Model for the online frontal image selection of silkworm pupae using machine vision

Feng Guo[1,2], Jing Li[1], Wei Qin[2], Chunjiang Zhao[2,3], Guanglin Li[2*]

(1. *Modern Agricultural Equipment Research Institute, Xihua University, Chengdu 610039, China*;
2. *College of Engineering and Technology, Southwest University, Chongqing 400715, China*;
3. *National Engineering Research Center for Information Technology in Agriculture, Beijing 100097, China*)

Abstract: The sorting of male and female silkworm pupae is an essential process of silkworm breeding, with its accuracy directly affecting the quality of hybrid silkworm eggs and silk. Gonadal characteristics serve as a reliable basis for sex identification in silkworm pupae; however, the gonads only exist on the positive side of the tail. Due to the unique geometry of silkworm pupae, online sex recognition based on machine vision requires flipping and taking many photos of the same silkworm pupae. Thus, accurately selecting the frontal image from multiple images of the same silkworm pupae in different poses is a prerequisite for subsequent sex identification. To address this challenge, we proposed SPNet-GS (Silkworm Pupae Network for Gonad Selection), a lightweight model for online selection of frontal silkworm pupae images. The model first employed a large kernel convolution to enhance the receptive field and capture the relevant information between adjacent pixels. Then the correlation between long-distance pixels under multi-scale information can be obtained by dilated convolutions. Finally, the correlation information between near and far pixels was fused to enhance feature extraction. Experimental results demonstrated that our method outperforms other models with an average accuracy of 98.41% and an average $F1$ score of 99.02%. The average inference time of each image was 0.03 s, which can fully meet the requirements of online selection of male and female silkworm pupae. Moreover, the gender identification accuracy rates using the selected frontal image and gonad region image reached 84.68% and 94.58%, respectively. These results were 10% and 19.90% higher than using multi-pose images for sex identification, demonstrating the effectiveness of the frontal image selection strategy. The findings of this investigation may provide a valuable reference for the machine vision-based intelligent online sorting of silkworm pupae by gender.

Keywords: frontal image selection, gonad image, large kernel convolution, sex identification of silkworm pupae
DOI: 10.25165/j.ijabe.20261901.8409

Citation: Guo F, Li J, Qin W, Zhao C J, Li G L. Model for the online frontal image selection of silkworm pupae using machine vision. Int J Agric & Biol Eng, 2026; 19(1): 163–169.

1 Introduction

As the birthplace of the mulberry silkworm cocoon industry, China not only has a long history and culture of silkworm breeding, but also plays an important role in the mulberry silkworm cocoon market[1]. The classification accuracy of male and female silkworm pupae can directly affect the yield and quality of silkworm cocoons in silkworm pupae breeding[2]. Nowadays, sex identification of silkworm pupae still suffers from heavy labor intensity and high breeding costs. Therefore, it is very important for the silkworm cocoon production industry to separate male and female pupae quickly and accurately.

Hitherto, researchers have been actively exploring non-destructive and reliable techniques for silkworm pupae gender differentiation. Sumriddetchkajorn and Kamtongdee[3-6] presented a method for silkworm pupae sex identification utilizing optical penetration. Liu et al.[7] conducted a study on silkworm gender classification based on magnetic resonance imaging (MRI) technology. Cai et al.[8] presented a method for silkworm pupae sex recognition based on X-ray imaging technology. Raj et al.[9] introduced a multi-sensor system for silkworm cocoon gender classification and separation by integrating weight and shape-related features. Furthermore, many researchers have also employed spectral technology for gender identification in silkworm pupae[10-12]. Unfortunately, the aforementioned studies have been limited by factors such as equipment cost, accuracy, and model generalizability, which have hindered their widespread adoption.

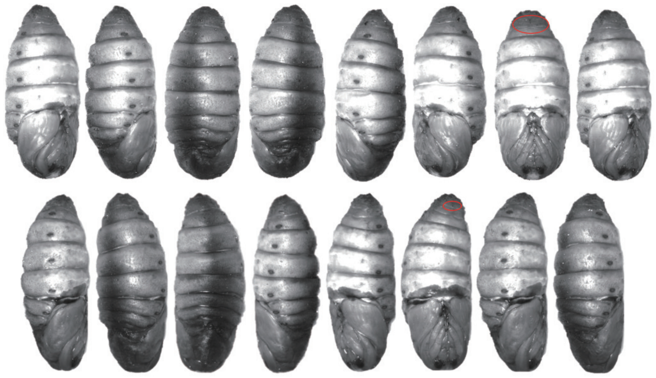
With the rapid advancement of artificial intelligence, machine vision technology has found widespread application in the field of agriculture[13,14]. The female silkworm pupae have an X-shaped line on their frontal tails, while the male silkworm pupae show a dot[15,16]. Based on the different gonad features of male and female silkworm pupae, we conducted research on online sex identification using machine vision technology. However, the posture of silkworm pupae was random when transported to the image acquisition area, making it impossible to obtain a frontal image at one time. Therefore, a rotational mechanism was employed to rotate each silkworm pupa approximately 360° while capturing eight images simultaneously, obtaining multiple pose images of the same individual. As illustrated in Figure 1, it is obvious that accurate gender identification of silkworm pupae is very difficult from the dorsal or lateral sides. Therefore, accurately selecting the frontal

Received date: 2023-07-09 **Accepted date:** 2025-12-12

Biographies: Feng Guo, PhD, Lecturer, research interest: machine vision, Email: gfchengdu@163.com; **Jing Li**, PhD, Lecturer, research interest: intelligent detection, Email: lijing@mail.xhu.edu.cn; **Wei Qin**, MS, research interest: intelligent agricultural equipment, Email: 1523670619@qq.com; **Chunjiang Zhao**, PhD, Researcher, research interest: smart agriculture, Email: zhaocj@nercita.org.cn.

***Corresponding author: Guanglin Li**, PhD, Professor, research interest: intelligent agricultural equipment. College of Engineering and Technology, Southwest University, Chongqing 400715, China. Tel: +86-23-68250807, Email: liguanglin@swu.edu.cn.

image from multi-pose images of silkworm pupae is critical for machine vision-based sex identification.



Note: The first row shows female silkworm pupae, the second row shows male silkworm pupae, and the red areas indicate the gonad features.

Figure 1 Images of female and male silkworm pupae in different postures

To address this issue and improve the accuracy of male and female identification of silkworm pupae, we proposed SPNet-GS (Silkworm Pupae Network for Gonad Selection), a novel network structure based on Xception. In the proposed method, we used the combination of large kernel convolution and dilated convolution to capture the far and near relationship between pixels, so as to enhance feature extraction. The proposed SPNet-GS not only has good generalization ability, but also has fewer parameters and faster calculation speed, which can fully meet the online sex identification requirements of silkworm pupae.

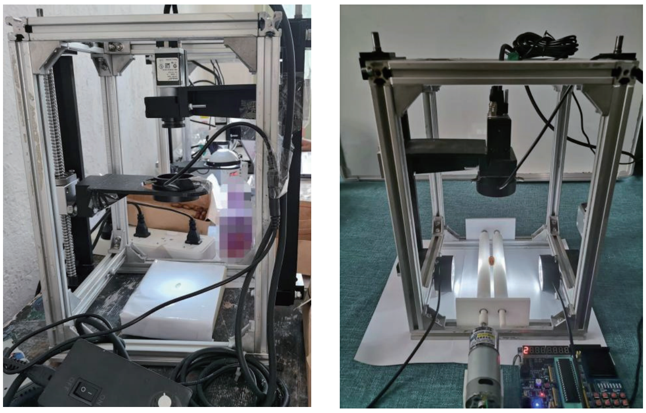
2 Materials and methods

2.1 Datasets

The silkworm pupae image datasets included static collection and online collection. The statically acquired images were for model training and verification, and online acquisition images were used for model testing. The collection locations were the Chongqing Sericulture Science and Technology Research Institute and Southwest University in China. The static images of silkworm pupae were collected using a Basler acA1300-30gc industrial camera, and the captured image size was 1280×960 pixels. The online silkworm pupae images were collected using a Basler acA1920-50gm industrial camera, and the captured image size was 1920×1200 pixels. In the static silkworm pupae image collection, experimenters manually placed samples in the image collection area of the device, then manually changed the posture of the pupae and collected images of the front, back, and side of each silkworm pupae. The online acquisition was after the silkworm pupae entered the camera's field of view, then the rotation structure drove the silkworm pupae to rotate for about 360°, and a total of eight images were collected for each sample in 1.6 s. The silkworm pupae image acquisition device is shown in Figure 2.

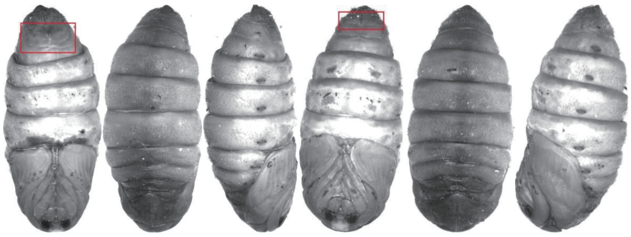
In this study, static clear silkworm pupae datasets were used for model training and validation, which included 3827 frontal, 3479 dorsal, and 3515 lateral images, and were randomly divided into a training set and a validation set according to the ratio of 9:1. Subsequently, both sets were expanded three times through random rotation and flipping. To evaluate the generalization ability of the model, the test set was composed of 16 additional silkworm varieties, with a total of 6544 images captured online. Since there are a large number of white background pixels in the image, the contours of silkworm pupae were segmented employing the thresholding and the findContours function of the OpenCV library[17] to better extract pupal features. The segmented images of different postures of female and male silkworm pupae are shown in Figure 3. The principle of threshold segmentation is defined as follows: src(x, y) represents the pixel value at position (x, y) in the input image, and dst(x, y) represents the pixel value at position (x, y) after threshold processing.

$$\mathrm{dst}(x,y)=\begin{cases}0 & \text{if } \mathrm{src}(x,y)>220\\ 255 & \text{otherwise}\end{cases} \tag{1}$$



a. Static image acquisition device b. Rotating imaging mechanism

Figure 2 Silkworm pupae image acquisition devices



Note: From left to right, the front, back, and side of female and male silkworm pupae are shown in sequence, in which the red box shows the gonad characteristics.

Figure 3 Images of female and male silkworm pupae in different postures

2.2 The proposed CNN architecture: SPNet-GS

Xception[18], a powerful convolutional neural network architecture, is built upon depthwise separable convolutions. Therefore, we proposed SPNet-GS, an Xception-based model for online front silkworm pupae image selection. However, Xception relies on stacking multiple 3×3 convolutions for feature extraction, which not only escalates computational load but also fails to effectively model long-range dependencies. To address this limitation, SPNet-GS discards the extensive stacking of 3×3 convolutional operations and instead adopts a combination of large convolutional kernels and dilated convolutions to capture both long- and short-range dependencies, thereby forming a more efficient model with enhanced generalization capability. The SPNet-GS model structure is shown in Figure 4. BN and SConv are Batch Normalization and Depthwise Separable Convolution[19], respectively. For the input image, we first use two convolution operations with a step size of 2 for feature extraction and downsampling. Then the Large Kernel Convolution and Dilated Convolution (LKCDC) module is employed for key feature extraction, as shown in Figure 5. After two LKCDC modules, SConv operation is used for feature optimization with a max pooling layer for downsampling. Meanwhile, the residual connection[20] is adopted to avoid the degradation of network. Finally, the classification results (front, back, side) are output through global average pooling, full connection, and logistic regression operations.

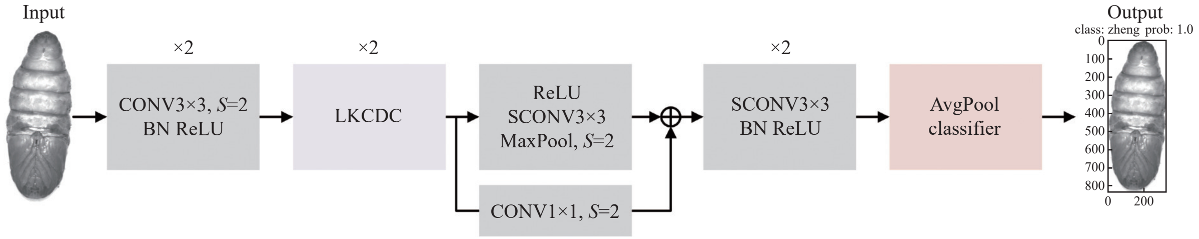


Figure 4 SPNet-GS structure

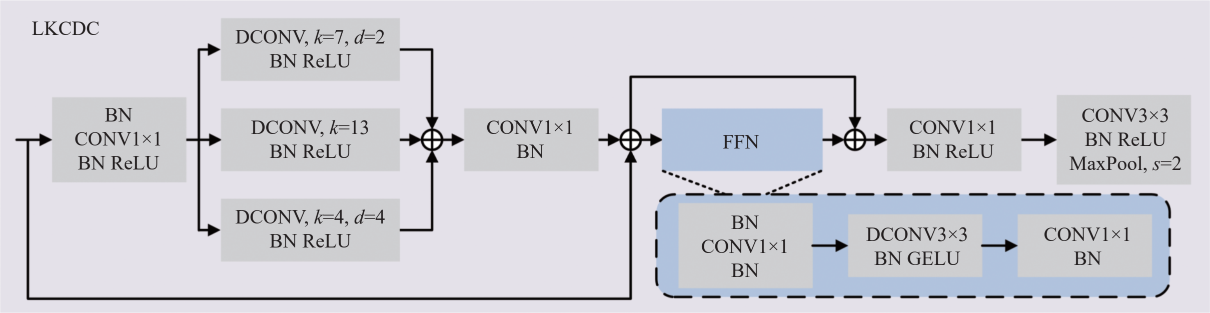


Figure 5 Illustration of the proposed LKCDC module

2.2.1 Depthwise separable convolution

Depthwise separable convolution consists of depthwise convolution and pointwise convolution. The depthwise separable convolution is different from standard convolution in that the network no longer aims to realize the joint mapping of channel and spatial correlation, but decouples the spatial and channel correlation and maps them separately to achieve better results. Depthwise convolution achieves spatial fusion and pointwise convolution achieves channel fusion. Depthwise separable convolution can effectively expand the network width, which can not only extract enough high-dimensional features, but also reduce the total number of parameters and improve the network performance.

Given that the input size of the feature map is set to $H_i×W_i×C_i$ and the convolution kernel size is $K×K$, then the output size of the feature map is $H_o×W_o×C_o$. For the standard convolution, the computation is $K×K×C_i×H_o×W_o×C_o$. For the depthwise separable convolution, the computation of depthwise convolution is $K×K×C_i×H_o×W_o$, the computation of pointwise convolution is $C_i×H_o×W_o×C_o$, and then the total computation is $K×K×C_i×H_o×W_o+C_i×H_o×W_o×C_o$. Therefore, the parameter ratio between the depthwise separable convolution and the standard convolution is shown in Equation (2).

$$\frac{K\times K\times C_i\times H_o\times W_o+C_i\times H_o\times W_o\times C_o}{K\times K\times C_i\times H_o\times W_o\times C_o}=\frac{1}{C_o}+\frac{1}{K^2} \qquad (2)$$

When the output feature map channel is unchanged, the number of depthwise separable convolution parameters used decreases more obviously with the increase of convolution kernel size.

2.2.2 Dilated convolution

Dilated convolution is simply the process of enlarging the convolution kernel by adding spaces (zeros) between the elements of the convolution kernel. These spaces we call the dilation rate[21]. The main advantages of dilated convolution are summarized as follows: 1) The dilated convolution has a low computational cost while increasing the receptive field. Assuming that the size of the convolution kernel is $K×K$ and the dilation rate is R, the corresponding receptive field of the convolution kernel is shown in Equation (3). For example, when the size of the convolution kernel and the dilation rate are set as 3×3 and 2, respectively, then the corresponding receptive field of the convolution kernel is 5×5. Now the calculation is the same, but the receptive field is enlarged compared with the standard convolution 3×3. 2) Dilated convolution can obtain multi-scale context information. When multiple dilated convolution kernels with different dilation rates are superimposed, different receptive fields will bring multi-scale information, which is very important for feature extraction.

$$K_1=(K+(K-1)\times(R-1))\times(K+(K-1)\times(R-1)) \qquad (3)$$

2.2.3 LKCDC block

To achieve a larger receptive field for better capturing long-range dependencies between pixels, the LKCDC module was proposed. Through a combination of large-kernel and dilated convolutions, LKCDC captures both short- and long-range dependencies between pixels, thereby gaining broader contextual information, as shown in Figure 5, where DConv and d stand for depthwise convolution and dilation, respectively. In the LKCDC block, the large convolution kernel size is 13×13, while the dilated convolutional kernels use sizes of 7×7 and 4×4 with dilation rates of 2 and 4, respectively. Then, we use element-wise addition for spatial information fusion and 1×1 convolution for channel information fusion. Meanwhile, we also introduce the Feed-Forward Network (FFN)[22] to fuse features. In the FFN, a 1×1 convolution is first applied to improve the network width. Subsequently, a 3×3 depthwise convolution performs spatial information fusion. Then, a GELU[23] activation and another 1×1 convolution are used to reduce the network width and enable channel information fusion. Finally, we use two convolution layers and one maximum pooling layer to increase the channel dimension and downsampling.

2.3 Evaluation metrics

To evaluate the performance of our model SPNet-GS, Accuracy, $F1$-score, Parameters, FLOPs, and Model Size were used as model evaluation metrics, as follows:

$$Accuracy=\frac{N_g}{All_g} \qquad (4)$$

$$F1=\frac{2PR}{P+R} \tag{5}$$

$$P=\frac{\text{TP}}{\text{TP+FP}} \tag{6}$$

$$R=\frac{\text{TP}}{\text{TP+FN}} \tag{7}$$

In this study, N_g represents the number of correctly selected frontal images (only one optimal front image is selected from eight images of the same silkworm pupa). All_g represents the total number of silkworm pupae. TP (True Positive) means the front of the silkworm pupa is present in the image, and the model correctly classifies it as the frontal side. FP (False Positive) refers to the fact that another side of the silkworm pupa is present in the image, and the model classifies it as the frontal side. FN (False Negative) indicates that the frontal side of silkworm pupa is present in the image, and the model classifies it as another side. *P* (Precision) represents the proportion of all the silkworm pupae images predicted to be frontal that are actually frontal. *R* (Recall) stands for the proportion of frontal images that are correctly predicted in all frontal images of silkworm pupae. The *F*1-score reflects the model' s capacity in Precision and Recall, and the larger the *F*1-score (its value ranges from 0 to 1), the better the model.

Parameters are the total number of parameters to be trained in model training, and the smaller the number of parameters, the lighter the model. FLOPs (Floating Point Operations per second) reflect the complexity of the model. Model Size is a measure of model portability.

2.4 Experiment settings

All models in this study are based on pytorch[24] deep learning framework, and the test environment is Win10 system, where the CPU (Intel(R) Xeon(R) Gold 6242R) and the GPU (NVIDIA Tesla V100) are used. The learning rate, epochs, batch size, and optimizer are 0.0001, 130, 16, and Adam[25], respectively.

3 Results and discussion

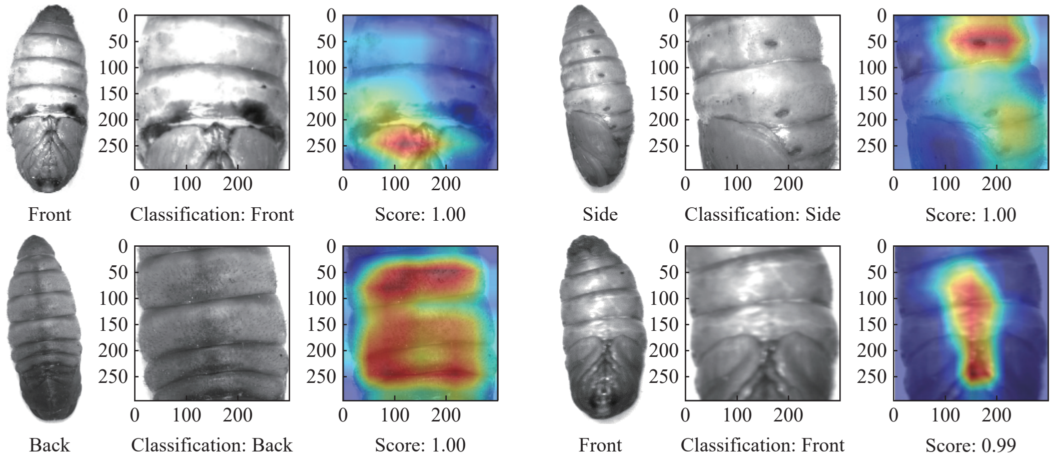
3.1 Front image identification result of silkworm pupae

To evaluate the performance of the SPNet-GS model in this study, AlexNet[26], VGG16[27], DenseNet-121[28], ResNet-50, MobileNetV2[29], GoogLeNet[30], Xception, and ViT[31] models were trained under the same experimental environment and dataset for comparison. Table 1 shows a performance comparison on the test set. In terms of accuracy, the constructed SPNet-GS model achieves a strong result of 98.41%, which is 30.81%, 24.69%, 3.91%, 6.23%, 3.06%, 4.52%, 6.72%, and 16.75% higher than that of AlexNet, VGG16, DenseNet-121, ResNet-50, MobileNetV2, GoogLeNet, Xception, and ViT, respectively. By analyzing the sorting results of AlexNet, VGG16, and ViT, it was found that they all exhibited a significant number of misclassifications, leading to lower accuracy. Additionally, the limited dataset size was an important reason for the suboptimal performance of the transformer-based ViT model. Regarding the *F*1-score, the SPNet-GS model also performs a strong result of 99.02%, with an improvement achieved about 18.35%, 14.15%, 2.65%, 3.51%, 2.13%, 2.23%, 4.02%, and 9.12% over benchmark models, respectively, indicating that SPNet-GS has excellent recall and precision capabilities for frontal images of silkworm pupae. Meanwhile, the SPNet-GS model has over 8, 80, 4, 14, 1, 3, 12, and 51 times fewer parameters than the reference models, respectively. In the FLOPs, while the SPNet-GS model is not the most compact, its accuracy and *F*1-score are both higher than those of MobileNetV2 and AlexNet. Simultaneously, the SPNet-GS model size is the smallest compared to other models, only 6.46 M. To sum up, the SPNet-GS model combines high recognition accuracy with low parameters and fast inference speed (approximately 0.03 s per image). Crucially, its performance is not limited by the silkworm pupae species, making it fully suitable for the online selection of the silkworm pupa front image.

Table 1 Performance comparison of different models on the test set

Model	Accuracy/%	*F*1/%	Parameters/M	FLOPs/G	Model Size/M
AlexNet	67.60	80.67	14.59	0.31	55.60
VGG16	73.72	84.87	134.27	15.47	512.00
DenseNet-121	94.50	96.37	6.95	4.90	27.00
ResNet-50	92.18	95.51	23.51	7.57	89.90
MobileNetV2	95.35	96.89	2.23	0.60	8.72
GoogLeNet	93.89	96.79	5.97	2.58	39.30
Xception	91.69	95.00	20.81	8.43	79.60
ViT	81.66	89.90	85.65	16.86	327
SPNet-GS	98.41	99.02	1.67	2.34	6.46

Class Activation Mapping (CAM) is a commonly used visualization tool for distinguishing regions (attention maps). We use Grad-CAM[32] to visualize the attentions on the silkworm pupa image by the SPNet-GS model. Additionally, validation was conducted on specific challenging frontal images that are prone to misclassification, including those with motion blur, distortion, exposure issues, pseudo-frontal views, and speckle patterns. Figure 6 shows partial visualization results. The images from top to bottom and left to right respectively show: frontal, dorsal, lateral, blurred, distorted, underexposed, pseudo-frontal, and speckled images of silkworm pupae. Visualization results demonstrate that the model can effectively capture key features for classification in both clear and special-case images, which is of significant importance for the accurate subsequent identification of silkworm pupae gender.



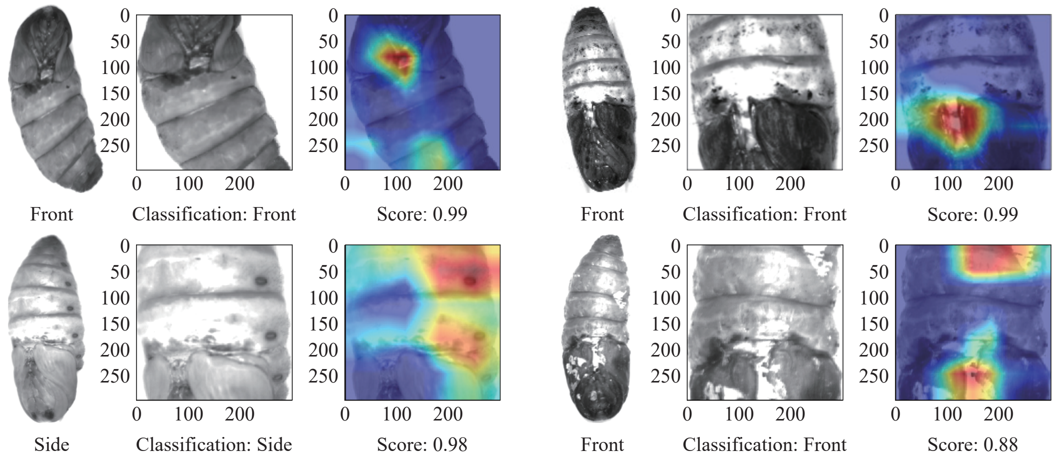


Figure 6 Partial visualization results

3.2 Ablation study

To evaluate the impact of key components in the proposed SPNet-GS model for online front image selection of silkworm pupae, a series of ablation studies were conducted, with results summarized in Table 2.

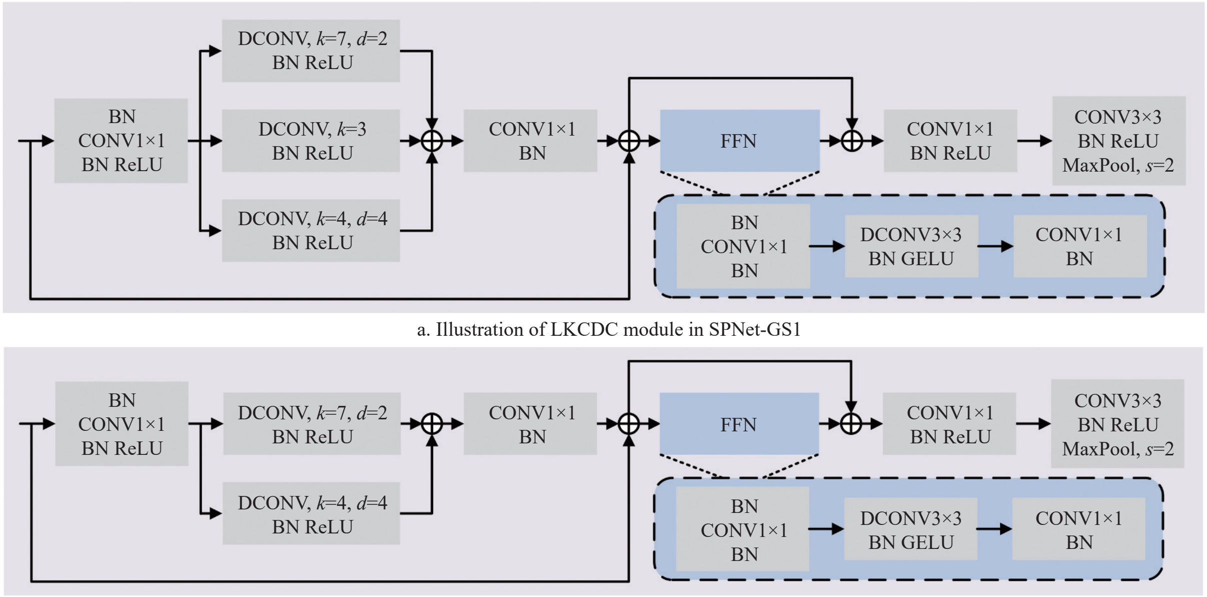
Table 2 Results of the ablation study

Model	Accuracy/%	*F*1/%	Parameters/M	FLOPs/G
SPNet-GS1	95.48	97.32	1.64	2.26
SPNet-GS2	88.75	93.68	1.63	2.25
SPNet-GS3	93.52	96.04	1.65	2.30
SPNet-GS4	73.11	84.47	1.64	2.27
SPNet-GS5	94.50	96.51	1.64	2.28
SPNet-GS6	92.42	94.92	1.65	2.29
SPNet-GS7	94.13	96.73	1.66	2.32
SPNet-GS8	95.35	97.50	1.68	2.37
SPNet-GS	98.41	99.02	1.67	2.34

Evaluating the superiority of large-kernel convolution in the LKCDC module: As shown in Figure 7a, the large-kernel convolution is replaced with a standard 3×3 convolution, called SPNet-GS1. Although this modification reduces both parameters and FLOPs compared to SPNet-GS, it decreases test accuracy by 2.93%. This demonstrates that large-kernel convolution effectively captures long-range dependencies between pixels, which is crucial for extracting key features. Furthermore, as depicted in Figure 7b, the large-kernel convolution is completely removed, named SPNet-GS2. While this reduction decreases both parameters and FLOPs, it also reduces test accuracy by 9.66%. This indicates that capturing only long-range dependencies while ignoring short-range interactions degrades the performance of the online frontal image selection model.

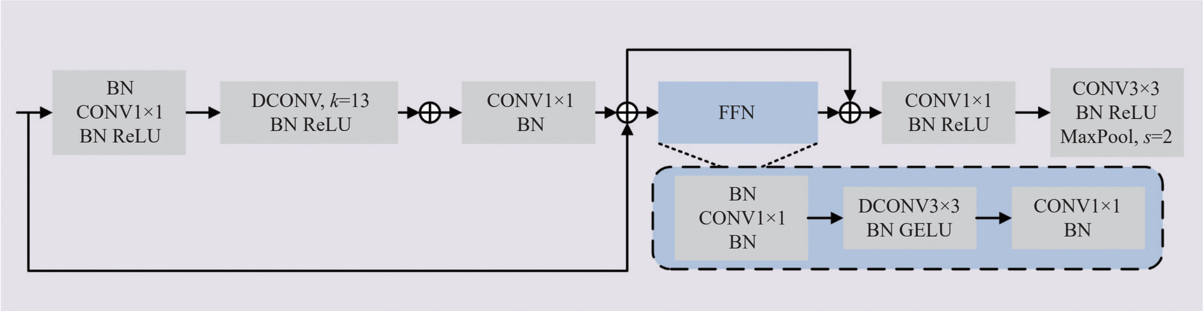
Assessing the combined effect of dilated and large-kernel convolution in the LKCDC module: As presented in Figure 7c, the dilated convolution is removed, dubbed SPNet-GS3. Although this modification similarly reduces parameters and FLOPs, it decreases test accuracy by 4.89%. This confirms that the synergistic integration of long-range dependencies across pixels and feature diversity serves as an effective mechanism for boosting model performance.

Determining the optimal kernel size in the LKCDC module: The large-kernel convolution is configured with sizes of 5×5, 7×7, 9×9, 11×11, and 15×15, with corresponding models named SPNet-GS4 through SPNet-GS8. As shown in Table 2, the performance results indicate that larger kernel sizes do not consistently yield better accuracy. Therefore, a 13×13 kernel is ultimately selected as the optimal configuration.



a. Illustration of LKCDC module in SPNet-GS1

b. Illustration of LKCDC module in SPNet-GS2



c. Illustration of LKCDC module in SPNet-GS3

Figure 7 Ablation on the different designs of LKCDC module

3.3 Sex identification result of silkworm pupae

To investigate the superiority of sex identification based on gonadal characteristics, the ResNet50 model was separately trained using multi-pose images, front images, and gonad images for sex classification. The strategy we adopt for sex identification without front selection of silkworm pupae is shown in Equation (8), where x_i and F are the input image and the ResNet50 network. First, each image of the same silkworm pupa is classified by gender, and the prediction score is recorded. A positive score indicates female, while a negative score indicates male. The final sex classification result is obtained by summing the gender identification scores from all images.

$$S_{\text{core}} = \sum_{i=1}^{n} F(x_i) \tag{8}$$

The experimental results demonstrate that the model trained with multiple orientation images of silkworm pupae achieves a test accuracy of only 74.68%. This is because gender distinction is difficult from lateral and dorsal views, and females in lateral views are often misclassified as males. In comparison, using only front images for training improved the test accuracy to 84.68%, indicating that significant gender-related differences are observable in the frontal view. However, the accuracy of male and female identification still cannot meet the breeding requirements. In contrast, the model trained solely on gonad images achieves a further improved accuracy of 94.58%, outperforming both the frontal and multi-orientation image approaches. This superiority stems from the effective elimination of irrelevant background interference, enabling the model to focus specifically on gonad characteristics. Therefore, study for front image selection of silkworm pupae is the foundation for machine vision-based sex sorting and holds significant practical value for engineering applications.

4 Conclusions

In this study, an online frontal image selection model of silkworm pupae is introduced based on large kernel convolution, dilated convolution, and depthwise separable convolution, which is both lightweight and accurate for choosing the frontal image of silkworm pupae. Experimental results demonstrate that the proposed SPNet-GS model achieves average accuracy and $F1$-scores of 98.41% and 99.02%, respectively, outperforming benchmark methods and meeting the requirements for frontal image selection in silkworm pupae. Meanwhile, compared to using multi-pose silkworm pupae images for gender identification, the selected frontal and gonad images achieve higher accuracy rates by 10.00% and 19.90%, respectively, demonstrating the advantage of the proposed selection strategy. In summary, the frontal image selection strategy proposed in this paper is of crucial significance for enhancing the accuracy of online gender identification in silkworm pupae.

Acknowledgments

This study was financially supported by the National Natural Science Foundation of China (Grant No. 31971782) and the Postgraduate Research Innovation Project of Southwest University (Grant No. SWUB23055).

[References]

[1] Ma Y, Xu Y, Yan H, Zhang G. On-line identification of silkworm pupae gender by short-wavelength near infrared spectroscopy and pattern recognition technology. Journal of Near Infrared Spectroscopy, 2021; 29(4): 207–215.

[2] He H, Zhu S, She L, Chang X, Wang Y, Zeng D, et al. Integrated analysis of machine learning and deep learning in silkworm pupae (*Bombyx mori*) species and sex identification. Animals, 2023; 13(23): 3612. doi: 10.3390/ani13233612.

[3] Sumriddetchkajorn S, Kamtongdee C. Optical penetration-based silkworm pupa gender sensor structure. Applied Optics, 2012; 51(4): 408–412.

[4] Kamtongdee C, Sumriddetchkajorn S, Sa-ngiamsak C. Feasibility study of silkworm pupa sex identification with pattern matching. Computers and Electronics in Agriculture, 2013; 95: 31–37.

[5] Sumriddetchkajorn S, Kamtongdee C, Chanhorm S. Fault-tolerant optical-penetration-based silkworm gender identification. Computers and Electronics in Agriculture, 2015; 119: 201–208.

[6] Kamtongdee C, Sumriddetchkajorn S, Chanhorm S, Kaewhom W. Noise reduction and accuracy improvement in optical penetration-based silkworm gender identification. Applied Optics, 2015; 54(7): 1844–1851.

[7] Liu C, Ren Z H, Wang H Z, Yang P Q, Zhang X L. Analysis on gender of silkworms by MRI technology. In: 2008 International Conference on BioMedical Engineering and Informatics, IEEE: Sanya, China, 2008; pp.8-12. doi: 10.1109/BMEI.2008.49.

[8] Cai J, Yuan L, Liu B, Sun L. Nondestructive gender identification of silkworm cocoons using X-ray imaging with multivariate data analysis. Analytical Methods, 2014; 6(18): 7224–7233.

[9] Raj A N J, Sundaram R, Mahesh V G V, Zhuang Z M, Simeone A. A multi-sensor system for silkworm cocoon gender classification via image processing and support vector machine. Sensors, 2019; 19(12): 2656.

[10] Lin X, Zhuang Y, Tao D, Li G L, Yang X D, Song J, et al. The model updating based on near infrared spectroscopy for the sex identification of silkworm pupae from different varieties by a semi-supervised learning with pre-labeling method. Spectroscopy Letters, 2019; 52(10): 642–652.

[11] Tao D, Wang Z R, Li G L, Xie L. Sex determination of silkworm pupae using VIS-NIR hyperspectral imaging combined with chemometrics. Spectrochimica Acta Part A, Molecular and Biomolecular Spectroscopy, 2019; 208: 7–12.

[12] Qiu G Y, Tao D, Xiao Q, Li G L. Simultaneous sex and species classification of silkworm pupae by NIR spectroscopy combined with chemometric analysis. Journal of the Science of Food and Agriculture, 2021; 101(4): 1323–1330.

[13] Hu C H, Shi Z F, Wei H L, Hu X D, Xie Y N, Li P P. Automatic detection of pecan fruits based on Faster RCNN with FPN in orchard. Int J Agric &

Biol Eng, 2022; 15(6): 189–196.

[14] Wang T, Chen B, Zhang Z, Li H, Zhang M. Applications of machine vision in agricultural robot navigation: A review. Computers and Electronics in Agriculture, 2022; 198: 107085.

[15] Tao D, Wang Z R, Li G L, Qiu G Y. Radon transform-based motion blurred silkworm pupa image restoration. Int J Agric & Biol Eng, 2019; 12(2): 152–159.

[16] Guo F, He F, Tao D, Li G. Automatic exposure correction algorithm for online silkworm pupae (*Bombyx mori*) sex classification. Computers and Electronics in Agriculture, 2022; 198: 107108.

[17] OpenCV. Achieve: https://opencv.org. Accessed on [2023-05-05].

[18] Chollet F. Xception: Deep learning with depthwise separable convolutions. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE: Honolulu, HI, USA, 2017; pp.1800–1807. doi: 10.1109/CVPR.2017.195.

[19] Howard A G, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv, 2017; arXiv: 1704.04861. doi: 10.48550/arXiv.1704.04861.

[20] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE: Las Vegas, 2016; pp.770–778. doi: 10.1109/CVPR.2016.90.

[21] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv, 2016. doi: 10.48550/arXiv.1511.07122.

[22] Ding X, Zhang X, Han J, Ding G. Scaling up your kernels to 31×31: Revisiting large kernel design in CNNs. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE: New Orleans, LA, USA, 2022; pp.11953–11965. doi: 10.1109/CVPR52688.2022.01166.

[23] Hendrycks D, Gimpel K. Gaussian error linear units (GELUs). arXiv preprint arXiv: 2020. doi: 10.48550/arXiv.1606.08415.

[24] PyTorch. Achieve: https://pytorch.org. Accessed on [2023-05-04].

[25] Kingma D P, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv, 2015. doi: 10.48550/arXiv.1412.6980.

[26] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks. Communications of the ACM, 2017; 60(6): 84–90.

[27] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint, 2014. doi: 10.48550/arXiv.1409.1556

[28] Huang G, Liu Z, Maaten L V D, Weinberger K Q. Densely connected convolutional networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017; pp.2261–2269. doi: 10.1109/CVPR.2017.243.

[29] Sandler M, Howard A, Zhu M, Zhmoginov A, Chen L C. MobileNetV2: Inverted residuals and linear bottlenecks. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE: Salt Lake City, USA, 2018; pp.4510–4520. doi: 10.1109/CVPR.2018.00474.

[30] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE: Boston, 2015; pp.1–9. doi: 10.1109/CVPR.2015.7298594.

[31] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16×16 words: Transformers for image recognition at scale. arXiv preprint, 2021. doi: 10.48550/arXiv.2010.11929.

[32] Selvaraju R R, Das A, Vedantam R, Cogswell M, Parikh D, Batra D. Grad-CAM: Why did you say that? Visual explanations from deep networks via gradient-based localization. arXiv preprint arXiv: 2016. doi: 10.48550/arXiv.1610.02391.