer

The reasons for the low gesture recognition rate of individual elderly people have been found: the palm deformation of the elderly leads to the high similarity of some gestures. We produced an array of the overall features of these gestures that are similar. Next, we needed to verify our idea and shift our focus to the characteristic matrix of special channels, find the differences between similar gestures, and establish an error-correction algorithm.

According to Table 3, when the predicted gesture number is 01, the possible input gesture numbers are 00, 01, 02, and 03. Because the probability of 00 is less than 10%, we only extract the feature map of gesture numbers 01, 02, and 03 in the fifth layer of the volume base (the reason for choosing the fifth layer will be given in the experiment section), and record them as 01, 02, and 03. The size of this feature map is 256 × 12 × 12, which is converted into a 192 × 192 matrix. In the neural network shown in Figure 1, the convolution method of edge pixel filling [28] is adopted. Although the size of each channel of the fifth layer of convolutional layers is 13 × 13, the useful matrix size is only 12 × 12. The size of the fifth convolution layer mentioned below is 12 × 12 × 256. β, represents a 12 × 12 matrix, then the feature map of the fifth layer can be expressed as \[
\begin{bmatrix}
\beta_1 & \beta_2 & \cdots & \beta_{256}
\end{bmatrix}.
\]

The conversion mode is shown in Equation (1).

\[
\begin{bmatrix}
\beta_1 & \beta_2 & \cdots & \beta_{256}
\end{bmatrix} \Rightarrow \begin{bmatrix}
\beta_1 & \cdots & \beta_{16} \\
\vdots & \ddots & \vdots \\
\beta_{241} & \cdots & \beta_{256}
\end{bmatrix}.
\]

To determine whether there is an obvious gap between the data of the three arrays, this paper uses the spline function interpolation method [29] to fit them into three-dimensional surface graphs, respectively, as shown in Figure 3 (in this paper, only 01 and 02 gestures are selected, and 03 is similar, so we will not repeat them). The x-direction represents the first-dimension information of the feature, and the value range is 1 to 192. The y-direction represents the second-dimension information of the feature, with a value range of 1 to 192. The z-direction represents the eigenvalue.

![Figure 3](image-url)

**Figure 3.** 3d surface map of feature map of the fifth convolutional layer: (a) is 01 gesture, (b) is 02 gesture.
It can be seen from Figure 3 that the feature arrays of the two gestures in the fifth layer of convolutional layers are very similar, so one gesture is likely to be mistakenly recognized as another. However, fundamentally, they are still two different gestures. To find the most representative feature of each gesture, we study each channel of the layer 5 feature array deeply. Firstly, three feature maps representing 01, 02, and 03 are selected and divided into 256 12 × 12 matrices, respectively, and then the differences of channels with the same number are calculated. Although we need to calculate the difference degree between the data, we cannot only calculate the matrix as data, because these data have specific meaning in the specific position of the matrix. In order not to lose the location characteristics of data and to facilitate calculation, we convert 256 12 × 12 matrices into 256 1 × 144 one-dimensional arrays in row order, as shown in Equation (2).

\[
\begin{bmatrix}
\gamma_1 \\
\gamma_2 \\
\gamma_12
\end{bmatrix} \Rightarrow \begin{bmatrix}
\gamma_1 & \gamma_2 & \gamma_12
\end{bmatrix}.
\]  

(2)

where \(\gamma_1\) represents a 12-dimensional row vector. In order to increase the degree of difference between the two groups of data, the data is fitted with a curve, assuming that the point set is \(D = \{(0, \gamma_{1,i}), (1, \gamma_{2,i}), \ldots (143, \gamma_{12,i})\}\), finding approximate Equation (3).

\[y = \phi(x) = a_0 + a_1x + \cdots a_kx^k.\]  

(3)

Using the method of minimum deviation absolute Equation (4), we calculate the Pearson similarity [30] of the fitted curve. Equation (5) is the calculation equation of the Pearson similarity. Because there are 256 channels, we need to calculate 256 \(a_i\).

\[
\min_{\phi} \sum_{i=1}^{144} |\delta_i| = \sum_{i=1}^{144} |\phi(x_i) - y_i|,
\]  

(4)

\[a_i = \text{getSimilarity}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}.
\]  

(5)

To select the two groups of matrices with the biggest difference, it is necessary to select the one with the smallest Pearson correlation coefficient, and use this method to calculate the layer numbers of the matrices with the biggest difference between 01 and 02, 01 and 03, 02 and 03, which are counted as \(x_1\), \(x_2\) and \(x_3\), and stored in the database, as shown in Figure 4.

\[x_1, x_2, x_3\]  

Figure 4. Matrix layer number selection diagram.

The error-correction algorithm will use a kind of game rule, that is, 01, 02, 03 to play a game, and use the round-robin rule. The winner who wins two games is the final winner.
The advantage is that the data with the greatest difference can make the experimental results clear. The general process is shown in Figure 5.

![Flow chart of error correction algorithm based on game rules.](image)

**Figure 5.** Flow chart of error correction algorithm based on game rules.

The feature matrices with layer numbers $x_1$, $x_2$, and $x_3$ are extracted from the feature array of real-time recognition. First, we need to calculate the similarity between the $x_1$ layer characteristic matrix and the $x_1$ layer matrix of 01 and 02 and record the gesture numbers with greater similarity. If the correct identification number is 01 at this time, we choose the feature matrix layer with the biggest difference between 01 gesture and 02 gesture, so this game is equivalent to comparing the strengths of 01 with the weaknesses of 02, and the result is very obvious. Using the same method to calculate the similarity between the $x_2$ layer feature matrix and the $x_2$ layer matrix of 01 and 03, and the similarity between the $x_3$ layer matrix and the $x_3$ layer matrix of 02 and 03. Finally, the two winning gesture numbers are selected as the error correction result. Detailed algorithm steps are shown in Algorithm 1.

Algorithm 1 Error correction algorithm based on game rules

1. **Input**: Gesture recognition number $m^*$; Initialization $n = 0$.  
2. **Output**: Correct gesture type number $n$.  
3. **If** (Input (?) ⇒ Output $m^*$)  
4. **Search** (?) = $(m, p, q)$;  
5. **Search** Matrix layer number with the largest difference from $(m$ and $p) = x_1$, $(m$ and $q) = x_2$, $(p$ and $q) = x_3$;  
6. $a_1 \leftarrow \text{getSimilarity} (m^*, x_1, m.x_1) / *$ The similarity between layer $x_1$ channel of $m^*$ and layer $x_1$ channel of $m$ is calculated. */  
7. $a_2 \leftarrow \text{getSimilarity} (m^*, x_1, p.x_1)$;  
8. **If** ($a_1 > a_2$) $M++;$ **else** $P++;$ / * $M$ is the number of times $m$ wins, $P$ is the number of times $p$ wins, and if $a_1 > a_2$, $m$ wins. */  
9. $a_3 \leftarrow \text{getSimilarity} (m^*, x_2, m.x_2)$;  
10. $a_4 \leftarrow \text{getSimilarity} (m^*, x_2, q.x_2)$;  
11. **If** ($a_3 > a_4$) $M++;$ **else** $Q++;$  
12. $a_3 \leftarrow \text{getSimilarity} (m^*, x_3, m.x_3)$;  
13. $a_4 \leftarrow \text{getSimilarity} (m^*, x_3, q.x_3)$;  
14. **If** ($a_3 > a_4$) $P++;$ **else** $Q++;$  
15. **For** $(M, P, Q)$ is 2 / * Find out who won two games, because everyone can play two games at most. */  
16. **If** $(M == 2) n = m; / *$ Correct the recognition gesture category to $m$. */  
17. **If** $(P == 2) n = p;  
18. **If** $(Q == 2) n = q;  
19. **If** $(n == 0)$ Reenter gesture command; / * No best match template found. */  
20. **Else** output $n;  
21. **end**

4. Experimental Results and Analysis
4.1. Experimental Environment Setting

The host processor selected in the experiment is Intel (R) Core (TM) i7-10700k CPU, 3.2 GHz, the graphics card is NVIDIA 2080Ti, running under a 64-bit windows10 system, using Kinect real-time acquisition of RGB images of the hand. The commands of the elderly
were finally completed by the humanoid intelligent robot Pepper, developed by SoftBank. The development languages are C++, Python, and MATLAB, and the development platforms are VS2015, PyCharm, and MATLAB 2018.

4.2. Experimental Methods

To verify whether the error-correction algorithm can improve the gesture recognition rate, this paper designed a tea-drinking escort prototype system combined with the error-correction algorithm. It is hoped that the elderly can send instructions to the robot through gestures, and the robot will perform corresponding operations after recognition. Finally, the robot will help the elderly finish the activity of drinking tea. A daily family living environment was simulated in the laboratory, which included a round tea table, chair with back, teacup, tea, kettle, and other tea service necessities. The detailed scene diagram is shown in Figure 6.

![Figure 6. Schematic diagram of the experiment scene.](image)

In the experimental operation method, this study also added other channel information, such as speech recognition and target detection. The purpose was to improve the real experience of the elderly experiment and improve the authenticity and reliability of the experimental data. The main operation was still that the elderly gave hand instructions to the robot through hand gestures. For example, when the elderly made gesture 01, they could command the robot to turn 90° to the left. In addition, when the elderly made gesture 07, they could command the robot to take the cup. In this process, we added the function of target detection. Only when the robot detected the cup would it execute the command to take the cup. The experiment was composed of 20 old people. They needed to make seven gestures to experience the tea service. Each gesture was tested 200 times. Some gesture instructions are shown in Table 4.

### Table 4. Instructions of experimental operation.

| Instructions of experimental operation. |
|-----------------------------------------|
| **Experimenter:** Peper: Make gesture 01 (left turn command). |
| **Experimenter:** Make gesture 03 (right turn command). |
| **Peper:** The robot turns left 90° without obstacles on the left. |
| **Experimenter:** Make gesture 05 (forward command). |
| **Peper:** The robot turns 90° to the right without obstacles on the right side. |
| **Experimenter:** Make gesture 06 (take the cup command). |
| **Peper:** If there is an obstacle ahead, the robot will stop automatically to ensure absolute safety. |
| **Peper:** The robot determines that the cup is in front of the robot through target detection, and then performs the grab operation. |

In this experiment, the whole tea-drinking service scene is divided into several steps. The start switch of each step is the corresponding gesture. That is to say, only when the
robot’s gesture recognition is correct can the corresponding service be started. On the contrary, if a certain step is wrong, the whole service will not be completed:

- The experimenters should interact with natural gestures as in daily life, and the speed should not be too fast;
- The experimenters only make gestures related to tea drinking service to avoid affecting the experiment time;
- After each gesture instruction, the experimenter should make the second gesture instruction after the robot finishes;
- The experimenters conducted ten tea service experiments based on a behavior-mechanism-error-correction algorithm and tea service experiments based on a robot cognitive-error-correction algorithm.

The purpose of the experiment is to verify whether the algorithm achieves the expected goal. Therefore, it is necessary to show the authenticity of the experimental process through on-site photos, and to illustrate the effect of the algorithm through the number of errors found and the number of errors corrected. In this study, two counters were added to the experimental code to record the number of errors found and the number of errors corrected, respectively. Readers may misunderstand “discovery error,” and think that as long as the recognition result number is in the error gesture library number, the error is found, but it is not. The robot begins to suspect that the gesture of the elderly is wrong, but is not sure, so it does not make much sense to set the counter here. Only when the robot calculates the matching degree between the real-time gesture recognition eigenvalue and the feature array in the matrix library can it find the error and set the first counter at this position. Now that the scene has been designed, the number to be transmitted to the robot is also determined, so it is easy to calculate the number of errors to be corrected. Therefore, it is necessary to compare the identification number transmitted to the robot with the set number to calculate the number of errors to be corrected. This is the second counter.

4.3. Experiment Results
4.3.1. Demonstration of Experimental Results

To show the feasibility of this algorithm, after the end of the experiment, several representative pictures are selected to show the experimental environment and experimenters.

In Figure 7, the old man in picture (a) is expressing the command of drinking tea to the robot through gesture, and the robot in picture (b) is handing the cup to the old man.

![Figure 7. Experimental scene. (a): middle-aged and old people are making gestures; (b): the robot is executing the corresponding command.](image-url)
4.3.2. Algorithm Feasibility Verification

To explore whether the error-recognition gesture-correction algorithm in this paper achieves the effect of finding and correcting errors, after the experiment, taking 01 gesture as an example, the gray images of the elderly wrong gesture and the gesture that was mistakenly recognized during the experiment are selected, as shown in Figure 8.

![Figure 8. Grayscale diagram of gesture of the elderly.](image)

Through experimental analysis, gesture 01 can be recognized as gesture 00, 01, 02, and 03, and the recognition probability is 17.6%, 76.4%, 5%, and 1%, respectively. It can be found that the probability of recognizing 01 is still the largest. If we look at the gray image of gesture, 00 and 01 have high similarities. 01 is similar to 02 and 03, but the similarity is low. This paper aims to find out these easily misunderstood gestures and use the intelligent error-correction algorithm to correct them. This process includes two tasks: one is to find errors and the other is to correct errors, so the experiment finally summarizes the number of errors found and corrected by the two algorithms into a histogram, as shown in Figure 9.

![Figure 9. Bar chart of experimental effect.](image)

As can be seen from Figure 8, in the process of 200 01 gesture recognitions, the correct recognition is 152 times and the wrong recognition is 48 times. The number of errors corrected is 34.

4.4. Contrast Experiment

To better reflect the advantages and persuasiveness of the algorithm in this paper, comparative experiments of five error-correction algorithms were conducted in this section, which is the identification algorithm of the original AlexNet network without adding any
error-correction algorithm, including the error-correction algorithm based on game rules, the error-correction algorithm based on the Hausdorff algorithm [31], the error-correction algorithm based on the Fréchet algorithm [32], and the error-correction algorithm based on Gaussian distribution [22]. In the same experimental environment as the previous paper, the comparative experiments of these five algorithms are carried out, and then the advantages of this algorithm are analyzed by comparing the gesture-recognition rate of the five algorithms.

Here we first introduce the reasons for choosing the fifth convolution layer. Firstly, two samples were randomly selected from a certain gesture type, and the eigenvalue data from the first layer to the last fully connected layer were extracted, respectively. Then the similarity between each layer was calculated repeatedly, and the data volume of each layer was also recorded. The results are shown in Figure 10. It can be seen that in the case of maintaining a high matching rate, the amount of data in the fifth layer of convolutional layers is the smallest, so this paper selects the fifth layer of convolutional layers.

Both the error correction algorithm based on the Hausdorff and the error correction algorithm based on the Fréchet use the idea of matching the optimal value. Firstly, a feature template is established for each kind of gesture with low a recognition rate, and then the similarity between the recognition gesture and the template is calculated by the Hausdorff algorithm and the Fréchet algorithm, respectively. The highest similarity is the final recognition result. The error-correction algorithm based on Gaussian distribution adopts the method of regional classification. Firstly, the algorithm judges whether to correct the error according to the gesture recognition number. If the gesture recognition number belongs to the error-prone gesture number, the three-dimensional surface peaks of channel 6 and channel 58 of the volume base layer of the fifth layer will be calculated, and finally, the recognition results after error correction will be output through the membership range of the peaks. Because the probability problem requires a large amount of data, the comparative experiment uses 3000 pictures mentioned in Section 3.1 as the test set. Finally, the recognition rate of each error-correction algorithm is counted, and the results are shown in Figure 11.
Figure 11. Comparison of recognition rates of different algorithms.

As can be seen from Figure 11, the error-correction algorithm based on the Hausdorff and the Fréchet not only does not improve the recognition rate but reduces the recognition rate. The recognition rate of the error-correction algorithm based on game rules and error-correction algorithm based on Gaussian distribution almost reach more than 90%, reaching the expected standard, but the error-correction algorithm based on game rules is better. In addition, the application range of the error-correction algorithm based on game rules is wider than that based on Gaussian distribution.

4.5. User Experience and Cognitive Load

To test whether the elderly escort robot based on the error correction algorithm meets the design requirements, 20 elderly people mentioned above were invited to this part. The evaluation of the elderly on the convenience, helpfulness, experience, and intelligence of the escort system was investigated. The evaluation adopts the 5-point system; 0–1 indicates that the evaluation is very low, 1–2 indicates that the evaluation is relatively low, 2–3 indicates that the evaluation is general, 3–4 indicates that the evaluation is relatively high, and 4–5 indicates that the evaluation is very high. The evaluation results are shown in Figure 12.

Figure 12. User experience evaluation.
As can be seen from Figure 12, in the aspect of helpfulness evaluation, the probability of the robot providing correct help is greatly improved because the error-correction algorithm based on game rules improves the gesture recognition rate. In terms of experience evaluation, the error rate of the CNN method is high, which greatly reduces the fluency of the whole process and the experience of the elderly. Intelligent evaluation is for the same reason, strong error-correction ability, intelligent evaluation is naturally high.

At the same time, 20 elderly people who participated in the experiment were invited to conduct a NASA evaluation. User evaluation indicators are divided into mental demand (MD), physical demand (PD), performance (P), effort (E), and frustration (f). NASA’s evaluation index adopts a 5-point system; 0–1 indicates that the cognitive burden is small, 1–2 indicates that the cognitive burden is relatively small, 2–3 indicates that the cognitive burden in general, 3–4 indicates that the cognitive burden is relatively large, and 4–5 indicates that the cognitive burden is very large. As can be seen from Figure 13, the game rule-based error-correction algorithm proposed in this paper has a low user cognitive load and smoother accompanying process. At the same time, it also brings novelty to the elderly and high user evaluation.

Figure 13. NASA user evaluation.

5. Conclusions

Aiming at the problems that the robot cannot correctly recognize the gestures of the elderly and the elderly cannot get good service in the accompanying process, an intelligent error-correction algorithm applied to the elderly-accompanying robot is proposed in this paper. The intelligent error-correction algorithm is based on the misread-gestures database and uses game rules to correct the misrecognized gesture, and achieves the goal of improving the gesture recognition rate.

This paper also establishes a prototype system of tea-drinking escort. The experimental results and the feedback from the experimenters show that the intelligent error correction algorithm greatly improves the recognition rate of gestures of the elderly, improves the smoothness of the accompanying process, and makes the elderly more dependent on the accompanying robot.

At the same time, the intelligent error-correction algorithm also needs to be improved. This algorithm only studies static gestures, hoping to add dynamic gestures to increase the diversity of interaction in the future.

Author Contributions: Conceptualization, X.Z. and X.Q.; Data curation, X.Z.; Formal analysis, X.Z. and Y.H.; Funding acquisition, Z.F.; Investigation, X.Z., X.Y. and X.Q.; Methodology, X.Z.; Project administration, Z.F.; Resources, X.Z.; Software, X.Z.; Writing—original draft, X.Z.; Writing—review and editing, X.Y. and T.X. All authors have read and agreed to the published version of the manuscript.
Funding: This paper is supported by the Independent Innovation Team Project of Jinan City (No. 2019GXRC013), and the Shandong Provincial Natural Science Foundation (No. ZR2020LZH004).

Institutional Review Board Statement: Not applicable for this study.

Informed Consent Statement: Not applicable for this study.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author. The data are not publicly available due to the data copyright belongs to the laboratory.

Conflicts of Interest: This paper has produced a patent, but it is still in the trial stage (August 2021). The authors declare no conflict of interest.

References
1. Qiannan, Z. Research on Gesture Recognition Technology Based on Computer Vision; Dalian University of Technology: Dalian, China, 2020.
2. Yi, J.; Cheng, J.; Ku, X. Review of visual gesture recognition. Comput. Sci. 2016, 43, 103–108.
3. Du, Y.; Yang, R.; Chen, Z.; Wang, L.; Weng, X.; Liu, X. A deep learning network-assisted bladder tumour recognition under cystoscopy based on Caffe deep learning framework and EasyDL platform. Int. J. Med. Robot. Comput. Assist. Surg. 2020, 17, e2169. [CrossRef]
4. Wu, H.; Ding, X.; Li, Q.; Du, L.; Zou, F. Classification of women’s trousers silhouette using convolution neural network CaffeNet model. J. Text. Res. 2019, 40, 117–121.
5. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. Commun. ACM 2017, 60, 84–90. [CrossRef]
6. Dutta, A.; Islam, M.M. Detection of Epileptic Seizures from Wavelet Scalogram of EEG Signal Using Transfer Learning with AlexNet Convolutional Neural Network. In Proceedings of the 2020 23rd International Conference on Computer and Information Technology (ICCIT), Dhaka, Bangladesh, 19–21 December 2020.
7. Gaur, A.; Yadav, D.P.; Pant, G. Morphology-based identification and classification of Pediastrum through AlexNet Convolution Neural Network. IOP Conf. Ser. Mater. Sci. Eng. 2021, 1116, 012197. [CrossRef]
8. Mitra, S.; Acharya, T. Gesture Recognition: A Survey. IEEE Trans. Syst. Man Cybern. Part C 2007, 37, 311–324. [CrossRef]
9. Fang, W.; Ding, Y.; Zhang, F.; Sheng, J. Gesture Recognition Based on CNN and DCGAN for Calculation and Text Output. IEEE Access 2019, 7, 28230–28237. [CrossRef]
10. Grimes, G.J. Digital Data Entry Glove Interface Device. U.S. Patent US4414537, 8 November 1983.
11. Takahashi, T.; Kishino, F. Hand gesture coding based on experiments using a hand gesture interface device. ACM SIGCHI Bull. 1991, 23, 67–74. [CrossRef]
12. Lee, J.; Lee, Y.; Lee, E.; Hong, S. Hand region extraction and gesture recognition from video stream with complex background through entropy analysis. In Proceedings of the 26th Annual International Conference of the IEEE, San Francisco, CA, USA, 1–5 September 2004; pp. 1513–1516.
13. Marouane, H.; Abdelhak, M. Hand Gesture Recognition Using Kinect’s Geometric and HOG Features. In Proceedings of the 2nd international conference on Big Data, Cloud and Applications, Tetouan, Morocco, 29–30 March 2017; p. 48.
14. Hu, D.; Wang, C.; Nie, F.; Li, X. Dense Multimodal Fusion for Hierarchically Joint Representation. In Proceedings of the ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 3941–3945.
15. Gao, Q.; Liu, J.; Ju, Z.; Li, Y.; Zhang, T.; Zhang, L. Static Hand Gesture Recognition with Parallel CNNs for Space Human-Robot Interaction International Conference on Intelligent Robotics & Applications; Springer: Cham, Switzerland, 2017.
16. Tewari, A.; Taetz, B.; Granddidier, F.; Stricker, D. A Probabilistic Combination of CNN and RNN Estimates for Hand Gesture Based Interaction in Car. In Proceedings of the IEEE International Symposium on Mixed & Augmented Reality IEEE International Symposium on Mixed & Augmented Reality IEEE Nantes, France, 9–13 October 2017.
17. Wu, Y.; Zheng, B.; Zhao, Y. Dynamic Gesture Recognition Based on LSTM-CNN. In Proceedings of the 2018 Chinese Automation Congress (CAC), Xi’an, China, 30 November–2 December 2018; pp. 2446–2450.
18. Wang, M.; Yan, Z.; Wang, T.; Cai, P.; Gao, S.; Zeng, Y.; Yan, C.; Wang, H.; Pan, L.; Yu, J.; et al. Gesture recognition using a bioinspired learning architecture that integrates visual data with somatosensory data from stretchable sensors. Nat. Electron. 2020, 3, 563–570. [CrossRef]
19. Aviles-Arriaga, H.H.; Sucar, L.E.; Mendoza, C.E. Visual Recognition of Similar Gestures. In Proceedings of the International Conference on Pattern Recognition IEEE, Hong Kong, China, 20–24 August 2006.
20. Elmezain, M.; Al-Hamadi, A.; Michaelis, B. Hand gesture recognition based on combined features extraction. J. World Acad. Sci. Eng. Technol. 2009, 60, 395.
21. Ding, Z.; Chen, Y.; Chen, Y.L.; Wu, X. Similar hand gesture recognition by automatically extracting distinctive features. Int. J. Control. Autom. Syst. 2017, 15, 1770–1778. [CrossRef]
22. Sun, K. Research on Error Recognition Gesture Detection and Error Correction Algorithm Based on Convolutional Neural Network; University of Jinan: Jinan, China, 2019. (In Chinese)

23. Stanceski, S.; Zhang, J. A Simple and Effective Learning Approach to Motion Error Correction of an Industrial Robot. In Proceedings of the 2019 International Conference on Advanced Mechatronic Systems (ICAMechS), Kusatsu, Japan, 26–28 August 2019; pp. 120–125.

24. Akinola, I.; Wang, Z.; Shi, J.; He, X.; Lapborisuth, P.; Xu, J.; Watkins-Valls, D.; Sajda, P.; Allen, P. Accelerated Robot Learning via Human Brain Signals. In Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 31 May–31 August 2020.

25. Bested, S.R.; de Grosbois, J.; Crainic, V.A.; Tremblay, L. The influence of robotic guidance on error detection and correction mechanisms. Hum. Mov. Sci. 2019, 66, 124–132. [CrossRef] [PubMed]

26. Kim, S.K.; Kirchner, E.A.; Stefes, A.; Kirchner, F. Intrinsic interactive reinforcement learning—Using error-related potentials for real world human-robot interaction. Sci. Rep. Nat. 2017, 7, 17562. [CrossRef] [PubMed]

27. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A.A.; Veness, J.; Bellemare, M.G.; Graves, A.; Riedmiller, M.; Fidjeland, A.K.; Ostrovski, G.; et al. Human-level control through deep reinforcement learning. Nature 2015, 518, 529–533. [CrossRef] [PubMed]

28. Brilliant. Detailed Explanation of Opencv Convolution Principle, Edge Filling Mode and Convolution Operation. Blog Garden. 2020. Available online: https://www.cnblogs.com/wojianxin/p/12597328.html (accessed on 30 March 2020).

29. Abidi, A.; Nouira, I.; Assali, I.; Saafi, M.A.; Bedoui, M.H. Hybrid Multi-Channel EEG Filtering Method for Ocular and Muscular Artifact Removal Based on the 3D Spline Interpolation Technique. Comput. J. 2021, bxa175. [CrossRef]

30. Benesty, J.; Chen, J.; Huang, Y.; Cohen, I. Pearson Correlation Coefficient. Noise Reduction in Speech Processing; Springer: Berlin/Heidelberg, Germany, 2009; pp. 1–4.

31. Kraft, D. Computing the Hausdorff Distance of Two Sets from Their Distance Functions. Int. J. Comput. Geom. Appl. 2020, 30, 19–49. [CrossRef]

32. Biran, A.; Jeremic, A. Segmented ECG Bio Identification using Fréchet Mean Distance and Feature Matrices of Fiducial QRS Features. In Proceedings of the 14th International Conference on Bio-inspired Systems and Signal Processing, Vienna, Austria, 11–13 February 2021.