Apple Grading Model Based on Improved ResNet-50 Network

Lei Zhao[1], Qinjun Zhao[1](✉), Tao shen[1,2], and Shuhui Bi[1]

[1] University of Jinan, Jinan 250000, China
cse_zhaoqj@ujn.edu.cn
[2] HIT Robot Group Co., Ltd, Harbin 150040, China

Abstract. In this paper, we study an apple grading model based on the convolutional neural network to classify Red Fuji apples according to features of size, color and external defects. Firstly, Red Fuji apple images are collected by professional equipment, and the RGB model of apple image is extracted and transformed into HSI model. Secondly, the segmentation between apple and background is realized by Otsu method in the S channel. Thirdly, the ResNet-50 network is improved by convolutional block attention module and LeakyReLU activation function. Finally, improved ResNet-50 network is applied to apple grading and compared with other mainstream convolutional neural networks. The experimental result shows that improved ResNet-50 network reaches the highest accuracy 95.1% in apple grading experiment, which is higher than AlexNet, VGG-16, GoogleNet, Mobilenet-V2 and the ResNet-50 network.

Keywords: Apple grading · ResNet network · Attention mechanism · LeakyReLU activate function

1 Introduction

Apple grading is an important part of the apple industry. In the apple growth, picking, transportation will be more or less rot, insect pests, crushing and other damage to the quality of apple. Apple shape, diameter and color will affect the sales, and then affect the profit, so apple grading is particularly important. The early apple grading method was manual sorting, which not only consumed a lot of manpower, but also slowly and inefficiently. Therefore, fast and accurate grading of apple is of great significance to the development of apple industry.

At present, neural network, machine learning and machine vision method has been widely used in fruit quality detection and classification due to its advantages in image processing [1]. Yuhui Ji [2] et al. applied the SVM model to the grading of apple based on features of defect and color, and the accuracy reached 91%. Payman Moallem [3] et al. detected Calyx region by K-means clustering method and applied the SVM model to apple grading, and the best accuracy reached 92.5%. Maoyong Nie [4] et al. applied Canny edge detection and SVM model based on particle swarm optimization, and the

S.-H. Wang and Y.-D. Zhang (Eds.): ICMTEL 2022, LNICST 446, pp. 749–760, 2022.
https://doi.org/10.1007/978-3-031-18123-8_59

accuracy reached 92%. Zheng Xu [5] et al. applied GA-SVM model to realize the grading of apple, and the accuracy reached 92.3%. Jinquan Li [6] et al. designed a shallow neural network through the Caffe framework to classify apple, and the accuracy reached 92%. To sum up, automatic classification has been widely used in apple grading, but there is still room for improvement in the accuracy of apple automation. In order to improve the accuracy of apple grading, this paper proposes improved ResNet-50 network to classify Red Fuji apples, and the experiment result proves that this network can effectively improve apple grading accuracy.

2 Overall Design of the Apple Grading Method

In this paper, the apple grading method mainly includes data acquisition, data expansion, image segmentation, improving ResNet-50 network, data set setting and setting comparison experiment to verify the accuracy of improved ResNet-50 network. Improved ResNet-50 network joins CBAM and LeakyReLU activation function. CBAM is Convolutional Block Attention Module [7, 8], which makes it easier for networks to capture significant information by focusing more on channel and spatial features. LeakyReLU activation function is used to solve the problem of neurons failing to learn to update (sudden death of neurons problem) [9]. In this paper, improved ResNet-50 network model is built on Pytorch framework, and its accuracy is tested through the Red Fuji apples data set. The apple grading system design is shown in Fig. 1. Apple grading system design.

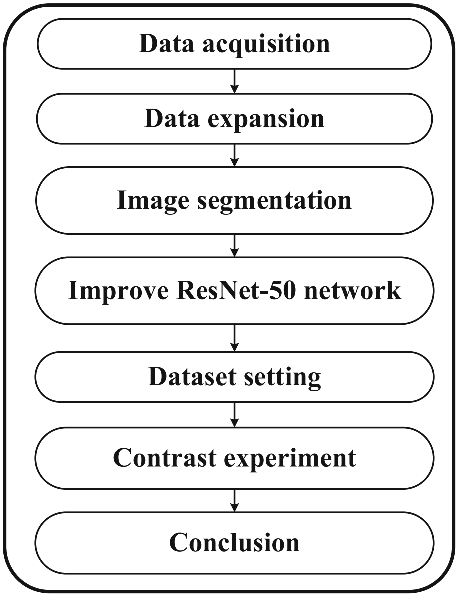


Fig. 1. Apple grading system design

2.1 Data Acquisition

The experimental material is Red Fuji apples, and apple images are collected by the apple image acquisition system, as shown in Fig. 2. Each apple is measured three times, and collected from the top, left side and right side respectively. A total of 1030 images of Red Fuji apple are collected, as shown in Fig. 3.

Fig. 2. Apple image acquisition system

Fig. 3. Red fuji apple image (left 1, 2, 3 sides; 4 top)

2.2 Data Expansion

Convolutional neural network requires a large amount of data for training and validating, so this paper applies the following methods to expand the data of apple images: (1) Vertical mirror flip, (2) Horizontal mirror flip, (3) Shrink or enlarge an image to a certain scale, (4) Using random rotating on image between $-60^\circ - 60^\circ$, (5) Using random cut on image.

After the above picture expansion method, the data set is expanded to 6800 images. The data set after expansion is shown in Fig. 4.

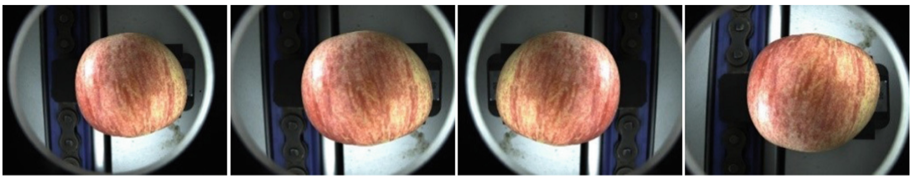


Fig. 4. Expanded red fuji apple data set (From left to right: 1. Original picture 2. Expand at random scale 3. Horizontal mirror flipping after random scaling 4. Vertical mirror flipping after random scaling)

2.3 Image Segmentation

After data expansion, Red Fuji apple images of the data set is segmented to extract apple information. In this paper, the Otsu method is used for image segmentation, which is an algorithm to obtain the global threshold of the image proposed by the Japanese scholar Otsu [10, 11]. The algorithm principle is as follows:

Assuming that the image contains L gray levels, the threshold T divides the pixels of the image into two types B_1 (less than T) and B_2 (greater than T). The mean values of the two types of pixels are m_1 and m_2 respectively, and the mean values of the total pixels is m. The probability of pixel point to B_1 and B_2 are p_1 and p_2 respectively. When σ^2 reaches its maximum value, the gray level k value is the threshold T. The calculation formula of OTSU method is as follows:

$$p_1 = \sum\nolimits_{i=0}^{k} p_i \tag{1}$$

$$p_2 = \sum\nolimits_{i=k+1}^{L} p_i \tag{2}$$

$$p_1 m_1 + p_2 m_2 = m \tag{3}$$

$$p_1 + p_2 = 1 \tag{4}$$

$$\sigma^2 = p_1(m_1 - m)^2 + p_2(m_2 - m)^2. \tag{5}$$

The specific steps of apple image segmentation are as follows:

RGB model is extracted and transformed into HSI model by the following conversion formula.

$$H = \begin{cases} \theta & G \geq B \\ 360^\circ - \theta & G < B \end{cases} \tag{6}$$

$$S = 1 - \frac{3min(R, G, B)}{R + G + B} \tag{7}$$

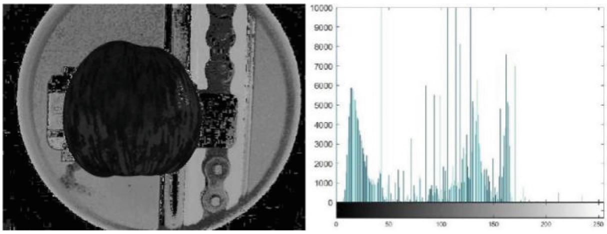
$$I = \frac{R + G + B}{3} \tag{8}$$

where

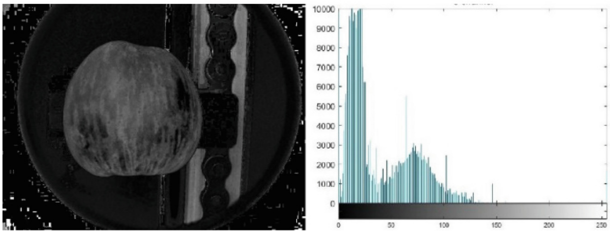
$$\theta = arccos\left\{ \frac{0.5[(R - G) + (R - B)]}{\left[(R - G)^2 + (R - B)(G - B)\right]^{\frac{1}{2}}} \right\}. \tag{9}$$

The HSI channel gray scale map and histogram information map are shown as Fig. 5.

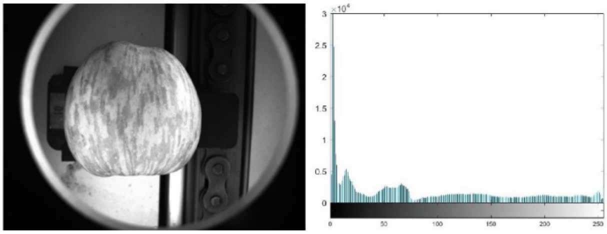
After the experiment and comparison of histogram information in channels H, S and I, the difference between apple and background information in channel S is more obvious, so S channel is chosen for image segmentation. Firstly, 3 × 3 mean-filter is selected to process the S channel gray image, and then the segmentation between apple



(a) H channel image (left) H channel histogram information (right)



(b) S channel image (left) S channel histogram information (right)



(c) I channel image (left) I channel histogram information (right)

Fig. 5. HSI channel grayscale image and histogram information

and background is realized by Otsu method. After this step, a small part of images have incomplete segmentation, so using the hole filling, expansion and corrosion operation to remove noise points, and making the white area as close as the shape of the original apple. Finally, the image is completely segmented, the middle area is filled with RGB color to obtain the final result. The image segmentation process is shown in Fig. 6.

Fig. 6. Image segmentation process (From left to right:1. Image after OTSU method 2,3. Morphologic process 4. Image filling with RGB channel)

2.4 Improving ResNet-50 Network

Deep convolutional neural network has the problem of degradation, that is, with the increase of network depth, network accuracy becomes saturated or even decreases. ResNet Network is Residual Network, which adopts unique jump connection technology, that greatly alleviates the degradation problem of deep neural Network [12].

Convolutional Block Attention Module

Convolutional Block Attention Module (CBAM) is an attentional mechanism module, which is divided into Channel Attention Module and Spatial Attention Module. The structure of CBAM is shown in Fig. 7.

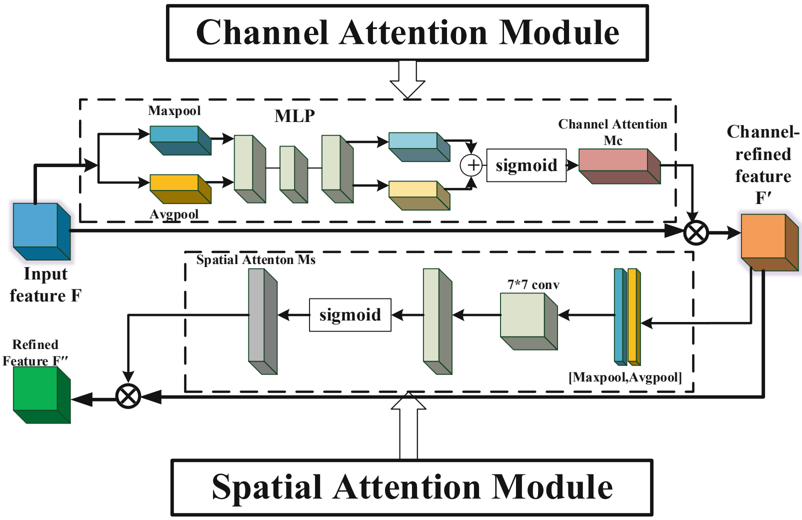


Fig. 7. CBAM structure

In the Channel Attention Module, the input feature is F, , which goes through global Max-pooling and global Avg-pooling based on image size respectively, and then enter MLP (Muti-Layer Perception). The MLP output two features are added based on elements. The combinative feature graph Mc is generated with sigmoid activation function. Mc and input feature F are multiplied based on elements to generate feature F', which is the output feature of the Channel Attention Module. The Mc calculation formula is as follows: (σ is the sigmoid activation function)

$$Mc = \sigma\{MLP[MaxPool(F)] + MLP[AvgPool(F)]\} \tag{10}$$

$$F' = Mc \otimes F. \tag{11}$$

In the Spatial Attention Module, the input is F', which goes through channel-based global Max-pooling and global Avg-pooling respectively, and then goes through a channel-based merge operation and a convolution layer to reduce the dimension to one channel. The feature graph Ms is generated with sigmoid activation function. Ms and input feature F' are multiplied based on elements to generate feature F'', which is the output feature of the Spatial Attention Module. Ms calculation formula is as follows: ($f^{7\times7}$ is the convolution layer of 7×7; σ is the sigmoid activation function)

$$Ms = \sigma\left\{f^{7\times7}\left[MaxPool\left(F'\right); AvgPool\left(F'\right)\right]\right\} \tag{12}$$

$$F'' = \mathrm{Ms} \otimes F'. \tag{13}$$

ResNet-50 Network and CBAM

In this paper, CBAM is added behind the block activation function. The output of the activation function after the first convolutional network in the block is set as F, and CBAM is added sequentially. The output is denoted as F''. The input F'' into the next convolution layer of the block until the optimization of the whole model is completed. By adding CBAM, the accuracy of the network can be effectively improved. Improved ResNet-50 network is shown in Fig. 8**.**

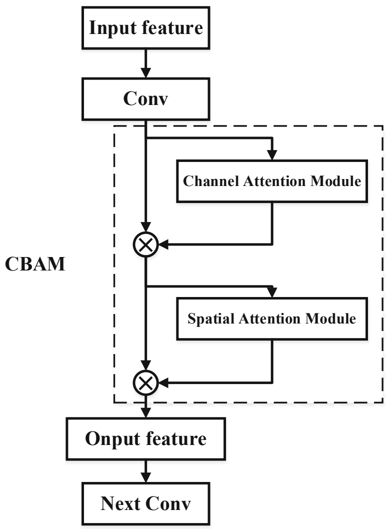


Fig. 8. Improved ResNet-50 network (Conv is convolution network)

Activation Function

In this paper, LeakyReLU activation function is used. Compared to the ReLU activation function, the LeakyReLU activation function has the following characteristics: When the input x is less than 0, there is still a very small gradient of the value $a * x$ output (a is a very small positive number, custom), thus avoiding the problem of neurons failing to update and learn. ReLU and LeakyReLU activation functions are shown in Fig. 9.

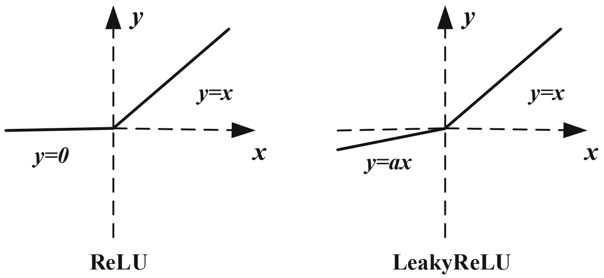


Fig. 9. Activation functions (left: ReLU; right LeakyRelu)

ReLU and LeakyReLU activation functions is calculated by the following formula: (a is a small positive number)

$$F(x) = \begin{cases} 0 \; x \leq 0 \\ x \; x > 0 \end{cases} \tag{14}$$

$$F(x) = \begin{cases} a * x \; x \leq 0 \\ x \qquad x > 0 \end{cases}. \tag{15}$$

3 Experiment and Conclusion

3.1 Data Set

In this paper, Red Fuji apple data set contains 3054 images. Firstly, according to our demand, we classify the overall data of apples into three parts: first-class, second-class and third-class. After being selected by several professionals according to the grading standard of Red Fuji apples as shown in Table 1, a total of 781 first-class apples, 1241 s-class apples and 1032 third-class apples are sorted out.

Then, the above data are randomly divided into training-set and verification-set in a ratio of 8:2. Training-set 2448, validation-set 606.The apple grading data set is shown in Table 2.

Table 1. Red fuji apple grading standard

Feature\Grade	First-class	Second-class	Third-class
Apple shape	Apple shape index ≥ 0.85	Apple shape index ≥ 0.80	Basic shape with this apple class
Apple color	Red colored area ≥ 75%	Red colored area ≥ 50%	Red colored area ≥ 25%
Diameter	≥ 80 mm	≥ 70 mm	≥ 65 mm
Diseases	No	No	No
Wound	No	No	No or slight
Chapped	No	No	No or slight
Sunburnt apple	No	No	No or slight

Table 2. Apple grading data set

	First-class	Second-class	Third-class	Total number
Training-set	625	997	826	2448
Validation-set	156	244	206	606
Total number	781	1241	1032	3054

3.2 Experiment Set

The apple grading experiment is carried out under the Ubuntu 18.04 system, and two GPU 2080Ti are used to accelerate the training of the program.

In order to validate the accuracy of improved ResNet-50 network in Red Fuji apples grading, five mainstream convolutional neural networks is selected for comparison, including, AlexNet, VGG-16, GoogleNet, Mobilenet-V2 and unimproved ResNet-50 network. We choose the highest accuracy rate and the average of the top five accuracy rate respectively as evaluation indexes to evaluate the accuracy of each network to apple

Table 3. Red fuji apples grading accuracy

Network\Aaccuracy	Highest-accuracy (%)	Average of the top five accuracy (%)
AlexNet	88.0	87.5
VGG-16	88.2	87.5
GoogleNet	91.8	90.7
Mobilenet-V2	87.2	86.8
ResNet-50	89.5	88.6
Improved ResNet-50	95.1	94.8

grading. Red Fuji apples grading accuracy is shown in Table 3. The accuracy comparison curve of each network is shown in Fig. 10.

The experimental result shows that the highest accuracy of improved ResNet-50 network is 95.1%, 3.3% higher than that of the second-rank GoogleNet network in Red Fuji apples grading, and the average of the top five accuracy of improved ResNet-50 network is 94.8%, 4.1% higher than that of the second-rank GoogleNet network. Improved ResNet-50 network achieves 95.1% accuracy in apple grading. To sum up, improved ResNet-50 network combined with combination of CBAM and LeakyReLU activation function has greater accuracy in apple grading.

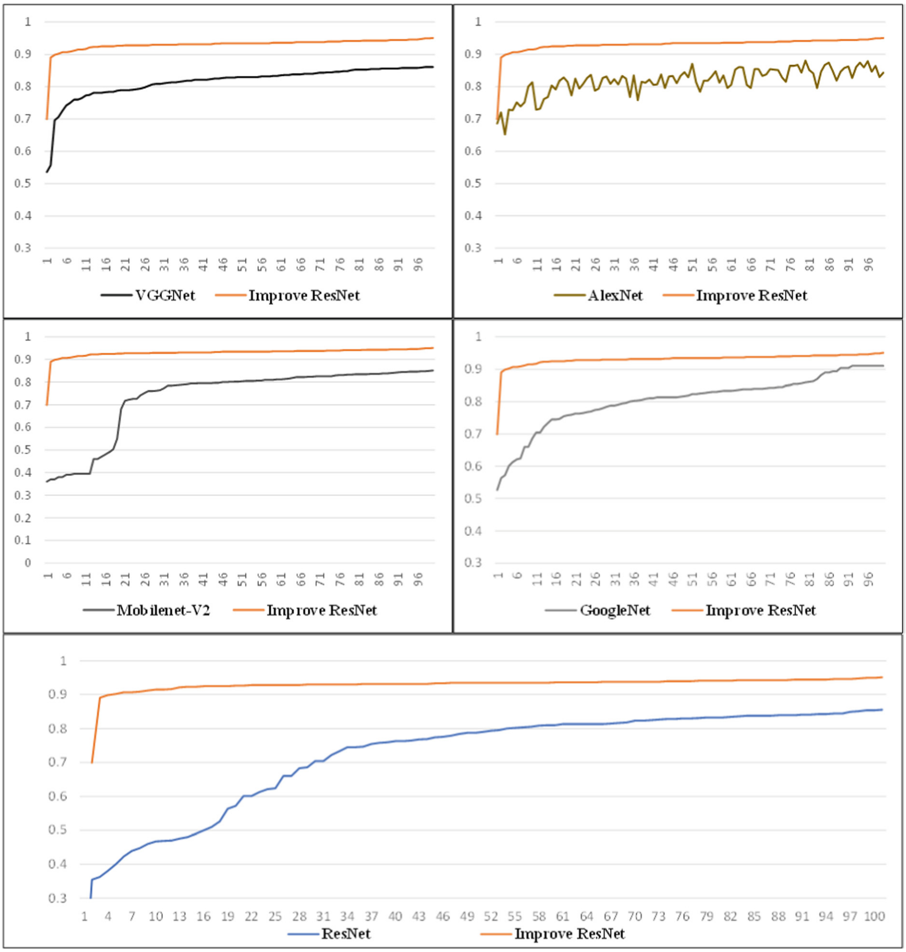


Fig. 10. Improved ResNet-50 network accuracy curve compared with VGGNet, AlexNet, Mobilenet-V2, GoogleNet, ResNet-50 (from left to right, from top to bottom).

3.3 Conclusion

This paper proposes an apple grading model based on improved ResNet-50 network by comparing the external features of the apple and draws the following conclusion: The addition of CBAM which contains a Channel Attention Module and a Spatial Attention Module, makes the network more aware of channel and spatial features and easier to extract key information. The addition of the LeakyReLU activation function alleviates the problem of neurons failing to update and learn. The above two improvements make ResNet-50 network more accurate for apple grading. In the apple grading experiment, the improved ResNet-50 network reaches high accuracy 95.1%, 5.6% improvement over the unimproved ResNet-50 network and 3.3% improvement over the second-rank network GoogleNet.

Acknowledgements. This work is partly supported by the National Key R&D Program of China (2018AAA0101703), the Key R & D project of Shandong province (2019GNC106093), Shandong agricultural machinery equipment R & D innovation plan project (2018YF011) and major scientific and technological innovation projects of Shandong province (2019JZZY021005).

References

1. Roy, K., Chaudhuri, S.S., Pramanik, S.: Deep learning based real-time Industrial framework for rotten and fresh fruit detection using semantic segmentation. Microsyst. Technol. **27**(9), 3365–3375 (2021)
2. Ji, Y., Zhao, Q., Bi, S., et al.: Apple grading method based on features of color and defect. In: Proceedings of the 2018 37th Chinese Control Conference (CCC), pp. 5364–5368. IEEE (2018)
3. Moallem, P., Serajoddin, A., Pourghassem, H.: Computer vision-based apple grading for golden delicious apples based on surface features. Inf. Process. Aagric. **4**(1), 33–40 (2017)
4. Nie, M., Zhao, Q., Xu, Y., et al.: Machine vision-based apple external quality grading. In: Proceedings of the 2019 Chinese Control and Decision Conference (CCDC), pp. 5961–5966. IEEE (2019)
5. Xu, Z., Zhao, Q., Zhang, Y., Zhang, Y., Shen, T.: Apple grading method based on GA-SVM. In: Liu, Qi., Liu, X., Shen, T., Qiu, X. (eds.) CENet 2020. AISC, vol. 1274, pp. 79–89. Springer, Singapore (2021). https://doi.org/10.1007/978-981-15-8462-6_9
6. Li, J., Xie, S., Chen, Z., et al.: A shallow convolutional neural network for apple classification. IEEE Access **2020**(8), 111683–111692 (2020)
7. Woo, S., Park, J., Lee, J.-Y., Kweon, I.S.: CBAM: convolutional block attention module. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11211, pp. 3–19. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01234-2_1
8. Ma, B., Wang, X., Zhang, H., Li, F., Dan, J.: CBAM-GAN: generative adversarial networks based on convolutional block attention module. In: Sun, X., Pan, Z., Bertino, E. (eds.) ICAIS 2019. LNCS, vol. 11632, pp. 227–236. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-24274-9_20
9. Dubey, A.K., Jain, V.: Comparative study of convolution neural network's relu and leaky-relu activation functions. In: Mishra, S., Sood, Y.R., Tomar, A. (eds.) Applications of Computing, Automation and Wireless Systems in Electrical Engineering. LNEE, vol. 553, pp. 873–880. Springer, Singapore (2019). https://doi.org/10.1007/978-981-13-6772-4_76

10. Shahabi, F., Poorahangaryan, F., Edalatpanah, S.A., et al.: A multilevel image thresholding approach based on crow search algorithm and otsu method. Int. J. Comput. Intell. Appl. **19**(02), 2050015 (2020)
11. Goh, T.Y., Basah, S.N., Yazid, H., et al.: Performance analysis of image thresholding: otsu technique. Measurement **2018**(114), 298–307 (2018)
12. McNeely-White, D., Beveridge, J.R., Draper, B.A.: Inception and ResNet features are (almost) equivalent. Cogn. Syst. Res. **2020**(59), 312–318 (2020)