IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 26 February 2024, accepted 9 March 2024, date of publication 25 March 2024, date of current version 2 April 2024.

*Digital Object Identifier 10.1109/ACCESS.2024.3380839*

RESEARCH ARTICLE

# Improved VGG Algorithm for Visual Prosthesis Image Recognition

**ZHIGUANG LI[1], BAITAO LI[2], SURNG GAHB JAHNG[1], AND CHANGYONG JUNG[1]**
[1]School of Advanced Imaging Science, Multimedia and Film, Chung-Ang University, Seoul 06974, South Korea
[2]School of Journalism and Communication, Henan University, Kaifeng 475000, China

Corresponding author: Zhiguang Li (makefilmcau@126.com)

**ABSTRACT** With the development of bio-informatics, visual prostheses have become an effective method for low vision people to restore visual function. To meet the visual needs of visual implant recipients, this study explored the problem of in vitro image processor image recognition and classification. It selected the convolutional neural network framework VGG as the technical core, introduced the fruit fly optimization algorithm for optimizing the VGG recognition model, and constructed a visual prosthesis image recognition model on the ground of improved VGG. The experiment demonstrated that the improved fruit fly search algorithm had an average absolute error and root mean square error values lower than 0.4, which was superior to other intelligent optimization algorithms. The performance of the improved image recognition model has been significantly improved, with a maximum AUC value of 0.942, a recognition accuracy range of 68.29%-97.23%, and a stable fitness curve of around 97.00. The maximum F1 value of the image recognition model designed in the study reached 91.47%, and the loss function curve converged to the minimum value. In the application of visual prostheses, the recognition accuracy and R-squared performance of this model were both the best. Compared with natural human vision, the contrast and functional visual effects matched well, the processing speed was faster, and the delay time did not affect the actual application of visual prostheses, achieving high user satisfaction. This study can enrich and improve the theoretical foundation of visual image analysis and visual prosthetics, and help visually impaired groups improve their lives and quality of life.

**INDEX TERMS** Visual prostheses, convolutional neural network, VGG, image recognition, Drosophila optimization algorithm.

## I. INTRODUCTION

Vision is the main sensory pathway for humans to obtain external information, and the presence of visual impairments can seriously affect individuals' normal life and work. For human ocular visual pathologies, traditional visual restoration techniques include contact lenses, laser therapy, corneal transplantation, retinal surgery and retinal lasers. For example, Viberg et al. performed corneal limbal stem cell transplantation, Van der Wekken-Pas et al. also conducted a study on the treatment of Candida keratitis and endophthalmitis after corneal transplantation, and Obata et al. performed anti-vascular endothelial growth factor therapy and laser therapy for Retinopathy of Prematurity. However, physical assistance as well as pharmacologic surgical treatment can only slow down the progression of ocular lesions and do not accomplish complete repair of the visual senses [1], [2], [3].

Visual Prosthesis is an implantable medical electronic device that is currently the most promising technical means to assist patients with visual impairment or severe low vision in restoring visual function [4]. Visual prostheses perceive and recognize light and images through the transmission and stimulation of electrical signals, and use tiny electrodes to convert the light and images captured by external devices into electrical signals. They are then transmitted to the patient's optic nerve to activate the visual cortex [5]. Meanwhile, the internal device of the visual prosthesis is implanted in the patient's eye or visual system to directly stimulate the retina or optic nerve, thereby achieving the patient's perception of light and image [6], [7]. However, the implementation and clinical practice of visual prostheses still face some difficulties and limitations. People's understanding of the brain's

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

visual processing process is limited, and the surgical risks of artificial visual prostheses implanted in the eye or brain, as well as the stability and safety issues of neural interfaces, limit the application of visual prostheses. Due to the limitation of the number of implanted chips and electrodes, the resolution and clarity of visual prosthetic images cannot reach natural visual levels [8], [9]. Therefore, to meet the visual needs of visual prosthesis users, an in vitro image processor is usually used to extract image features using computer algorithms to complete image recognition and assist in the work of visual prostheses. Recently, Convolutional Neural Networks (CNNs) have been extensively utilized in many terms like detection, recognition, and classification. However, with the complexity of their structure, the improvement of computational efficiency and accuracy of network models remains an urgent challenge in image processing [10], [11]. On the ground of this, this study takes the classic CNN architecture Visual Geometry Group (VGG) designed by the VGG as the technical core, and introduces the Fruit Fly Optimization Algorithm (FOA) for meeting the application requirements of visual prostheses. Firstly, the FOA is optimized and improved using Gaussian backbone and abandonment mechanism. Then, on the ground of the improved FOA, the image classification model VGG was improved, and an improved VGG recognition algorithm for visual prosthetic image recognition technology was developed. The contribution of the research mainly lies in optimizing the FOA while improving the VGG network. This enriches the theoretical study of FOA and VGG, and is conducive to the visual prosthesis for better in vitro image processing tasks. At the same time, it promotes the progress and development of image processing technology and expands the potential of image recognition applications.

The research consists of four parts. The first is a review of the current research status of visual prosthesis technology and image recognition processing both domestically and internationally. The second part proposes an image recognition and classification model on the ground of FOA and VGG. The third part conducts testing experiments and analysis on the performance of the recognition algorithm. The fourth part summarizes and summarizes the experimental results. This study is expected to improve the technical difficulties of visual prostheses and promote their clinical application.

## II. RELATED WORKS

The research on visual prostheses involved multiple fields such as computer vision, machine learning, and pattern recognition, which helped to promote innovation and progress in medical research and eye disease treatment. The related research has received widespread attention worldwide, and numerous researchers have conducted extensive research on visual prostheses. Visual prosthetic devices are commonly used to treat low vision. To determine patient acceptance and the risk-reward balance of visual prostheses, Karadima et al. conducted a study on the attitudes of low vision people in Athens, Greece. Through questionnaires, guided discussions, and quantitative data reporting, the study found that patients still perceive the risks of visual prostheses to outweigh the benefits. The retinal approach was found to have the least negative perception, while the cortical approach was perceived the most negatively [5]. The use of visual prostheses currently had limitations in restoring vision. Further research and pre-clinical work was necessary to improve the devices and stimulation strategies used to induce neural activity for visual perception. Spencer et al. devised a methodology for evaluating candidate strategies and devices. The experimental results showed that when there is a large amount of overlap in the neural activity generated by different electrodes, the methodology devised in the study can provide accurate assessment results. This can help to select faster and more accurate improved devices and neuro stimulation strategies [12]. Wang et al. designed an energy-efficient dynamic scene processing framework for implantable visual prosthesis. The framework used a combination of spike representation coding technique and bio-inspired spike recurrent neural network to achieve smart processing and low power consumption. The experimental results showed that the Pearson correlation coefficient of this method was 0.93, which was better than the state-of-the-art retinal prosthesis processing framework. Compared with convolutional recurrent neural network processing, the power consumption was reduced by a factor of 8, alleviating the accuracy and power consumption problems of retinal prosthesis [13]. Facial perception and cognition are key functions of low vision retinal prostheses, and the limitations of electrode arrays result in poor image pixels provided by visual prostheses, seriously affecting the relevant expression. For enhancing the comprehensibility of limited pixel facial images with visual prostheses, Xia et al. constructed a facial semantic information conversion model on the ground of deep neural networks to convert real faces into pixel faces, solving the semantic loss under limited pixels. Then the researchers built a pixel face database designed for visual prostheses. The experiment showcased that the model has better recognizability in features and expressions than existing technologies [14]. The optic nerve was the first choice for the placement of visual prostheses, and the effectiveness of visual prostheses was closely related to the combination of stimulus parameters. It is necessary to determine the target cortical activation mode of visual prostheses and associate them with the visual stimuli of the subjects. Regarding this, De et al. performed ten different visual stimuli on three mice, using wide field calcium imaging technology to record the primary visual cortex responses of the mice, and used CNN to classify visual stimuli in wide field images. The experiment showcased that the classification accuracy of the CNN was $75.38\% \pm 4.77\%$, and the classification model had excellent generalization [15]. The human eye had excellent adaptive abilities such as wide field of view and low aberration, as well as the ability to pre-process perceived visual information. Long et al. attempted to integrate unfiltered color vision, in device pre-processing, and optical adaptability into visual prostheses, and achieved the ability to reconstruct full-color

images by adjusting color selectivity through bias [16]. Choi et al. designed a visual prosthesis chip with temporal intra pixel image processing function, which can effectively generate edge extraction stimuli. To reduce time-domain processing power consumption, a pulse width modulator on the ground of leakage current was designed. The visual prosthesis chip adopted local echoes on the ground of bipolar stimuli, minimizing image dispersion and surpassing existing advanced image processing visual prosthesis chips [17].

Image processing technology was the key to the application of visual prostheses, and various experts and scholars have conducted extensive research on visual image processing technology. To optimize the performance of multi differential image fusion in detecting ultra high resolution remote sensing images with complex targets and rich texture information, Luo et al. proposed a multi difference image fusion change detection model, which used change vector analysis and spectral gradient difference to construct differential images, and used a visual attention model to obtain differential saliency images. The model was combined with a wavelet transform fusion algorithm to fuse two salient images. After verification with two different datasets, the model had fewer missed alarms and false alarms, and had strong robustness and generalization ability. The three feature extraction modules of the model played a key role [18]. Ping et al. designed an intelligent prosthetic hand on the ground of image recognition. The prosthetic hand used CNN to identify objects that need to be grasped, and the recognition results served as control trigger signals. The experiment demonstrated that the new intelligent bionic hand can achieve five types of human body movements, and the relevant recognition control accuracy was high [19]. Hyperspectral image classification methods on the ground of graph convolutional networks often ignore pixel level spectral spatial features. To address this issue, Ding et al. proposed a multi feature fusion network using different convolutional networks, which reduced computational power requirements while extracting multi-scale spatial and pixel local features to complete hyperspectral image classification. The experiment validated the superiority of the network model [20]. The loss of photonic devices limited the scalability of optical deep networks. Ashtiani et al. designed an integrated end-to-end photon deep neural network that directly processed light waves from pixel arrays on impact plates and performed sub nanosecond image classification when light waves propagate through neural layers. This achieved the scalability of large-scale photon deep neural networks, with an image classification accuracy of up to 93.8% [21]. The detection and recognition of traffic signs had potential value as well as application prospects. Yu et al. on the ground of the fusion model of YOLO-V3 and VGG19 networks, utilized the relationships between multiple images for recognizing traffic signs. After verification on a public dataset, the proposed model can achieve an accuracy of over 90%, which was superior to existing advanced image recognition models [22]. Khaldi et al. designed a framework for generating adversarial networks and CNN models on the ground of conditional deep convolution, which was used to complete the coloring and classification of grayscale and dark images. The importance of this framework has been validated through dataset evaluation [23].

In summary, the analysis of various application technologies and image processing technologies for visual prostheses has been very extensive, but there are still insufficient research on image recognition and processing technologies in the field of visual prostheses. And image recognition and classification are very important for the application of visual prostheses, so research has explored the application of CNN models in visual prosthetic image recognition.

## III. DESIGN OF A VISUAL PROSTHESIS IMAGE RECOGNITION MODEL ON THE GROUND OF IMPROVED VGG ALGORITHM

Visual cortex visual prosthesis is an artificial organ that artificially processes and encodes image information, implants micro-electrode arrays to stimulate the relevant system, as well as helps visually impaired patients recover partial vision. The in vitro image processor of visual prostheses belongs to the key research areas in the application of visual prostheses. To improve the image recognition rate of visual prostheses, this study used the CNN framework VGG to build an image recognition algorithm for visual prosthetic image processing modules. Then it further optimizes the network model with the help of an improved fruit fly (FF) algorithm.

### A. DESIGN AND CONSTRUCTION OF IMAGE RECOGNITION NETWORK STRUCTURE ON THE GROUND OF VGG MODEL

CNN is a deep learning model including convolutional layers, pooling layers(PL), and fully connected layers (FCL). It could markedly learn the characteristics of images as well as other high-dimensional data, and has excellent performance in areas. The classification process of CNN contains an input layer, a multi-layer convolutional layer, a PL, and a FCL. The sliding convolutional kernel (CK) performs pixel multiplication and summation calculations on the feature image, and the depth of the CK is decided by the depth of the input feature image [24], [25]. A schematic of the VGG model, one of the CNN architectures, is shown in Figure 1.

The size calculation of the output feature map (FM) of the convolutional layer as expressed by equation (1), where $N_{out}$ and $N_{in}$ respectively represent the size of the output and input FM. $k$ represents the size of the CK. $p$ represents the filling unit of the FM, which avoids the loss of edge position features in the output image. $s$ represents the sliding step size of the convolutional window.

$$N_{out} = \frac{N_{in} - k + 2p}{s} + 1 \tag{1}$$

After activation, the final output FM is obtained, and the PL performs a down-sampling operation on the FM. The PL decreases the size of the FM through taking the maximum or average values at different windows, i.e. average pooling

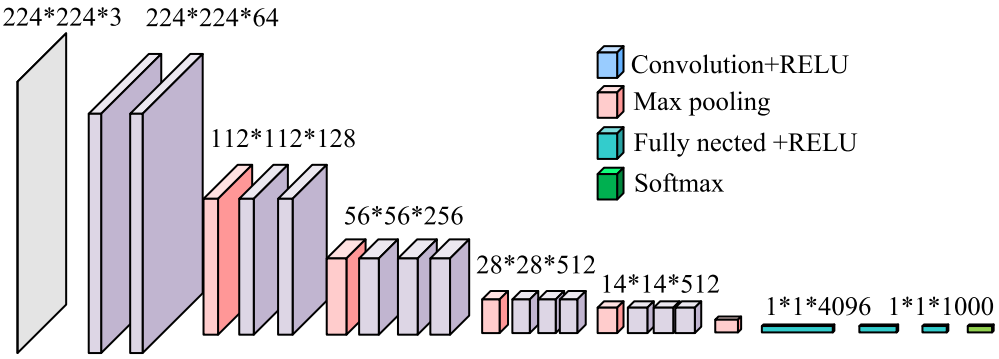


**FIGURE 1.** Schematic diagram of CNN structural models.

or maximum pooling. The schematic of the operations of convolution and pooling is shown in Figure 2. The size of the pooling window determines the size of the output feature image. Through the down-sampling operation of the PL, the image can extract more stable features. In CNN, the number of CK gradually increases with the complexity of the network model and the depth of the network, resulting in an explosive increase in the quantity of FM with the calculation of the network model. The role of PL greatly reduces the dimensionality of feature images, reduces network parameters and computational complexity. The FCL connects the FM output from the convolutional and PL to one or more FCL for final classification or regression [26].

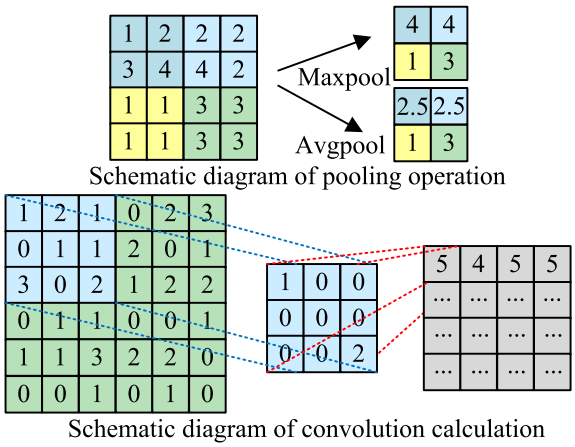


**FIGURE 2.** Schematic diagram of convolutional and pooling operations.

The commonly used CNN architectures include LeNet-5, AlexNet, GoogLeNet (Inception v1, Inception v3), VGG-16, ResNet-50, etc. LeNet-5 belongs to the earliest CNN architectures, including alternating convolutional and PL, FCL, with fewer network parameters and lower complexity, making it suitable for simple handwritten digit classification and recognition. AlexNet is a CNN architecture that has made a breakthrough in the ImageNet image classification competition. Compared to the LeNet-5 architecture, AlexNet introduces ReLU nonlinear activation functions, Dropout regularization, and GPU parallel computing acceleration modules. GoogLeNet adopts the Inception module, which improves the stability and accuracy of the network by serializing and paralleling multiple CK of different sizes. ResNet is an architecture that solves the gradient vanishing problem in deep CNN by introducing residual connections [27].

The VGG model is a classic CNN model proposed by a research team at the University of Oxford. The major characteristics is the utilization of multiple $3 \times 3$ smaller CK and PL to deepen the quantity of layers in the network model. It is mainly divided into VGG16 and VGG19. The stacking of network structure layers enables the network for gradually learning more complex and abstract characteristics, thereby enhancing the classification. The VGG model possesses the merits of simple structure, easy understanding and implementation, and good generalization [28]. Compared to VGGNet-16, the depth of the network of VGGNet-19 is deeper, which makes VGGNet-19 capable of learning and capturing more complex image features with better performance on image visual recognition tasks, so the CNN architecture used in the study is VGGNet-19 considering the large-scale tedious task of visual prosthesis image recognition. The application process of VGGNet-19 in visual prostheses is shown in Figure 3. The external device worn by the user captures visual information through a micro-camera. This information is then processed and recognized using the VGGNet-19 architectural model. The encoded image recognition information is wirelessly transmitted to the micro-current stimulator, which decodes the electrical signals and transmits them to the patient's optic nerves. This activation of the visual cortex stimulates the phantom vision, resulting in artificial vision.

The VGGNet-19 network model uses a $3 \times 3$ small CK, which to some extent reduces the computational time and storage resources of the model. This study adopts a square CK, and the relevant calculation of the convolutional layer parameters as expressed by equation (2), where $D_i$ and $D_0$ respectively represent the quantity of input and output channels in the convolutional layer. $k$ represents the size of the CK.

$$Conv_{param} = D_0 \times \left(k^2 \times D_i + 1\right) \tag{2}$$

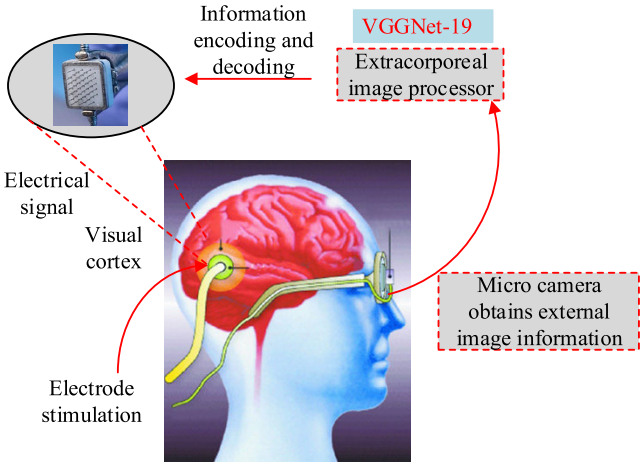


**FIGURE 3.** The working mechanism of visual prostheses.

The calculation amount of each convolutional layer as expressed by equation (3), where *FLOPs* represents the floating-point operation amount, while $W$ and $H$ serve as the elements of the feature image, respectively. Through the operations of equations (2) and (3), it can be seen that the computational resources of the $3 \times 3$ small CK are significantly lower than those of other sizes.

$$FLOPs = \left[\left(D_i \times k^2\right) + \left(D_i \times k^2 - 1\right) + 1\right] \times D_0 \times W \times H \tag{3}$$

The calculation of FCL parameters for VGGNet-19 as expressed by equation (4), where $I$ and $O$ respectively represent the input and output neurons of the FCL. $I \times O$ serves as the number of weights in the FCL. $+1$ represents bias.

$$FLOPs = [I + (I - 1) + 1] \times O = 2 \times I \times O \tag{4}$$

However, the images involved in the visual prosthesis project are mainly low resolution images. If there are too many convolutional layers designed for the VGGNet-19 network model, the size of the feature image in the later stage of network structure extraction will be reduced to $2 \times 2$, which is smaller than the size of the small CK, and the network model cannot capture the minimum receptive field. Meanwhile, in order to accelerate the training speed of network models, existing research usually optimizes the FCL, using batch processing computing techniques to process the FCL. However, this operation increases the pressure on storage and computing resources of computing hardware, and needs to achieve a balance with the transmission data bandwidth. In this regard, the study optimized the clipping of VGGNet-19 and constructed a four segment VGG convolution. And it used average pooling instead of FCL optimization. In order to reduce the negative impact of readjusting convolutional layer features on the slow convergence speed of the network, the study used $1 \times 1$CK to compress the number of FM channels, reduce a large amount of convolutional computation, and achieve the goal of improving network prediction speed. This is to achieve faster image recognition and classification speed

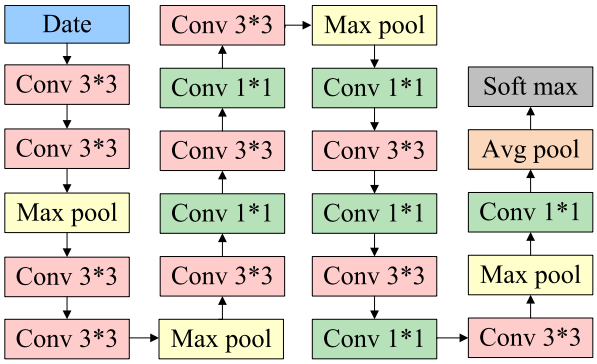


**FIGURE 4.** Schematic diagram of optimized VGG network structure.

while ensuring the visual needs of the visual prosthesis. The optimized VGGNet-19 network model structure is shown in Figure 4.

In the traditional VGGNet-19 model, the FCL parameters account for the majority of the entire network parameters. For low resolution images, the number of neurons output from the FCL can be reduced or the FCL could be further layered. This is to decrease the quantity of FCL parameters and accelerate the calculation speed of the FCL, but the hardware storage requirements are higher. Therefore, as shown in Figure 4, this study replaced the FCL with an average PL, reducing the number of parameters while also reducing the demand for hardware storage resources. Meanwhile, this study used a $1 \times 1$ size CK for compressing the quantity of FM channels, further reducing the convolutional computation. At this time, the parameter calculation of the convolutional layer as expressed by equation (5), where $m$ serves as the quantity of compressed input channels of the FM and the quantity of $1 \times 1$ CKs.

$$P_i = p_{i,m} + p_{i,k} = m \times (d_i + 1) + d_0 \times \left(k^2 \times m + 1\right) \tag{5}$$

Finally, the ReLu activation function in the traditional VGGNet-19 model is modified to ReLu6 to ensure the numerical resolution of the network model and reduce accuracy loss.

### B. DESIGN OF IMAGE RECOGNITION MODEL ON THE GROUND OF IMPROVED DROSOPHILA ALGORITHM OPTIMIZATION

For enhancing the image classification accuracy of the VGGNet-19 model and achieve accurate recognition of visual images, this study introduces the FOA to further improve the optimized VGGNet-19 model. FOA is a heuristic optimization algorithm that simulates the foraging behavior of FF. It seeks the optimal solution through simulating the foraging behavior and information transmission of FF during the foraging process. The foraging process of FF populations is showcased in Figure 5 [29], [30].

The FOA algorithm is simple and easy to implement, with little dependence on initial solutions. It has shown good application performance in fields such as function optimization,

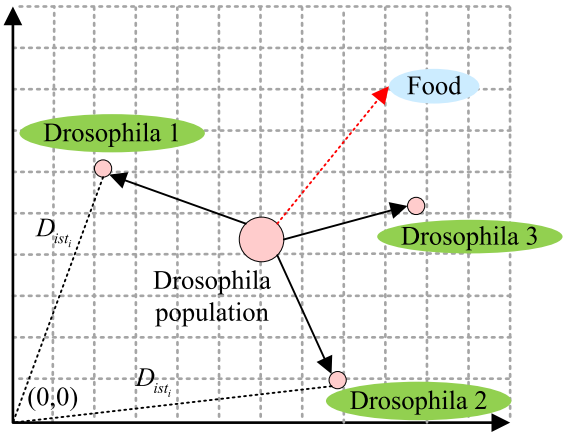


**FIGURE 5.** Foraging process of FF population.

image processing, and machine learning. The entire workflow of FOA is shown in Figure 6.

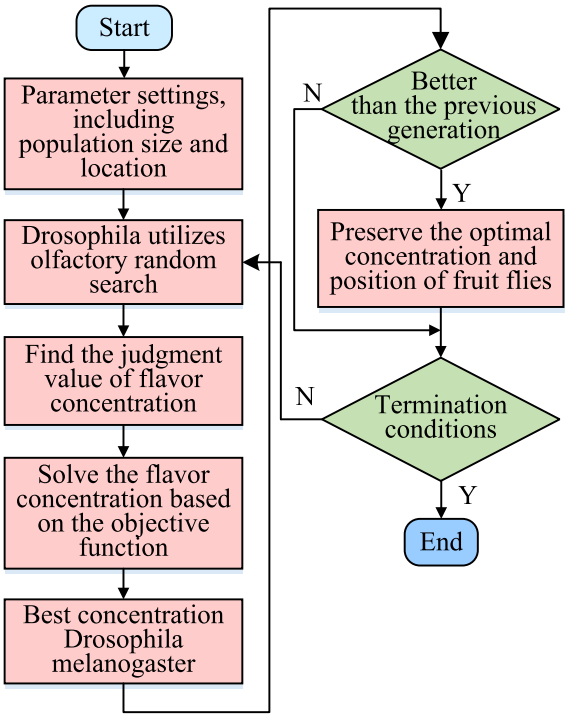


**FIGURE 6.** Schematic diagram of the workflow of the FF algorithm.

Firstly, it sets the population size and maximum iteration of FF. The FF population forages in the $j$-dimensional space, and the position coordinates of the FF are represented as $X_i = (X_{i,1}, X_{i,2}, \ldots, X_{i,j})$ and $Y_i = (Y_{i,1}, Y_{i,2}, \ldots, Y_{i,j})$. The population of FF is randomly searched, and the process of updating individual positions as expressed by equation (6), where $X_{-axis}$ and $Y_{-axis}$ respectively represent the position coordinates of the FF at the previous moment.

$$\begin{cases} X_i = X_{-axis} + \text{RandomValue} \\ Y_i = Y_{-axis} + \text{RandomValue} \end{cases} \tag{6}$$

The relevant calculation of FF as expressed by equation (7), where $D_{ist_i}$ represents the distance from the FF to the origin. $S_i$ represents the odor concentration judgment value.

$$S_i = \frac{1}{D_{ist_i}} = \frac{1}{\left(\sqrt{(X_{i,1})^2 + (Y_{i,1})^2}, \sqrt{(X_{i,2})^2 + (Y_{i,2})^2}\right)} \tag{7}$$

The odor concentration value is solved on the ground of the objective function set by the model, and the calculation process as expressed by equation (8).

$$Smell_i = function\,(S_i) \tag{8}$$

It identifies the FF with the best taste concentration value in the FF population, retains the odor concentration and position coordinates of this FF, and the FF population approaches the optimal FF. The calculation process as expressed by equation (9).

$$\begin{cases} Smellbest = bestSmell \\ X_{-axis} = X\,(bestindex) \\ Y_{-axis} = Y\,(bestindex) \end{cases} \tag{9}$$

The initial process of FOA requires a large quantity of parameters to be set. The parameters affect each other. However, there is no clear reference standard for setting parameters, and how to set reasonable parameters to achieve ideal optimization efficiency has become the key to the Drosophila algorithm. According to the fitness function of the FF algorithm, the mutation probability of the FF population is low. When the iterations grows, the diversity of the population gradually reduces, and the search space is limited, leading to the FF algorithm often falling into local optima. In addition, in traditional FF algorithms, the FF population searches for food within the search range with a fixed step size, resulting in slower convergence speed of the algorithm [31], [32]. Therefore, in combination with the needs of visual prosthetic image recognition models, this study introduces Gaussian backbone mutation mechanism, dynamic step size mechanism, and random abandonment strategy to improve FOA. It places the improved FF algorithm before the soft max classification layer to optimize the classification performance. The workflow diagram of the optimized FF algorithm for improving VGG is shown in Figure 7.

This study utilizes the Gaussian backbone mutation mechanism for enhancing the sustained search capability of the Drosophila algorithm. Firstly, it generates a random number $V\,(i, j)$ with a Gaussian distribution, and the calculation process as expressed by equation (10), where $\mu$ represents the parameter representing the mean. $\sigma$ represents the standard deviation of the parameter.

$$\begin{cases} \mu = (X\_axis\,(j) + X\,(i,j))\,/2 \\ \sigma = abs\,(X\_axis\,(j) + X\,(i,j)) \\ V\,(i,j) = normrnd\,(mu, sigma) \end{cases} \tag{10}$$

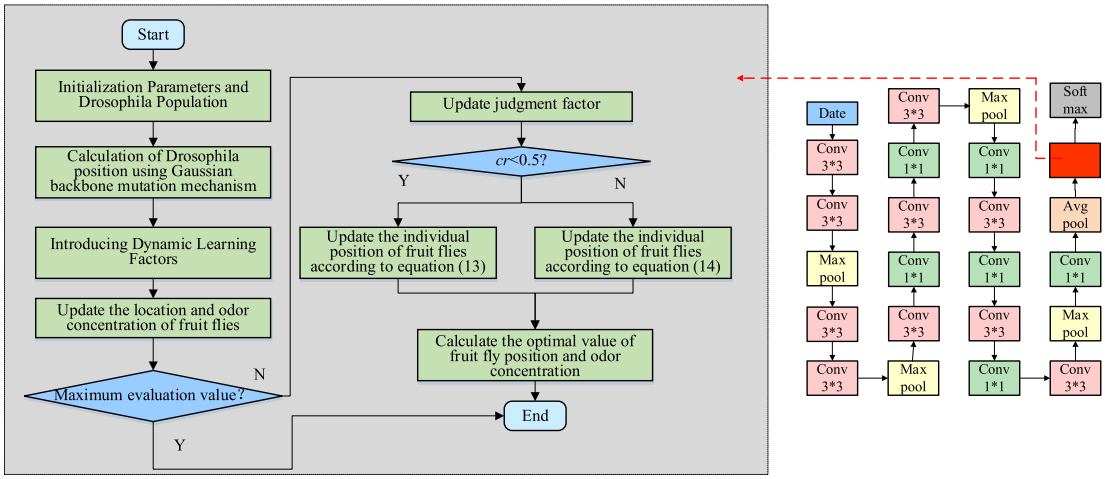


**FIGURE 7.** Optimized drosophila algorithm for improving VGG process.

After introducing the Gaussian backbone mutation mechanism, the position calculation of the Drosophila population as expressed by equation (11), where $k$ is a random number within the range of (0, 1). Both $k1$ and $k2$ represent random integers within the range of [1, *sizepop*]. Then it calculates the odor concentration of FF and updates the individual's fitness value. The Gaussian backbone mutation mechanism expands the search space of the Drosophila population, to some extent avoiding Drosophila individuals from falling into local optima during search.

$$V(i,j) = X\_axis(j) + k * (X(k1,j) - X(k2,j)) \quad (11)$$

A fixed step size can easily make it difficult for the algorithm to approach the location of the food source in the later stage of iteration. This study introduced a dynamic learning factor to convert the fixed step size in the traditional FF algorithm into a dynamic step size. The calculation of the learning factor *ds* as expressed by equation (12).

$$ds = 0.5 * ones(30, 1) + 0.1 \tan(pi * (rand(sizepop, 1) - 0.5)) \quad (12)$$

The calculation of the random distribution position of the FF population as expressed by equation (13), where $r1$ and $r2$ both represent random numbers within the range of [0, *popsize*]. The transformation of dynamic step size enhances the global search capability of the Drosophila algorithm, and collaborates with the Gaussian backbone mechanism to prevent the algorithm from falling into local optima too early.

$$\begin{cases} X_{i,j} = X_{axis} + \mathrm{ds.} * (X_{axis} - X(r1) + X(r2)) \\ Y_{i,j} = Y_{axis} + \mathrm{ds.} * (Y_{axis} - Y(r1) + Y(r2)) \end{cases} \quad (13)$$

When the FF algorithm calculates the optimal concentration value, the flight route determined by the FF population has strong randomness and fuzziness, and the convergence speed of the FF algorithm is slow, with low optimization accuracy. In this regard, the study introduced a abandonment strategy to select the location of FF. It utilizes a Gaussian distribution for generating a judgment factor $cr$, and determines the calculation method for updating the position of FF on the ground of the size of the judgment factor $cr$. When $cr < 0.5$, the current position of FF is calculated according to equation (13). When $cr > 0.5$, it calculates the position of the FF on the ground of the optimal solution, as expressed by equation (14).

$$\begin{cases} cr = normrnd(nu\_cr, 0.1) \\ X(i,j) = X\_axis(j) + cr(j) * tempx \\ Y(i,j) = Y\_axis(j) + cr(j) * tempy \end{cases} \quad (14)$$

The improved FF algorithm is optimized for the VGGNet-19 model, and after being optimized and solved by intelligent optimization algorithms, the image recognition accuracy is expected to be improved. Meanwhile, the improved VGG model subtracting the variance after subtracting the average of the three-channel pixels from the RGB when pre-processing low resolution images, and then inputs them into the Imp-FOA-VGGNet-19 model for image classification and recognition.

## IV. PERFORMANCE TESTING OF IMAGE RECOGNITION ALGORITHM ON THE GROUND OF VGG MODEL

For testing the image recognition model designed in the study, different performance testing experiments and image recognition effect analysis experiments were designed.

### A. PERFORMANCE TEST OF DROSOPHILA IMPROVED VGG IMAGE RECOGNITION MODEL

It selected the CIFAR-10 dataset, ImageNet, Category_Flower, and Caltech10 as the test experimental dataset. CIFAR10 is a small dataset for recognizing pervasive objects put together by Hinton students Alex Krizhevsky and Ilya Sutskever. The entire CIFAR-10 dataset includes 60000 $32 * 32$ sized RGB images of different categories. There are currently 14197122 images in ImageNet is a large computer vision dataset that was started by Stanford University

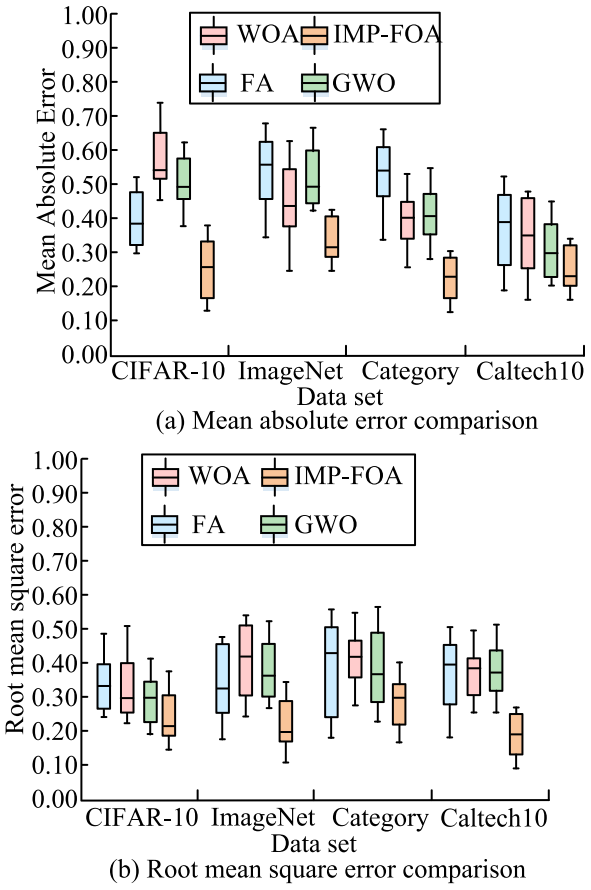


**FIGURE 8.** Error comparison of different intelligent optimization algorithms.

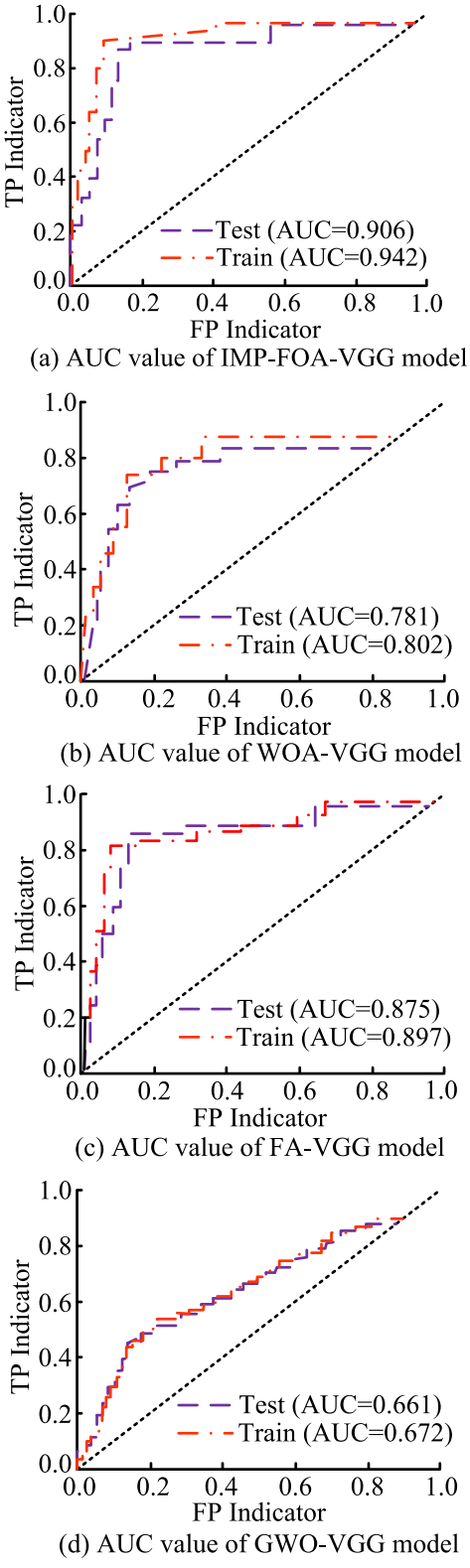


**FIGURE 9.** Comparison results of AUC values for VGG models optimized by different algorithms.

and other organizations in 2007. ImageNet, divided into 21841 categories, with an average image resolution of 469 × 387 pixels. The Caltech10 dataset was filmed and labeled for publication by the videographer, the Caltech10 dataset contains 101 types of images, each with approximately 40-800 images, and each image has a size of approximately 300× 200 pixels. Category_Flower is a dataset of 17 categories of common flowers in the UK, published by the VGG at the University of Oxford. Each containing 80 flower images. The dataset contains a training set and a testing set in a 7:3 ratio.

The step size of all convolutional layers in the Imp-FOA-VGGNet-19 model designed for research is 1, and the number of CKs for the four convolutional segments is 96, 192, 384, and 384, respectively. The pooling window size is 2 ∗ 2, and the sliding step size is 2. The third and fourth convolutional segments are used for channel compression. All network models and optimization algorithms are written on the ground of the Python platform. Firstly, an optimized VGG model performance analysis experiment was conducted to evaluate the application effect of the improved Drosophila algorithm in the model. It will compare the improved FOA studied and designed with Whale Optimization Algorithm (WOA), Firefly Algorithm (FA), and Grey Wolf Optimizer (GWO) to evaluate the optimal solution error of the intelligent optimization algorithm in the VGG model. Then it selects Mean Absolute Error (MAE) and Root Mean Squared Error

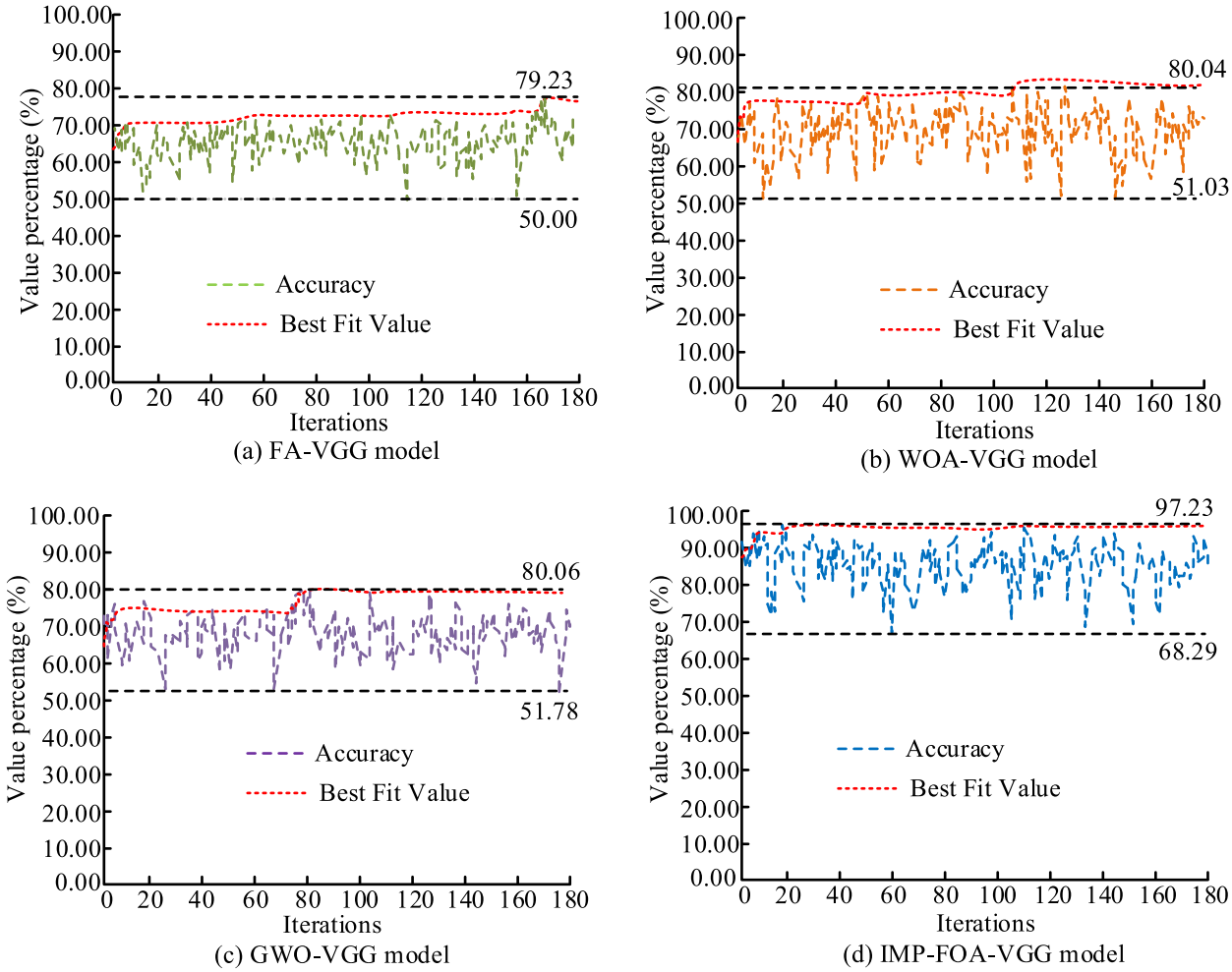


**FIGURE 10.** Performance comparison of VGG models improved by different intelligent optimization algorithms.

(RMSE) as evaluation indicators, and the relevant outcomes are showcased in Figure 8.

Both MAE and RMSE represent the size of measurement errors, and RMSE is greatly affected by outliers. RMSE is chosen as a supplement to the MAE index to compare the optimization and optimization errors of different intelligent optimization algorithms. Figure 8 shows that there are significant differences in MAE and RMSE among several intelligent optimization algorithms under different datasets. The two error indicators of the improved FOA are significantly lower than the other three algorithms. As shown in Figure 8 (a), the median values of the improved FOA solution for MAE on different datasets are all below 0.4. From Figure 8 (b), it can be seen that the median values of the RMSE solved by the improved FOA are all below 0.3 on different datasets, indicating a low level of error convergence. It can be seen that in the process of optimizing the VGG model, improving FOA can demonstrate lower optimization errors, which possesses an essential influence on improving the recognition accuracy of the image recognition model.

It selects Area Under the Curve (AUC) under the Receiver Operating Characteristic Curve (ROC) as the evaluation indicator to examine the AUC size of VGG models optimized by different algorithms on the CIFAR-10 dataset. The relevant outcomes are showcased in Figure 9. AUC is commonly used in binary classification problems, representing the probability that a correctly classified positive sample ranks before a mis-classified negative sample, with AUC values ranging from 0 to 1. Figure 9 (a)shows that the ROC curve of the Imp-FOA-VGGNet model is located at the top of the coordinate axis, with a maximum AUC value of 0.942. In Figures 9 (b) and (c), the AUC values of the WOA-VGG model and FA-VGG model only reached 0.802 and 0.897, respectively. As can be seen from Figure 9 (d), the ROC curve of GWO-VGG is located at the bottom of the coordinate axis, with a minimum AUC value of only 0.661. The larger the AUC value, the more excellent the classification effect of the model, and the model can distinguish between positive and negative samples well. It showcases that the performance of the Imp-FOA-VGGNet model designed in the study is

superior to the VGG recognition model optimized by other algorithms.

The iteration number of all VGG models is set to 180, and the accuracy and optimal fitness curve results of different optimized VGG image recognition models are showcased in Figure 10. Figures 10(a), (b), and (c) showcases that on the Category_Flower dataset, the fitness value curves of WOA, GWO, and FA intelligent optimization algorithms still improve in the middle and late stages of the iteration. The fitness curve of the Imp FOA algorithm converges at the beginning of the iteration and stabilizes at around 97.00% of the highest value. This indicates that different intelligent optimization algorithms have jumped out of the current fitness value in the optimization of improving the VGG model. The algorithm has optimized the VGG model's ability to recognize correct samples. But as shown in Figure 10 (d), the Imp FOA algorithm designed for research has a fast convergence speed and the optimal value. By comparing the improved VGG models with different algorithms, it can be found that the Imp-FOA-VGG model in Figure 10 (d)has the highest recognition accuracy range, ranging from 68.29% to 97.23%, and has better overall performance.

It compares the performance of the Imp-FOA-VGG model designed for research with existing advanced image recognition models, including the traditional CNN VGG, image classification model EfficientNet, and GoogLeNet, using the F1 value and loss curve of the model as evaluation indicators. The F1 indicator is the harmonic average of accuracy and recall, and classification models often find it difficult to achieve a balance between these conflicting indicators. Therefore, this study selected the F1 value to evaluate the image recognition model. The average test results of four different datasets are showcased in Figure 11. Figure 11 (a) showcases that the F1 value curve of the Imp-FOA-VGG model designed in the study is located at the top of the coordinate axis, reaching a maximum value of 91.47%, which is 14.06 percentage points higher than the minimum value of 77.41% in the Efficient Net model. The F1 values of the other two image recognition models are in the range of 80% -90%. As shown in Figure 11 (b),comparing the loss function curves of different models, the Imp-FOA-VGG model's loss function curve converges to the minimum value, and the convergence curve steadily decreases, with the fastest convergence speed of the loss curve. The loss function represents the degree of consistency between the predicted and true values of the model. The smaller the loss function value, the higher the fitting degree. Therefore, the overall performance of the Imp-FOA-VGG model is better.

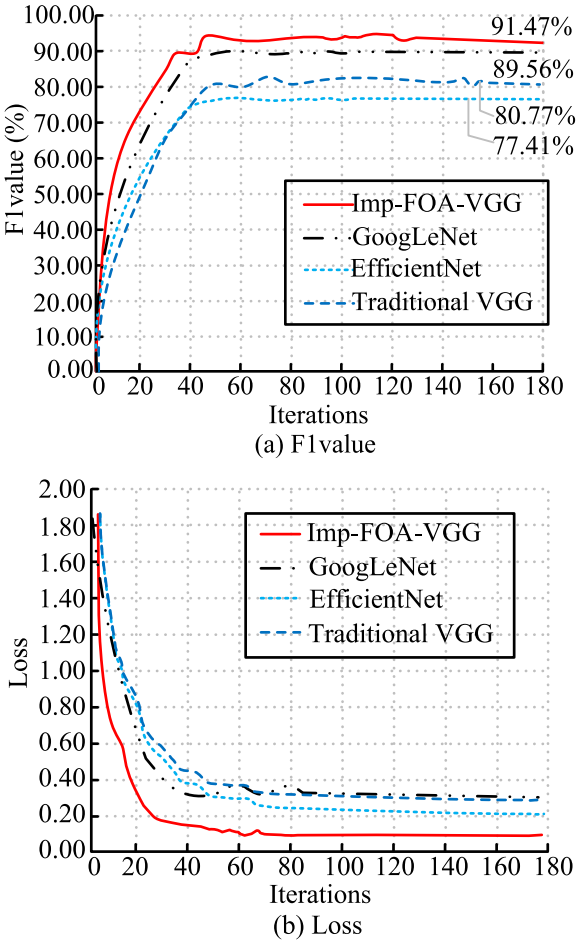


**FIGURE 11.** Performance comparison of different image recognition models.

### B. ANALYSIS OF THE APPLICATION EFFECT OF DROSOPHILA IMPROVED VGG IMAGE RECOGNITION MODEL

To verify the application effect of the image recognition model designed in the study on visual prostheses, and considering the daily needs of users, the recognition accuracy and R-squared index of different recognition models are analyzed. The R-squared index reflects the degree of fit between predicted values and true values. The relevant outcomes are showcased in Figure 12. Figure 12 showcases that the Imp-FOA-VGG model designed on four different publicly available datasets has the highest recognition accuracy and R-squared metrics. On the Category_Flower dataset in Figure 12(a), the highest accuracy of the model reached 96.45%. In Figure 12 (b), the model'sR-squared index reached 0.94. The evaluation index values of the other three models are mostly below 80%, significantly lower than the Imp-FOA-VGG model. This indicates that the Imp-FOA-VGG model has high practicality in image recognition in different fields, high recognition accuracy, and high consistency with real images. This is crucial for image conversion and clinical application of visual prostheses.

It applies the Imp-FOA-VGG model designed in the research to the visual prosthesis image processor and compares it with the visual effect of real vision. This study randomly selects 10 experimental personnel who did not have obvious eye diseases for human visual contrast testing, with visual examination angles including contrast vision and functional vision, and set a test sample size of 500. On the

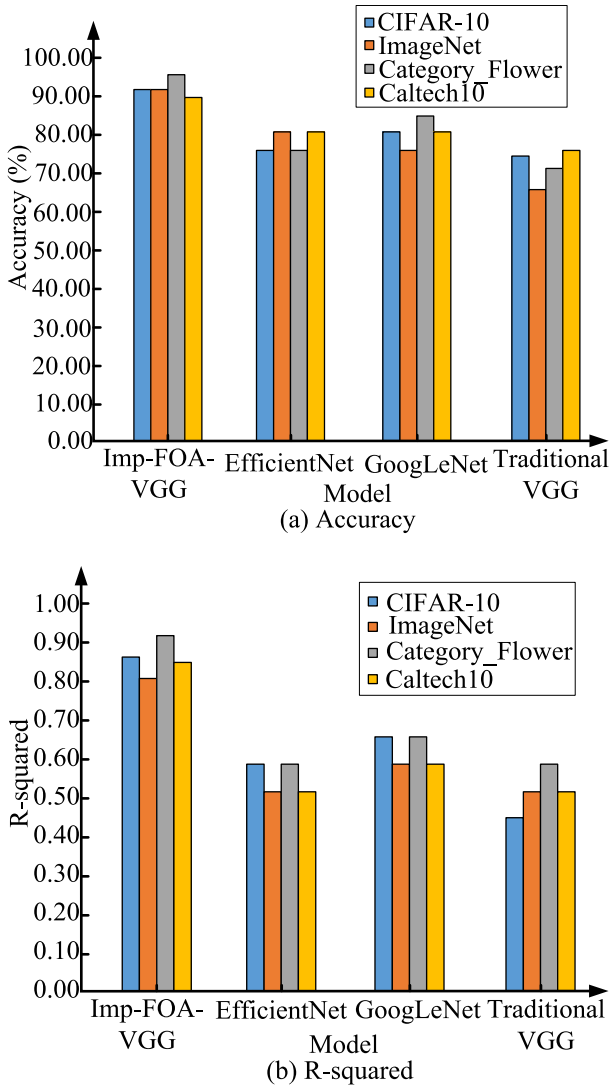


**FIGURE 12.** Application effects of different recognition models.

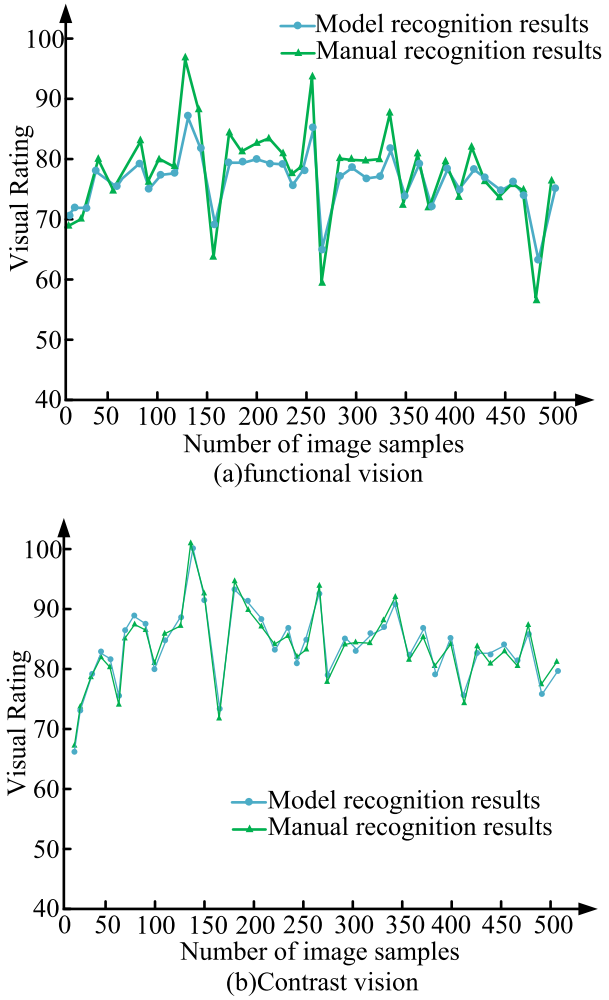


**FIGURE 13.** Comparison of visual prosthesis application and normal visual function.

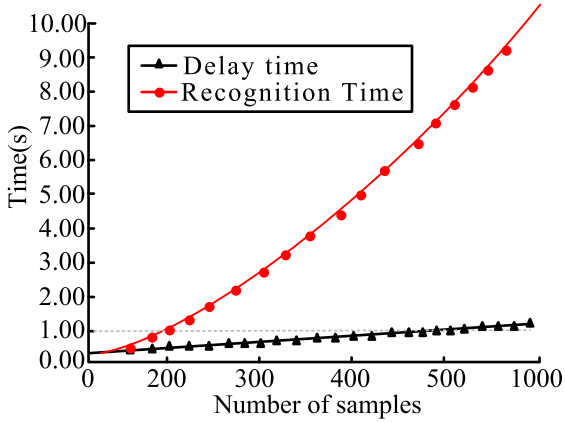


**FIGURE 14.** Recognition time and delay time of visual prosthesis algorithm application.

ground of the work experience of industry experts or doctors, the relevant outcomes are showcased in Figure 13. Figure 13 showcases that the algorithm designed in the experiment has achieved significant results in the application of visual prostheses. Compared with human natural vision, the contrast visual acuity in Figure 13(a) differs less from the functional visual acuity in Figure 13(b). The image processing and perception level are relatively consistent, and there is no significant deviation. This indicates that the model can play a key role in visual prosthetic image processors and promote the application of visual prostheses.

The information collected by normal human vision is transmitted through nerves to the brain for processing with only a very small time difference, and the human body can hardly perceive the delay in the data processing. The test results of image processing speed and brain perception delay time of visual prostheses are showcased in Figure 14. Figure 14 showcases that in visual prostheses, as the number of image processing samples increases, the model time tends to increase. Compared with artificial natural vision, the speed of image processing is significantly faster. However, as the number of samples increases, the processing delay time of visual prostheses also shows an upward trend, but

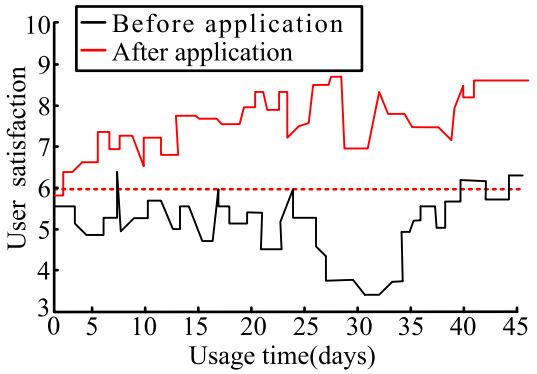


**FIGURE 15.** Long term user satisfaction survey.

the maximum value is still less than 1 second. This study suggests that this is a minor delay that does not affect the image recognition of visual prostheses and does not interfere with the user experience.

The image processing model designed in the study was simulated and used in visual prostheses. 10 patients who were implanted with visual prostheses were randomly selected and followed up for a long-term satisfaction period of 45 days. User experience was collected through a questionnaire survey. The relevant outcomes are showcased in Figure 15. Figure 15 showcases that before the introduction of the recognition model, the satisfaction curve of users with visual prostheses was generally below the 6-point range, indicating poor satisfaction. Satisfaction fluctuates significantly over time, with an increase in satisfaction in the later stages, possibly due to the patient's adaptation to visual prostheses. After introducing the image recognition processing algorithm, the user satisfaction score has significantly improved, all above 6 points, with a maximum satisfaction of 8 points. This indicates that the image recognition model has achieved good feedback in the practical application of visual prostheses.

## V. CONCLUSION

Due to limitations in electronic and optical technologies, the resolution and clarity of visual prosthetic images have been negatively affected to a certain extent. To improve the image processing performance of visual prostheses, the study selected the CNN framework VGG as the foundation, introduced three improvement mechanisms to optimize the FF algorithm, and completed the improvement of the VGG recognition model. A visual prosthetic image recognition model on the ground of the improved VGG was constructed. The experiment demonstrated that the average absolute error and root mean square error of the improved FOA were significantly lower than the other three algorithms, with values below 0.40. The improved image recognition model had the best performance, with the Imp-FOA-VGGNet model's ROC curve located at the top of the coordinate axis and the maximum AUC value of 0.942. The recognition accuracy fluctuated between 68.29% and 97.23%, and the fitness curve converges at the beginning of the iteration, stabilizing at around 97.00%. The F1 value curve of the Imp-FOA-VGG model was located at the top of the coordinate axis, reaching a maximum value of 91.47%, which was 14.06 percentage points higher than the minimum value. And the loss function curve converges to the minimum value, with the fastest convergence speed. In the application of visual prostheses, the recognition accuracy of this model and the R-squared index performed the best on different datasets, with an accuracy of 96.45% and an R-squared index of 0.94. Compared with natural human vision, the contrast and functional visual effects were highly consistent, and the processing speed was fast. The delay time did not affect the practical application of visual prostheses, and has achieved high user satisfaction in long-term monitoring testing. This study helped to promote the practical application of visual prostheses, but it did not focus on targeted improvements in image recognition efficiency, which can be a focus of future research work.

## REFERENCES

[1] A. Viberg, A. Vicente, B. Samolov, J. Hjortdal, and B. Byström, "Corneal transplantation in aniridia-related keratopathy with a two-year follow-up period, an uncommon disease with precarious course," *Acta Ophthalmologica*, vol. 101, no. 2, pp. 222–228, Mar. 2023, doi: 10.1111/aos.15229.

[2] L. van der Wekken-Pas, P. de Haas, R. Wisse, M. Rados, and K. van der Elst, "Candida keratitis and endopthalmitis after corneal transplantation; two case reports, a novel regimen and literature overview of therapeutic options," *Eur. J. Ophthalmol.*, vol. 33, no. 2, pp. NP73–NP78, Mar. 2023, doi: 10.1177/11206721211060140.

[3] S. Obata, R. Matsumoto, M. Iwasa, M. Kakinoki, O. Sawada, T. Sawada, Y. Saishin, and M. Ohji, "Visual field after anti-vascular endothelial growth factor therapy and laser treatment for retinopathy of prematurity," *Graefe's Arch. Clin. Experim. Ophthalmology*, vol. 261, no. 11, pp. 3207–3213, Sep. 2023, doi: 10.1007/s00417-023-06227-6.

[4] J. Wang, H. Zhu, J. Liu, H. Li, Y. Han, R. Zhou, and Y. Zhang, "The application of computer vision to visual prosthesis," *Artif. Organs*, vol. 45, no. 10, pp. 1141–1154, Jul. 2021, doi: 10.1111/aor.14022.

[5] V. Karadima, E. A. Pezaris, and J. S. Pezaris, "Attitudes of potential recipients toward emerging visual prosthesis technologies," *Sci. Rep.*, vol. 13, no. 1, pp. 10963–10978, Jul. 2023, doi: 10.1038/s41598-023-36913-8.

[6] L. Jiang, G. Lu, Y. Zeng, Y. Sun, H. Kang, J. Burford, C. Gong, M. S. Humayun, Y. Chen, and Q. Zhou, "Flexible ultrasound-induced retinal stimulating piezo-arrays for biomimetic visual prostheses," *Nature Commun.*, vol. 13, no. 1, pp. 3583–3596, Jul. 2022, doi: 10.1038/s41467-022-31599-4.

[7] N. R. B. Stiles, V. R. Patel, and J. D. Weiland, "Multisensory perception in Argus II retinal prosthesis patients: Leveraging auditory-visual mappings to enhance prosthesis outcomes," *Vis. Res.*, vol. 182, pp. 58–68, May 2021, doi: 10.1016/j.visres.2021.01.008.

[8] C. Wang, J. Yang, and M. Sawan, "NeuroSEE: A neuromorphic energy-efficient processing framework for visual prostheses," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 8, pp. 4132–4141, Aug. 2022, doi: 10.1109/JBHI.2022.3172306.

[9] A. Omisakin, R. Mestrom, G. Radulov, and M. Bentum, "Sub-milliwatt transceiver IC for transcutaneous communication of an intracortical visual prosthesis," *Electronics*, vol. 11, no. 1, p. 24, Dec. 2021, doi: 10.3390/electronics11010024.

[10] B.-Y. Wang, Z. C. Chen, M. Bhuckory, T. Huang, A. Shin, V. Zuckerman, E. Ho, E. Rosenfeld, L. Galambos, T. Kamins, K. Mathieson, and D. Palanker, "Electronic photoreceptors enable prosthetic visual acuity matching the natural resolution in rats," *Nature Commun.*, vol. 13, no. 1, pp. 6627–6637, Nov. 2022, doi: 10.1038/s41467-022-34353-y.

[11] K.-W. Hintzen, C. Simons, K. Schaffrath, G. Roessler, S. Johnen, F. Jakob, P. Walter, U. Schwaneberg, and T. Lohmann, "BioAdhere: Tailor-made bioadhesives for epiretinal visual prostheses," *Biomaterials Sci.*, vol. 10, no. 12, pp. 3282–3295, Jun. 2022, doi: 10.1039/d1bm01946e.

[12] M. Spencer, T. Kameneva, D. B. Grayden, A. N. Burkitt, and H. Meffin, "Quantifying visual acuity for pre-clinical testing of visual prostheses," *J. Neural Eng.*, vol. 20, no. 1, Jan. 2023, Art. no. 016030, doi: 10.1088/1741-2552/ac9c95.
[13] C. Wang, C. Fang, Y. Zou, J. Yang, and M. Sawan, "SpikeSEE: An energy-efficient dynamic scenes processing framework for retinal prostheses," *Neural Netw.*, vol. 164, pp. 357–368, Jul. 2023, doi: 10.1016/j.neunet.2023.05.002.
[14] X. Xia, X. He, L. Feng, X. Pan, N. Li, J. Zhang, X. Pang, F. Yu, and N. Ding, "Semantic translation of face image with limited pixels for simulated prosthetic vision," *Inf. Sci.*, vol. 609, pp. 507–532, Sep. 2022, doi: 10.1016/j.ins.2022.07.094.
[15] D. De Luca, S. Moccia, L. Lupori, R. Mazziotti, T. Pizzorusso, and S. Micera, "Convolutional neural network classifies visual stimuli from cortical response recorded with wide-field imaging in mice," *J. Neural Eng.*, vol. 20, no. 2, Apr. 2023, Art. no. 026031, doi: 10.1088/1741-2552/acc2e7.
[16] Z. Long, X. Qiu, C. L. J. Chan, Z. Sun, Z. Yuan, S. Poddar, Y. Zhang, Y. Ding, L. Gu, Y. Zhou, W. Tang, A. K. Srivastava, C. Yu, X. Zou, G. Shen, and Z. Fan, "A neuromorphic bionic eye with filter-free color vision using hemispherical perovskite nanowire array retina," *Nature Commun.*, vol. 14, no. 1, pp. 1972–1980, Apr. 2023, doi: 10.1038/s41467-023-37581-y.
[17] D.-H. Choi and D.-W. Jee, "A 1984-pixels, 1.26 nW/Pixel retinal prosthesis chip with time-domain in-pixel image processing and bipolar stimulating electrode sharing," *IEEE J. Solid-State Circuits*, vol. 58, no. 10, pp. 2757–2766, Oct. 2023, doi: 10.1109/JSSC.2023.3284460.
[18] J. Luo, Q. Chen, L. Wang, and Y. Huang, "Multi-difference image fusion change detection using a visual attention model on VHR satellite data," *Remote Sens.*, vol. 15, no. 15, p. 3799, Jul. 2023, doi: 10.3390/rs15153799.
[19] P. Shi, K. Fang, and H. Yu, "Design and control of intelligent bionic artificial hand based on image recognition," *Technol. Health Care*, vol. 31, no. 1, pp. 21–35, Jan. 2023, doi: 10.3233/thc-213320.
[20] Y. Ding, Z. Zhang, X. Zhao, D. Hong, W. Cai, C. Yu, N. Yang, and W. Cai, "Multi-feature fusion: Graph neural network and CNN combining for hyperspectral image classification," *Neurocomputing*, vol. 501, pp. 246–257, Aug. 2022, doi: 10.1016/j.neucom.2022.06.031.
[21] F. Ashtiani, A. J. Geers, and F. Aflatouni, "An on-chip photonic deep neural network for image classification," *Nature*, vol. 606, no. 7914, pp. 501–506, Jun. 2022, doi: 10.5061/dryad.q2bvq83mw.
[22] J. Yu, X. Ye, and Q. Tu, "Traffic sign detection and recognition in multiimages using a fusion model with YOLO and VGG network," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 9, pp. 16632–16642, Sep. 2022, doi: 10.1109/TITS.2022.3170354.
[23] Y. Khaldi and A. Benzaoui, "A new framework for grayscale ear images recognition using generative adversarial networks under unconstrained conditions," *Evolving Syst.*, vol. 12, no. 4, pp. 923–934, May 2020, doi: 10.1007/s12530-020-09346-1.
[24] H. Mano, S. Fujiwara, and N. Haga, "How children with congenital limb deficiencies visually attend to their limbs and prostheses: Eye tracking of displayed still images and visuospatial body knowledge," *Develop. Neurorehabilitation*, vol. 24, no. 8, pp. 547–554, Apr. 2021, doi: 10.1080/17518423.2021.1901151.
[25] Y. Yang and X. Song, "Research on face intelligent perception technology integrating deep learning under different illumination intensities," *J. Comput. Cognit. Eng.*, vol. 1, no. 1, pp. 32–36, Jan. 2022, doi: 10.47852/bonviewjcce19919.
[26] S. Sordillo, A. Cheikh, A. Mastrandrea, F. Menichelli, and M. Olivieri, "Customizable vector acceleration in extreme-edge computing: A RISC-V software/hardware architecture study on VGG-16 implementation," *Electronics*, vol. 10, no. 4, p. 518, Feb. 2021, doi: 10.3390/electronics10040518.
[27] A. Malini, P. Priyadharshini, and S. Sabeena, "An automatic assessment of road condition from aerial imagery using modified VGG architecture in faster-RCNN framework," *J. Intell. Fuzzy Syst.*, vol. 40, no. 6, pp. 11411–11422, Jun. 2021, doi: 10.3233/jifs-202596.
[28] M. Lachmann, E. Rippen, D. Rueckert, T. Schuster, E. Xhepa, M. von Scheidt, C. Pellegrini, T. Trenkwalder, T. Rheude, A. Stundl, R. Thalmann, G. Harmsen, S. Yuasa, H. Schunkert, A. Kastrati, M. Joner, C. Kupatt, and K. L. Laugwitz, "Harnessing feature extraction capacities from a pre-trained convolutional neural network (VGG-16) for the unsupervised distinction of aortic outflow velocity profiles in patients with severe aortic stenosis," *Eur. Heart J. - Digit. Health*, vol. 3, no. 2, pp. 153–168, Apr. 2022, doi: 10.1093/ehjdh/ztac004.
[29] S. S. Mohar, S. Goyal, and R. Kaur, "Optimum deployment of sensor nodes in wireless sensor network using hybrid fruit fly optimization algorithm and bat optimization algorithm for 3D environment," *Peer-to-Peer Netw. Appl.*, vol. 15, no. 6, pp. 2694–2718, Aug. 2022, doi: 10.1007/s12083-022-01364-x.
[30] Z. Tian, "Echo state network based on improved fruit fly optimization algorithm for chaotic time series prediction," *J. Ambient Intell. Humanized Comput.*, vol. 13, no. 7, pp. 3483–3502, Jul. 2022, doi: 10.1007/s12652-020-01920-4.
[31] J. He, Z. Peng, J. Qiu, D. Cui, and Q. Li, "A novel elitist fruit fly optimization algorithm," *Soft Comput.*, vol. 27, no. 8, pp. 4823–4851, Dec. 2022, doi: 10.1007/s00500-022-07621-8.
[32] Q. Liao, Q. Yin, L. Xie, and G. Yin, "Improved exponential model for thermal error modeling of machine-tool spindle based on fruit fly optimization algorithm," *Proc. Inst. Mech. Eng., C, J. Mech. Eng. Sci.*, vol. 236, no. 12, pp. 6912–6922, Feb. 2022, doi: 10.1177/09544062211066243.

**ZHIGUANG LI** was born in Henan, China. He received the Bachelor of Arts degree in journalism (new media) from Zhengzhou University, China, in 2020. He is currently pursuing the Ph.D. degree in media art and production with the Graduate School of Advanced Imaging Science, Multimedia and Film, Chung-Ang University, South Korea.

**BAITAO LI** was born in Shangcai, Henan, China. She received the bachelor's degree from Xinxiang University, China. She is currently pursuing the master's degree with Henan University, Henan. Her research interests include film special effects, film narratives, and virtual reality.

**SURNG GAHB JAHNG** received the B.S., M.S., and Ph.D. degrees in electronic engineering from Chung-Ang University, Seoul, South Korea, in 1988, 1990, and 2000, respectively. He joined the Faculty with Chung-Ang University, in 2003. His main research interests include cinema special effects and culture technology.

**CHANGYONG JUNG** received the Ph.D. degree from the Department of Advanced Image, Graduated School of Advanced Imaging Science, Multimedia and Film, Chung-Ang University. His main research interests include 3-D projection mapping, interface design, 3-D contents, media façade, new media art, and interactive media art.