

## Image Recognition of Modern Agricultural Fruit Maturity Based on Internet of Things

Dan Chen\*, Jiali Tang, Haixu Xi, Xiaorong Zhao

School of Computer Engineering, Jiangsu University of Technology, Changzhou 213001, China

Corresponding Author Email: [2003500009@jsut.edu.cn](mailto:2003500009@jsut.edu.cn)



<https://doi.org/10.18280/ts.380435>

### ABSTRACT

**Received:** 5 February 2021

**Accepted:** 12 June 2021

#### Keywords:

*internet of things (IoT), image processing, modern agriculture, fruit maturity*

The accurate judgement of fruit maturity is significant for modern agriculture. At present, few scholars have monitored and recognized fruit maturity based on the Internet of things (IoT) and image recognition technology. Therefore, this paper explores the image recognition of fruit maturity in the context of agricultural Internet of things (IoT). Firstly, the single shot multi-box detection (SSD) algorithm was improved for fruit recognition and positioning, and used to determine the size and position the fruits to be recognized. Next, an image fusion algorithm was designed based on improved Laplacian pyramid, which effectively compresses the large fruit monitoring images shot in the same scene. The proposed algorithm was proved feasible and effective through experiments.

## 1. INTRODUCTION

The fruit picking time mainly depends on the maturity of the fruit [1-3]. Fruits of different usages should be picked at different stages of maturity. The picking time needs to be properly advanced or delayed, according to the specific usage [4-7]. Due to the limitations of warehousing and logistics, fruits should not be picked when they are too mature or raw. The accurate judgement of fruit maturity is significant for modern agriculture [8, 9]. The Internet of things (IoT) helps to acquire information from fruit detection terminals in orchards, and evaluate the fruit maturity. Meanwhile, image recognition can determine the size and position fruits [10-14]. The advancement of the two techniques makes the monitoring of fruit growth and insect attack more intelligent and automatic in modern agriculture.

Manual measurement/grading often fails on oil palm fruits. Syaifuddin et al. [15] collected palm fruit colors based on gray-level co-occurrence matrices (GLCMs), extracted the color features of palm fruits through regression, and thus detected the maturity of these fruits. Zheng et al. [16] adopted artificial neural network (ANN) to recognize the maturity of citrus fruits in Xinhui, China, and experimentally verified that the proposed algorithm is a non-contact, low-cost, easy-to-achieve, and high-performance method for remote monitoring. The algorithm facilitates the record of citrus growth and forecast of market sales, contributing greatly to agricultural production.

Tomato maturity is closely related to its surface color. Dimatira et al. [17] treated color, size, and shape as the metrics of tomato maturity, evaluated the maturity level of tomatoes by machine vision, and divided the colors of the fruits into six phases: green phase, crushing phase, transition phase, pink phase, and light red phase. To realize real-time accurate recognition of the maturity of fruits and vegetables, Fadhel and Al-Shamma [18] proposed a field programmable gate array (FPGA) as the parallel hardware structure, aiming to reduce the high time cost of color thresholding and k-means clustering (KMC). Xiao et al. [19] predicted the maturation stage of

tomato fruits according to surface color, and managed to forecast the surface color variation based on temperature conditions. Cai and Zhao [20] developed a mature fruit recognition technique in natural scenes, which compares the color models of hue-saturation-intensity (HSI) system. The technique can fuse tone and saturation data into a fused image, in order to eliminate the effects of soil and sky backgrounds on recognition.

Overall, the fruit maturity recognition techniques at home and abroad are mostly theoretical. Few of them have been applied in practice. Not many scholars have combined the IoT with image recognition for fruit maturing monitoring and recognition in complex orchid environments. To solve the problem, this paper explores the image recognition of fruit maturity in the context of agricultural IoT. Section 2 improves the single shot multi-box detection (SSD) algorithm for fruit recognition and positioning, and completes the determination of the size and position the fruits to be recognized. Section 3 designs an image fusion algorithm based on improved Laplacian pyramid, which effectively compresses the large fruit monitoring images shot in the same scene. The proposed algorithm was proved feasible and effective through experiments.

## 2. IMPROVED SSD FOR FRUIT RECOGNITION AND POSITIONING

### 2.1 SSD model construction

The extensive research and application of machine vision and IoT effectively improve the efficiency of various tasks on modern agricultural product picking lines, including fruit type recognition, maturity identification, fast picking, rapid sorting, fast packing, etc. In the automatic picking scenes of agricultural products, there are a huge number of various fruits, and the picking environment is extremely complex. These pose an immense challenge to the size and maturity detection of fruits.

The automation of agricultural picking is premised on the effective recognition of fruits and their maturity in complex background. This paper chooses to construct an SSD target detection model, in order to determine the size and position of the fruits to be recognized. The SSD model, which is based on feature pyramid network (FPN), is an end-to-end target detection framework, capable of detecting fruits of various sizes. The model regards feature map cells of different resolutions as priori boxes with different sizes and aspect ratios, and predicts the targets' bounding boxes based on the priori boxes.

The SSD model consists of a feature layer  $CL_{4,3}$  and a series of additional convolutional layers. The feature layer is obtained through feature extraction of VGG16. The convolutional layers cover feature layers  $CL_7$ ,  $CL_{8-2}$ ,  $CL_{9-2}$ ,  $CL_{10-2}$ ,  $CL_{11-2}$ .

The SSD model can calculate the confidence for the position of the detection frame and the class of fruit size corresponding to each prior box. Let  $p=(p^{da}, p^{db}, p^q, p^s)$  be the position of a prior box;  $r=(r^{da}, r^{db}, r^q, r^s)$  be a bounding box. Then, the relationship of the predicted bounding box  $k=(k^{da}, k^{db}, k^q, k^s)$  for the target fruit with  $r$  and  $p$  can be expressed as:

$$k^{da} = (r^{da} - p^{da}) / p^q \quad (1)$$

$$k^{db} = (r^{db} - p^{db}) / p^s \quad (2)$$

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \quad (3)$$

$$k^s = \log(r^s / p^s) \quad (4)$$

To predict the bounding box of a fruit, the above encoding process needs to be reversed, that is, the bounding box needs to be decoded.

Let  $d$  be the number of predicted classes for the target fruit;  $N \times M$  be the size of the feature map;  $NM$  be the number of cells in the feature map;  $l$  be the number of prior boxes corresponding to each cell. Then, each unit will output  $(d+4)l$  predicted bounding boxes, and the feature map will output a total of  $(d+4)lNM$  predicted bounding boxes.

In this paper, the feature map for prediction has  $N$  feature layers of different sizes. Let  $E_{min}$  and  $E_{max}$  be the minimum and maximum dimension scales of the feature layers, respectively. The feature layers of different sizes generate prior boxes with different aspect ratios. The ratio of each prior box to the original image can be calculated by:

$$E_l = E_{min} + \frac{E_{max} - E_{min}}{N - 1}(l - 1), \quad l \in [1, N] \quad (5)$$

Let  $q_l^\beta$  and  $g_l^\beta$  be the width and height of the prior box, respectively;  $\beta_s$  be the default aspect ratio of the box. Then, the width of the prior box can be calculated by:

$$q_l^\beta = E_l \sqrt{\beta_s} \quad (6)$$

The height of the prior box can be calculated by:

$$g_l^\beta = E_l / \sqrt{\beta_s} \quad (7)$$

Let  $m$  be the number of positive samples for the matching between the target box and predicted box;  $a \in \{1, 0\}$  be the binary function reflecting whether the two boxes match each other (if yes,  $a=1$ ; otherwise,  $a=0$ );  $h$  be the position of actual target box;  $\zeta$  be the weight coefficient about the proportional relationship of position loss and confidence loss. In this paper, the sum of the position error  $K_L$  and the class confidence error  $K_C$  of the detection frame is defined as the loss function of SSD model:

$$Loss(a, d, k, h) = \frac{1}{m} (K_C(a, d) + \zeta K_L(a, k, h)) \quad (8)$$

Because the model adopts the cross-entropy loss function, the output probabilities of different fruit classes need to be calculated through softmax regression of class confidence. Let  $a_{ijo}$  be the indicator of whether the  $i$ -th prior box matches the  $j$ -th actual box in class  $O$ ;  $\dot{d}_i^o$  be the confidence of the  $i$ -th prior box belonging to class  $O$ ;  $\dot{d}_i^o$  and  $\dot{d}_i^0$  be the confidence of softmax regression and background of the  $i$ -th prior box, respectively;  $ZS$  and  $FS$  be the set of positive samples, and the set of negative samples, respectively. Then, the class confidence error  $K_C$  can be calculated by:

$$K_C(a, d) = -\sum_{i \in ZS} a_{ij}^o \log(\dot{d}_i^o) - \sum_{i \in FS} \log(\dot{d}_i^0) \quad (9)$$

$\dot{d}_i^o$  can be calculated by:

$$\dot{d}_i^o = \frac{\exp(\dot{d}_i^o)}{\sum_o \exp(\dot{d}_i^o)} \quad (10)$$

The position error of the detection frame can be characterized by *Smooth-L1* loss function:

$$K_{smooth-L_1}(a) = \begin{cases} 0.5a^2 & |a| < 1 \\ |a| - 0.5 & otherwise \end{cases} \quad (11)$$

Let  $p_i^{da}$ ,  $p_i^{db}$ ,  $p_i^q$ , and  $p_i^s$  be the central coordinates, width, and height of the  $i$ -th prior box, respectively;  $h_j^{da}$ ,  $h_j^{db}$ ,  $h_j^q$ , and  $h_j^s$  be the central coordinates, width, and height of the actual box of the target fruit, respectively. The deviations of the actual box from the prior box can be calculated by:

$$\dot{h}_j^{da} = (h_j^{da} - p_i^{da}) / p_i^q \quad (12)$$

$$\dot{h}_j^{db} = (h_j^{db} - p_i^{db}) / p_i^s \quad (13)$$

$$\dot{h}_j^q = \log(h_j^q / p_i^q) \quad (14)$$

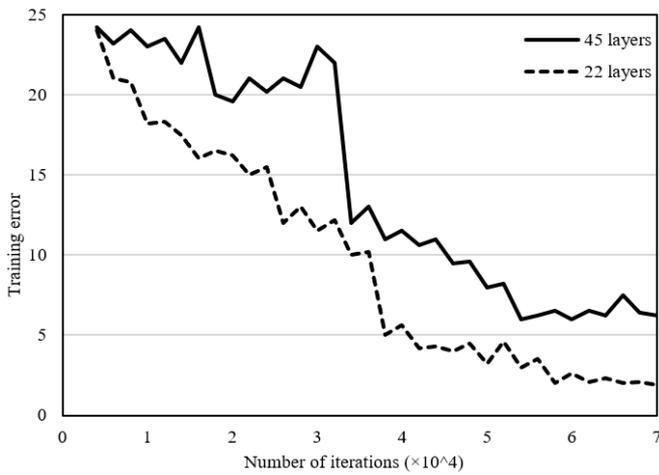
$$\dot{h}_j^s = \log(h_j^s / p_i^s) \quad (15)$$

The deviations obtained by formulas (12)-(15) can be transformed to derive the position of the prior box.

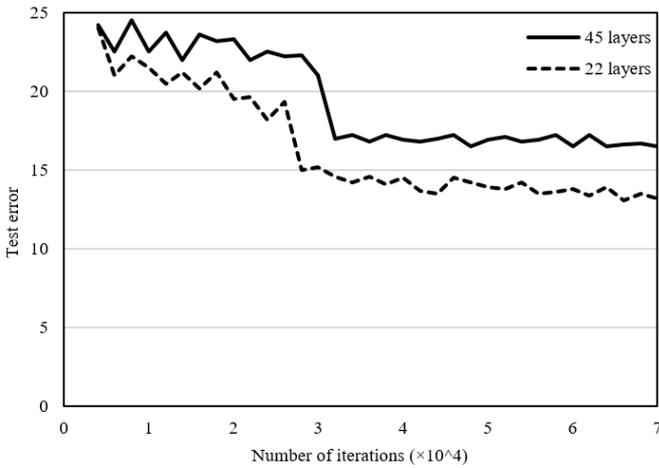
## 2.2 SSD model construction based on residual network

Before introducing the residual network, SSD target

detection model could contain 22 layers at the most. The network depth needs to be properly increased for better recognition and positioning effects. Figure 1 shows the relationship between training error, test error, and number of network layers. It can be seen that the training error and test error increase with the growing number of layers. In a network with more layers, the training and test errors of fruit recognition descend faster. Therefore, over-fitting will not cause the two errors to rise. In theory, it is not wise to build a deep network simply by stacking several nonlinear layers. A better strategy is to add an identity mapping layer to the constructed network, or fit the ideal identity mapping based on a series of nonlinear networks. Figure 2 shows the structure of the residual learning unit that helps the neural network fit identity mapping.

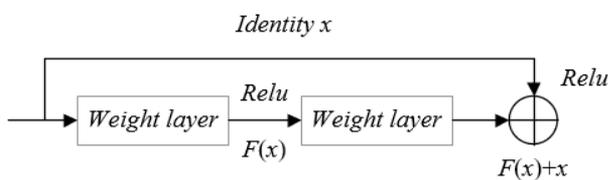


(a) Relationship between training error and number of network layers



(b) Relationship between test error and number of network layers

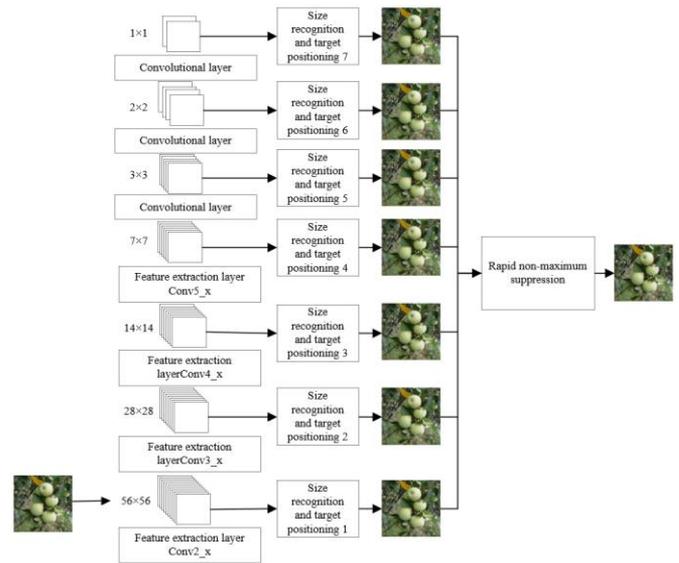
**Figure 1.** Relationship between training error, test error, and number of network layers



**Figure 2.** Structure of residual learning unit

Figure 3 shows the structure of SSD target detection model based on residual network. In the model, the area in the input image corresponding to a point on a feature map changes with the sizes of the feature map. To select the default box parameters for the target fruit, the size and position of different fruits should correspond to the default box at different positions. Based on formula (5),  $E_{min}$  was set to 0.2,  $E_{max}$  to 0.95, and default aspect ratio to  $\beta_s = \{1, 2, 3, 0.5, 0.33\}$ . Let  $s_l$  be the size of the  $l$ -th feature map. Then, the central coordinates of the default box can be configured as:

$$\left( \frac{i + 0.5}{s_l}, \frac{j + 0.5}{s_l} \right) \quad (16)$$



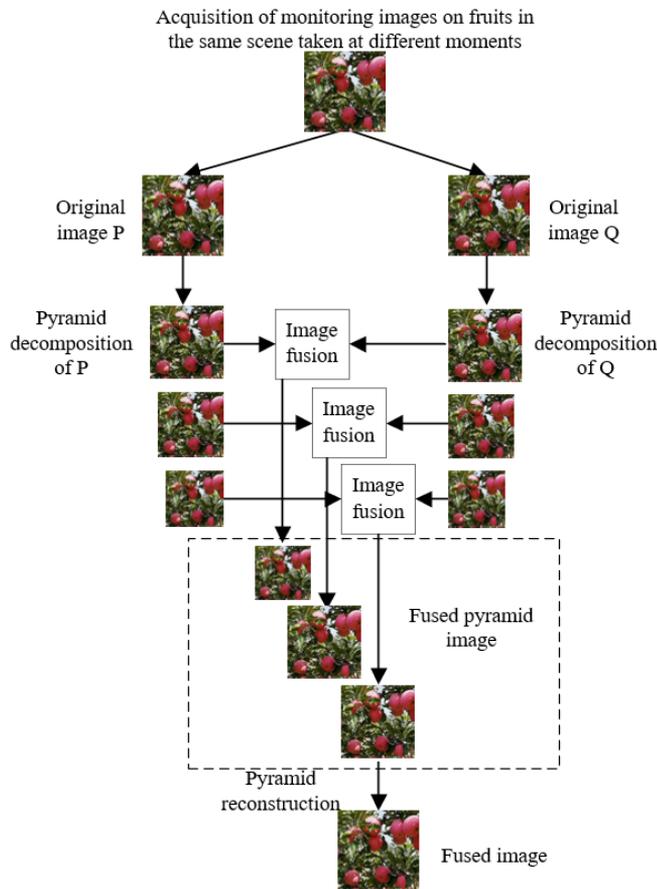
**Figure 3.** Structure of SSD target detection model based on residual network

### 3. FRUIT IMAGE FUSION ALGORITHM FOR AGRICULTURAL IOT

With the progress of science and technology, the integration between the IoT and image recognition open new directions to maturity monitoring of agricultural fruits. For example, wireless sensor network (WSN) can detect maturity information like fruit size and color via multimedia visual nodes, and transmit the data on fruit maturity to the data processing center through wireless communication, realizing the full detection of fruit maturity across the orchard. Fruit growth is a long process. During the growth period, the monitoring system shots a huge number of images on the same scene, calling for effective compression. This paper introduces image fusion to synthesize the multiple images taken by the monitoring system on the same scene into a new higher quality image, which demonstrates the fruit maturity in an all-round way. The synthesis effectively enhances the image usability.

The traditional image fusion approaches are based on blocks or pixels. Despite enhancing the direct correlations between adjacent pixels, the traditional methods affect the finer visual effects, such as the size changes of the targets. This paper proposes an image fusion algorithm based on improved Laplacian pyramid. The algorithm can fuse images on different feature layers decomposed from the original image, offering a better alternative for analyzing image details. Figure

4 shows the flow of image fusion algorithm based on pyramid decomposition.



**Figure 4.** Flow of fruit image fusion algorithm for agricultural IoT

Let  $H_0$  be an original monitoring image of fruits;  $N$  be the number of feature layers in each original image. Then, the process of Gaussian pyramid decomposition is detailed first:

Firstly, a low-pass window function  $WF(x, y)$  is convoluted with the  $k$ -1-th feature layer  $H_{k-1}$  of the original image. Then, down-sampling is performed on the convolution result every other row and every other column:

$$H_k = \sum_{x=-2}^2 \sum_{y=-2}^2 WF(x, y) H_{k-1}(2i+x, 2j+y) \quad (17)$$

where,  $WF(x, y)$  is a  $5 \times 5$  low-pass filter. Let  $DS_k$  and  $HS_k$  be the number of columns and rows in the sub-image on the  $k$ -th layer of Gaussian pyramid. Then,  $k, i$  and  $j$  satisfy  $0 < k \leq N, 0 \leq i < DS_k$ , and  $0 \leq j < HS_k$ , respectively.

$WF(x, y)$  meets four constraints at the same time: separability, normalizability, symmetry, and equal contribution between odd and even terms. Let  $g$  be the Gaussian density distribution function. The separability constraint can be expressed as:

$$WF(x, y) = g(x) * g(y) \quad (18)$$

The normalizability constraint can be expressed as:

$$\sum_{N=-2}^2 g(x) = 1 \quad (19)$$

The symmetry constraint can be expressed as:

$$g(i) = g(-i) \quad (20)$$

The constraint of equal contribution between odd and even terms can be expressed as:

$$g(0) + g(-2) + g(+2) = g(-1) + g(+1) \quad (21)$$

Under the above four constraints, we have:

$$\begin{cases} g(0) = 0.375 \\ g(-1) = g(+1) = 0.25 \\ g(-2) = g(+2) = 0.0625 \end{cases} \quad (22)$$

Under the separability constraint (18), the window function  $WF(x, y)$  can be calculated by:

$$WF = \begin{bmatrix} \frac{1}{256} & \frac{1}{64} & \frac{1}{42} & \frac{1}{64} & \frac{1}{256} \\ \frac{1}{64} & \frac{1}{16} & \frac{1}{11} & \frac{1}{16} & \frac{1}{64} \\ \frac{1}{42} & \frac{1}{11} & \frac{1}{7} & \frac{1}{11} & \frac{1}{42} \\ \frac{1}{64} & \frac{1}{16} & \frac{1}{11} & \frac{1}{16} & \frac{1}{64} \\ \frac{1}{256} & \frac{1}{64} & \frac{1}{42} & \frac{1}{64} & \frac{1}{256} \end{bmatrix} \quad (23)$$

The Gaussian pyramid image sequence can be described as  $H_0, H_1, \dots, H_N$ . Starting with each original image, the image  $H_k$  of the previous feature layer is processed through low-pass filtering and down-sampling to obtain the feature map  $H_{k-1}$  of the current layer. The final Gaussian pyramid image is composed of the feature maps in ascending order. The top and bottom layers are  $H_N$  and  $H_0$ , respectively. The total number of layers is  $N+1$ .

After obtaining the Gaussian image pyramid, interpolation is adopted to expand the image  $H_k$  on the  $k$ -th feature layer, while ensuring that the expanded image  $H'_k$  is of the same size as the feature image  $H_{k-1}$  on layer  $k-1$ :

$$H'_k = EX(H_k) \quad (24)$$

The expansion operator  $EX$  can be defined as:

$$H'_k(i, j) = 4 \sum_{x=-2}^2 \sum_{y=-2}^2 WF(x, y) H_k^* \left[ \frac{x+i}{2}, \frac{y+j}{2} \right] \quad (25)$$

where,

$$H_k^* \left( \frac{x+i}{2}, \frac{y+j}{2} \right) = \begin{cases} H_k \left( \frac{x+i}{2}, \frac{y+j}{2} \right), & \text{if } \frac{x+i}{2} \text{ and } \frac{y+j}{2} \text{ are integers} \\ 0, & \text{otherwise} \end{cases} \quad (26)$$

The image  $IM_k$  on the k-th feature layer in the Laplacian pyramid, which is generated from Gaussian pyramid, can be given by:

$$IM_k = H_k - H'_k \quad (27)$$

$$IM_N = H_N \quad (28)$$

The complete Laplacian pyramid is composed of the image sequence  $IM_0, IM_1, \dots, IM_m$ , which can be reconstructed by:

$$H_k = IM_k + H'_k \quad (29)$$

From the top of the Laplacian pyramid, each layer is deduced by formula (29) until  $k=0$ . In this way, the original image on the target fruit can be reconstructed. To a certain extent, the image reconstructed through traditional Laplacian transform is influenced by the coefficient noise of the transform domain. Therefore, the reconstructed image is not necessarily the optimal image.

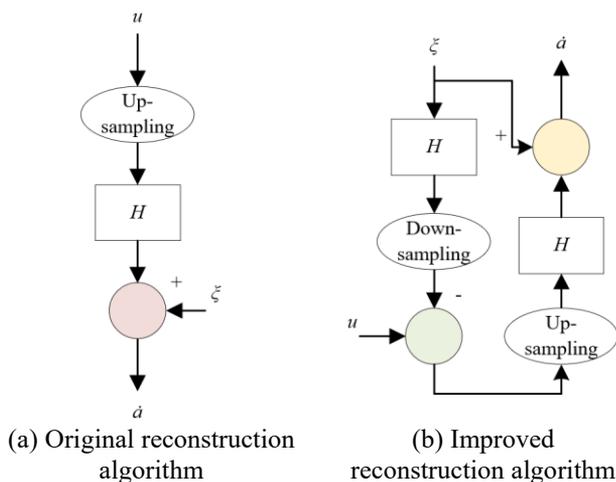
The traditional Laplacian reconstruction algorithm can be expressed as:

$$\hat{a}_T = H \cdot u + \xi \quad (30)$$

This paper improves the Laplacian pyramid algorithm for image reconstruction:

$$\hat{a}_2 = H \cdot u + (1 - HP)\xi \quad (31)$$

Figure 5 compares the reconstruction processes of the original and improved algorithms. The two algorithms process the difference signal  $\xi$  differently. The original algorithm directly superposes  $\xi$  values, while the improved algorithm superposes the projections of  $\xi$ . The improved algorithm outperforms the original algorithm, because it can effectively eliminate the influence of some errors, when the transform domain coefficients are noisy.



**Figure 5.** Reconstruction processes of the original and improved algorithms

After Laplacian pyramid transform, each original image is decomposed into different spatial frequency bands. Then, the detailed features are extracted from each layer of the decomposed image, using different fusion operators based on

regional features.

For original images P and Q, the images on the k-th layer decomposed by Laplacian pyramid are denoted as  $IP_k$  and  $IQ_k$ , respectively; the fused image is denoted as  $IS_k(0 \leq k \leq N)$ . For an image, the greater the mean gradient, the more obvious the changes to the edges of the target, and the easier it is to extract target size. Therefore, this paper fuses the top-level images through local mean gradient method. Suppose  $x$  and  $y$  are two odd numbers no smaller than 3. Then, the mean gradient is calculated for the  $x \times y$  area centering on each pixel of the image. Let  $\Delta J_a$  and  $\Delta J_b$  be the first-order differences of pixel  $PI(a, b)$  in directions  $a$  and  $b$ , respectively. Then, we have:

$$AG = \frac{1}{(x-1)(y-1)} \sum_{i=1}^{x-1} \sum_{j=i}^{y-1} \sqrt{(\Delta J_a^2 + \Delta J_b^2)} / 2 \quad (32)$$

$\Delta J_a$  and  $\Delta J_b$  can be respectively defined as:

$$\Delta J_a = PI(a, b) - PI(a-1, b) \quad (33)$$

$$\Delta J_b = PI(a, b) - PI(a, b-1) \quad (34)$$

The fused image can be expressed as:

$$IS_k(i, j) = \begin{cases} IP_k(i, j) & HP(i, j) \geq HQ(i, j) \\ IQ_k(i, j) & HW(i, j) \leq HQ(i, j) \end{cases} \quad (35)$$

This paper fuses the non-top-layer decomposed images with  $0 \leq k \leq N$  with area energy method. Firstly, the area size is set to  $3 \times 3$ , and weighting coefficient to  $\omega(\gamma, \theta)$ . Then, the energy of the local area  $LAE_k(i, j)$  centering at  $(i, j)$  on the k-th layer of Laplacian pyramid can be calculated by:

$$LAE_k(i, j) = \sum_{\gamma=-1}^1 \sum_{\theta=-1}^1 \omega(\gamma, \theta) [\psi_i(i + \gamma, j + \theta)]^2 \quad (36)$$

The matching degree of the local area  $LAM_k(i, j)$  can be expressed as:

$$LAM_k(i, j) = \frac{\sum_{\gamma=-1}^1 \sum_{\theta=-1}^1 \omega(\gamma, \theta) \psi_{i,P}(i + \gamma, j + \theta) \psi_{i,Q}(i + \gamma, j + \theta)}{LAE_{k,P}(i, j) + LAE_{k,Q}(i, j)} \quad (37)$$

Next is to define the matching threshold  $\varphi$  for images. In an area, if the matching degree between two images is smaller than  $\varphi$ , then the two images have great differences. In this case, the central pixel in an area with large local area energy can be selected as the central pixel of the image fused from the two images. If the matching degree between two images is greater than  $\varphi$ , then the two images have little differences. In this case, the gray value of the area in the fused image can be determined by weighted fusion algorithm. If  $LAM_k(i, j) < \varphi$ , then:

$$\begin{cases} \psi_{k,S}(i, j) = \psi_{k,P}(i, j), LAM_{k,P}(i, j) \geq LAM_{k,Q}(i, j) \\ \psi_{k,S}(i, j) = \psi_{k,Q}(i, j), LAM_{k,P}(i, j) < LAM_{k,Q}(i, j) \end{cases} \quad (38)$$

When  $LAM_k(i, j) \geq \varphi$ , if  $LAE_{k,P}(i, j) > LAE_{k,Q}(i, j)$ , we have:

$$\begin{aligned} \psi_{k,S}(i, j) &= \mu_{k,max}(i, j)\psi_{k,P}(i, j) \\ &+ \mu_{k,min}(i, j)\psi_{k,Q}(i, j) \end{aligned} \quad (39)$$

If  $LAE_{k,P}(i, j) < LAE_{k,Q}(i, j)$ , we have:

$$\begin{aligned} \psi_{k,S}(i, j) &= \mu_{k,min}(i, j)\psi_{k,P}(i, j) \\ &+ \mu_{k,max}(i, j)\psi_{k,Q}(i, j) \end{aligned} \quad (40)$$

where,  $\mu_{k,min}(i, j)$  and  $\mu_{k,max}(i, j)$  are weighted fusion operators:

$$\begin{cases} \mu_{k,min}(i, j) = \frac{1}{2} - \frac{1}{2} \left( \frac{1 - LAM_k(i, j)}{1 - o} \right) \\ \mu_{k,max} = 1 - \mu_{k,min}(i, j) \end{cases} \quad (41)$$

#### 4. EXPERIMENTS AND RESULTS ANALYSIS

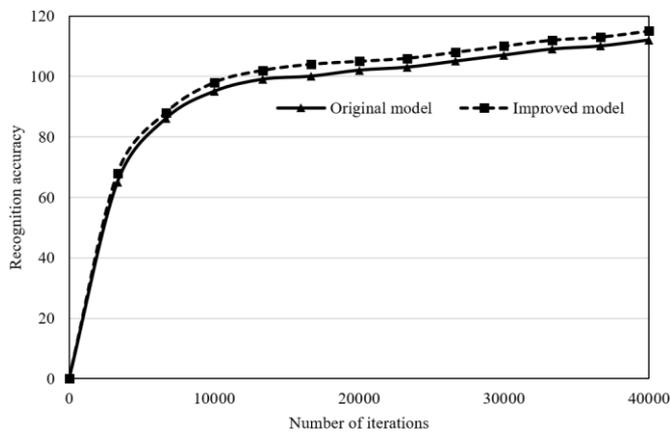


Figure 6. Recognition accuracy curves

To test the convergence and fruit recognition accuracy of the improved SSD target detection model, this paper designs a comparative experiment between the original SSD and the improved SSD on a dataset of fruit maturity monitoring samples in the context of agricultural IoT. Figures 6 and 7 show the recognition accuracy curves and the relationship between the number of fruits and recognition accuracy, respectively.

As shown in Figure 6, within 4,000 iterations, the original and improved SSD models differed slightly in the recognition rate curve of the fruits in the dataset. After 10,000 iterations, the improved SSD model achieved a slightly higher recognition rate than the original model. As shown in Figure 7, the recognition accuracy is negatively correlated with the number of fruits in each monitoring image. Therefore, a small number of fruits per image benefits the fruit recognition accuracy. The improved SSD model greatly enhances the accuracy of fruit recognition. When the number of fruits was greater than 2, the improved SSD achieved a much higher recognition rate than the original model. When the number increased to 7, the recognition rate of the improved SSD was still greater than 90%.

The accuracy and recall of positive samples (Figure 8) show that our agricultural IoT-driven fruit recognition method clearly outshined the traditional SSD and R-CNN, a CNN-based regional method, in the accuracy and recall of positive samples. Our method had no significant error, although it was

not as good as the traditional SSD in the extraction of the maturity features of a few fruits. The improved SSD optimizes the feature extraction from fruits with size and shape changes.

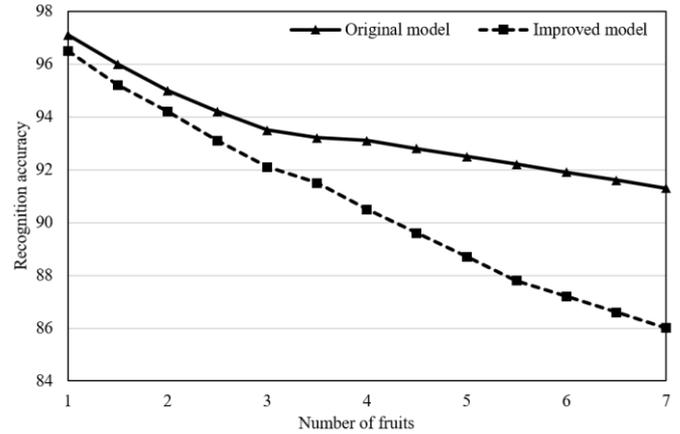
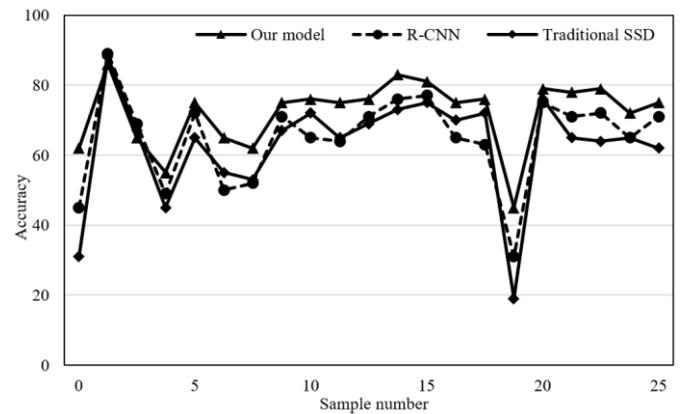
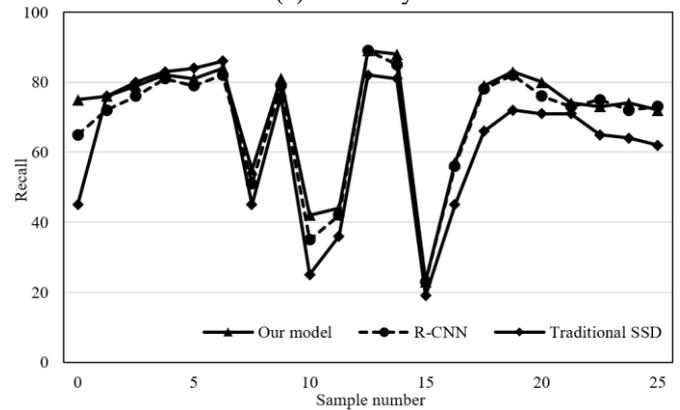


Figure 7. Relationship between the number of fruits and recognition accuracy



(a) Accuracy



(b) Recall

Figure 8. Accuracy and recall of positive samples

Note: R-CNN is short for region-based convolutional neural network.

Next, the above three models were compared against several metrics of size estimation and positioning effect (Table 1). On positive samples, the proposed improved SSD model achieved an accuracy of 93.74%, a recall of 92.65%, an F1-score of 93.56%, a Cohen's kappa coefficient of 93.89%, and a mean accuracy (MA) of 93.15%. The MA of our model was 3.33% and 1.26% higher than the original SSD and R-CNN, respectively. Overall, our SSD model slightly outperforms the other two models.

Next, the authors further examined how the number of layers decomposed by Laplacian pyramid in the agricultural IoT-driven fruit image fusion algorithm influences the fusion quality. Our method was applied to fuse the monitoring images on fruit maturity. Table 2 presents the fused image quality at different number of layers. The relationship between the number of layers and four objective metrics of image fusion effect can be observed from the table, including MSE, mutual information, cross entropy, and peak S/N. The four metrics gradually improved with the increasing number of layers. Considering the effect of computing load, this paper sets the number of layers decomposed by Laplacian pyramid to five.

**Table 1.** Performance metrics of three models

Metric	Traditional SSD	R-CNN	Our model
Accuracy	90.72	92.57	93.74
Recall	91.75	91.35	92.65
F1-score	91.46	91.46	93.56
Cohen’s kappa coefficient	91.34	91.52	93.89
MA	91.26	92.63	93.15

**Table 2.** Fused image quality at different number of layers

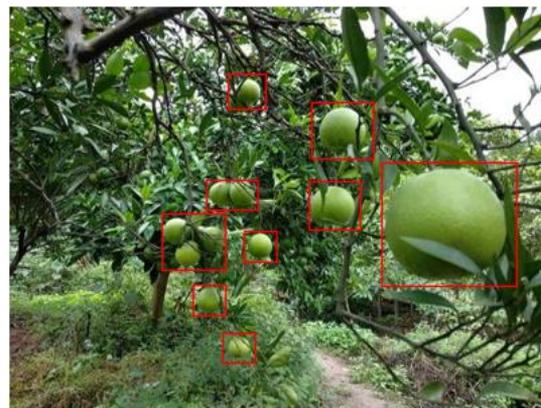
Number of layers	MSE	Mutual information	Cross entropy	Peak S/N
2	5.3246	5.0721	0.0600	33.6759
3	2.7984	5.3256	0.0185	39.5031
4	1.8823	5.2470	0.0075	42.8254
5	1.8675	5.2768	0.0068	42.1762

Note: MSE is short for mean squared error; S/N is short for signal-to-noise ratio.

After that, a contrastive experiment was carried out on two clear images taken with an interval of 1 week in the dataset of



(a)



(b)

**Figure 9.** Recognition effects of fruit maturity

## 5. CONCLUSIONS

This paper probes into the image recognition of fruit maturity in the context of agricultural IoT. The SSD was improved to accurately recognize the size and position of each target fruit. Next, the improved Laplacian pyramid was adopted to design an image fusion algorithm, which effectively compresses the large fruit monitoring images shot in the same scene. After that, the authors experimentally obtained the recognition accuracy curves, and further explored

fruit maturity monitoring images. The fusion results of five fusion algorithms were compared, including traditional Laplacian transform, wavelet transform, multiscale transform, Ridgelet transform, and our algorithm. The performance of these algorithms is compared in Table 3, using metrics like mean gradient, information entropy, and peak S/N.

**Table 3.** Image fusion performance of five different algorithms

Name of algorithm	Mean gradient	Information entropy	Peak S/N
Traditional Laplacian transform	2.4356	6.8572	30.1578
Wavelet transform	2.7523	6.7534	33.6842
Multiscale transform	2.6862	7.0239	33.7595
Ridgelet transform	2.6714	7.0458	33.8267
Our algorithm	2.7527	7.1275	35.7653

As shown in Table 3, the fused images obtained through traditional Laplacian transform and traditional wavelet transform had problems like dimness, unclarity, ghosting, and noises, i.e., the two algorithms achieved poor fusion effects. Multiscale transform realized better fusion effect than the two algorithms, but did not output sufficient clarity. The images fused by Ridgelet transform had fuzzy edges. Compared with the other methods, our algorithm can fully preserve the details of the original images, and improve the subjective visual effects of fruit edges.

A total of 250 fruit images taken in six months were fused by our model to verify its performance of maturity recognition. Our model spent 213.10s recognizing all images, that is, 0.852s on each image. Figure 9 visualizes some of the detection results. The recognition rates in the red boxes are the final recognition results on orange maturity.

the relationship between the number of fruits and recognition accuracy. Furthermore, the accuracy and recall on positive samples, and recognition performance of multiple models were compared. The comparison shows that our model outperformed the other models in positive sample accuracy, recall, F1-score, Cohen’s kappa coefficient, and MA. Finally, the performance of different image fusion algorithms was contrasted, revealing that our algorithm can fully preserve the details of the original images, and improve the subjective visual effects of fruit edges.

## ACKNOWLEDGEMENT

This work is funded by the National Social Science Fund of China (Project Title: Construction and Application of Mobile Library Service Model Based on Context-aware, Grant No.: 19BTQ045).

## REFERENCES

- [1] Hu, B., Sun, D.W., Pu, H., Wei, Q. (2019). Recent advances in detecting and regulating ethylene concentrations for shelf-life extension and maturity control of fruit: A review. *Trends in Food Science & Technology*, 91: 66-82. <https://doi.org/10.1016/j.tifs.2019.06.010>
- [2] Chandra, T.G., Erwandha, K.G., Aditya, Y., Priyantari, B.A., Fitriani, D., Hakim, M.H., Hidayat, A.S., Hatta, A.M., Irawati, N. (2019). Tomatoes selection system based on fruit maturity level using digital color analysis method. In *Third International Seminar on Photonics, Optics, and Its Applications (ISPhOA 2018)*, Surabaya, Indonesia. <https://doi.org/10.1117/12.2504987>
- [3] Sukrisno, E. (2019). Identification using the K-Means clustering and gray level co-occurrence matrix (GLCM) at maturity fruit oil head. *2019 Fourth International Conference on Informatics and Computing (ICIC)*, Semarang, Indonesia. <https://doi.org/10.1109/ICIC47613.2019.8985681>
- [4] Takahashi, M., Hirose, N., Ohno, S., Arakaki, M., Wada, K. (2018). Flavor characteristics and antioxidant capacities of hihatsumodoki (*Piper retrofractum* Vahl) fresh fruit at three edible maturity stages. *Journal of Food Science and Technology*, 55(4): 1295-1305. <https://doi.org/10.1007/s13197-018-3040-2>
- [5] Tolesa, G.N., Workneh, T.S., Melesse, S.F. (2018). Modelling effects of pre-storage treatments, maturity stage, low-cost storage technology environment and storage period on the quality of tomato fruit. *CyTA-Journal of Food*, 16(1): 271-280. <https://doi.org/10.1080/19476337.2017.1392616>
- [6] VanderWeide, J., Medina-Meza, I.G., Frioni, T., Sivilotti, P., Falchi, R., Sabbatini, P. (2018). Enhancement of fruit technological maturity and alteration of the flavonoid metabolomic profile in Merlot (*Vitis vinifera* L.) by early mechanical leaf removal. *Journal of Agricultural and Food Chemistry*, 66(37): 9839-9849. <https://doi.org/10.1021/acs.jafc.8b02709>
- [7] Aliteh, N.A., Mison, N., Aris, I., Mohd Sidek, R., Tashiro, K., Wakiwaka, H. (2018). Triple flat-type inductive-based oil palm fruit maturity sensor. *Sensors*, 18(8): 2496. <https://doi.org/10.3390/s18082496>
- [8] Javel, I.M., Bandala, A.A., Salvador, R.C., Bedruz, R.A.R., Dadios, E.P., Vicerra, R.R.P. (2019). Coconut fruit maturity classification using fuzzy logic. *2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*, Baguio, Philippines. <https://doi.org/10.1109/HNICEM.2018.8666231>
- [9] Abdullah, S., Supian, L.S., Arsad, N., Zan, S.D., Bakar, A.A.A. (2018). Assessment of palm oil fruit bunch maturity based on diffuse reflectance spectroscopy technique. *2018 IEEE 7th International Conference on Photonics (ICP)*, Langkawi, Malaysia. <https://doi.org/10.1109/ICP.2018.8533172>
- [10] Elavarasi, G., Murugaboopathi, G., Kathirvel, S. (2019). Fresh fruit supply chain sensing and transaction using IoT. *2019 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS)*, Tamilnadu, India. <https://doi.org/10.1109/INCOS45849.2019.8951326>
- [11] Ray, P.P., Dash, D., De, D. (2019). Approximation of fruit ripening quality index for IoT based assistive e-healthcare. *Microsystem Technologies*, 25(8): 3027-3036. <https://doi.org/10.1007/s00542-018-4238-y>
- [12] DiSalvo, C., Jenkins, T. (2017). Fruit are heavy: A prototype public IoT system to support urban foraging. *Proceedings of the 2017 Conference on Designing Interactive Systems*, Edinburgh United Kingdom, pp. 541-553. <https://doi.org/10.1145/3064663.3064748>
- [13] Ray, P.P., Pradhan, S., Sharma, R.K., Rasaily, A., Swaraj, A., Pradhan, A. (2016). IoT based fruit quality measurement system. *2016 Online International Conference on Green Engineering and Technologies (IC-GET)*, Coimbatore, India. <https://doi.org/10.1109/GET.2016.7916620>
- [14] Ruan, J., Shi, Y. (2016). Monitoring and assessing fruit freshness in IOT-based e-commerce delivery using scenario analysis and interval number approaches. *Information Sciences*, 373: 557-570. <http://doi.org/10.1016%2Fj.ins.2016.07.014>
- [15] Syaifuddin, A., Mualifah, L.N.A., Hidayat, L., Abadi, A.M. (2020). Detection of palm fruit maturity level in the grading process through image recognition and fuzzy inference system to improve quality and productivity of crude palm oil (CPO). *Journal of Physics: Conference Series*, 1581(1): 012003. <http://doi.org/10.1088/1742-6596/1581/1/012003>
- [16] Zheng, S., Lin, Z., Xie, J., Liao, M., Gao, S., Zhang, X., Qiu, T. (2021). Maturity recognition of citrus fruits by Yolov4 neural network. *2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)*, Nanchang, China, pp. 564-569. <https://doi.org/10.1109/ICBAIE52039.2021.9389879>
- [17] Dimatira, J.B.U., Dadios, E.P., Culibrina, F., Magsumbol, J.A., Cruz, J.D., Sumage, K., Tan, M.T., Gomez, M. (2016). Application of fuzzy logic in recognition of tomato fruit maturity in smart farming. *2016 IEEE Region 10 Conference (TENCON)*, Singapore, pp. 2031-2035. <https://doi.org/10.1109/TENCON.2016.7848382>
- [18] Fadhel, M.A., Al-Shamma, O. (2020). Implementing a hardware accelerator to enhance the recognition performance of the fruit mature. *International Symposium on Signal and Image Processing*, Kolkata, India, pp. 41-52. [https://doi.org/10.1007/978-981-33-6966-5\\_5](https://doi.org/10.1007/978-981-33-6966-5_5)
- [19] Xiao, Q., Niu, W., Zhang, H. (2015). Predicting fruit maturity stage dynamically based on fuzzy recognition and color feature. In *2015 6th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, Beijing, China, pp. 944-948. <https://doi.org/10.1109/ICSESS.2015.7339210>
- [20] Cai, J., Zhao, J.Z. (2005). Recognition of mature fruit in natural scene using computer vision. *Nongye Jixie Xuebao/Transactions of the Chinese Society of Agricultural Machinery*, 36(2): 61-64.