

PAPER • OPEN ACCESS

Spirits quality classification based on machine vision technology and expert knowledge


To cite this article: Mengchi Chen *et al* 2023 *Meas. Sci. Technol.* **34** 055405

View the [article online](#) for updates and enhancements.

You may also like

- [Hubble Space Telescope Imaging of Luminous Extragalactic Infrared Transients and Variables from the Spitzer Infrared Intensive Transients Survey](#)
Howard E. Bond, Jacob E. Jencson, Patricia A. Whitelock et al.
- [SPIRITS 15c and SPIRITS 14buu: Two Obscured Supernovae in the Nearby Star-forming Galaxy IC 2163](#)
Jacob E. Jencson, Mansi M. Kasliwal, Joel Johansson et al.
- [Dynamic aggregation evolution of competitive societies of cooperative and noncooperative agents](#)
Zhen-Quan Lin, , Gao-Xiang Ye et al.

Spirits quality classification based on machine vision technology and expert knowledge

Mengchi Chen¹ , Hao Liu², Suyi Zhang³, Zhiyong Liu⁴, Junpeng Mi¹, Wenjun Huang^{1,*} and Delin Li³

¹ College of Control Science and Engineering, Zhejiang University, Hangzhou, People's Republic of China

² State Key Laboratory of Fluid Power and Mechatronics Systems, School of Mechanical Engineering, Zhejiang University, Hangzhou, People's Republic of China

³ Luzhou Laojiao Group Co., Ltd, Luzhou, People's Republic of China

⁴ Zhejiang SUPCON Technology Co., Ltd, Hangzhou, People's Republic of China

E-mail: wjhuang@zju.edu.cn

Received 13 September 2022, revised 26 December 2022

Accepted for publication 13 January 2023

Published 8 February 2023



CrossMark

Abstract

By combining machine vision technology and expert knowledge, this paper proposes an online intelligent classification solution for Chinese spirits, which effectively improves the classification accuracy and production efficiency of spirits. Specifically, an intelligent spirits quality classification system is first designed, including spirits collectors, image sampling cameras, and computing devices. According to the principle that the size and shape of the bubbles in the spirits collector will change with the alcohol content in the spirits, a classification method of spirits quality based on the convolutional neural network (CNN) and bubble region of interest (ROI) selection is proposed. Furthermore, a post-processing method based on expert knowledge is proposed to improve the accuracy of the classification algorithm. A spirits quality classification dataset containing 139 119 images is created, and 15 CNNs are tested. Test results show that the highest spirits quality classification accuracy is 98.62% after using the bubble ROI selection method, and the highest classification accuracy reached 99.82% after adopting the post-processing method. Furthermore, practical application tests show that the solution proposed in this paper can improve spirits' production quality and efficiency.

Keywords: spirits classification, machine vision, expert knowledge, convolutional neural network, region of interest

(Some figures may appear in colour only in the online journal)

1. Introduction

In recent years, the application of artificial intelligence (AI) technology in industrial manufacturing has exploded, giving

birth to industrial AI [1], which has become the technical core of intelligent manufacturing. AI-driven manufacturing has significantly improved many aspects of the closed-loop production chain, from the manufacturing processes to the final product logistics [2]. In particular, the use of machine vision technology has greatly benefited the field of production monitoring [3–5].

Chinese spirits are distilled spirits with a long history, usually obtained by natural fermentation [6]. It is usually made from sorghum or a blend of barley, corn, rice, wheat, and sorghum and is rich in volatile components such as esters

* Author to whom any correspondence should be addressed.



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

and organic acids. Traditional solid-state fermentation and brewing mainly include the following steps: material preparation, starter preparation, fermentation, distillation, aging, and blending [7].

During the distillation process, the quality of spirits will change continuously with the distillation time due to the constantly changing content of ethylene glycol, ethyl butyrate, acetaldehyde, and other substances. Sorting and collecting spirits according to their quality is a key link in spirits production, directly related to their final quality and production efficiency. Traditional production lines classify different qualities of spirits by manually observing the characteristics of the bubbles (shape, size, and quantity) produced by pouring the spirits. Although this method is simple and easy to implement, it has the disadvantage of different classification standards and relies heavily on the personal experience of workers, which is not conducive to maintaining the stability of spirits quality and improving production efficiency. Therefore, studying the intelligent quality classification technology of spirits is very necessary and valuable.

There are few studies on online spirits quality classification based on machine vision technology. Existing studies on spirits quality classification are mainly based on analytical instruments (such as spectrometers, chromatography, and mass spectrometers) to analyze the chemical composition of spirits. Based on Fourier transform analysis and Raman spectroscopy, Mandrile *et al* [8] identified the variety, origin, and aging time of red wine. The identification accuracy of varietal and origin is 90%, and the identification accuracy of aging time is 84%. Wu *et al* [9] measured the fermentation parameters in rice wine by Raman spectroscopy and classified the fermentation stage of rice wine based on the cars-support vector machine (Cars-SVM) algorithm, with a classification accuracy of 94.90%. Pérez-Caballero *et al* [10] used ultraviolet-visible spectroscopy to analyze the chemical composition of tequila and used various machine learning algorithms such as random forest and support vector machine to classify the category of tequila. The best classification accuracy is greater than 94%.

To the authors' knowledge, there are no machine vision studies specifically focused on classification of spirits bubbles. Fortunately, several studies have used machine vision technology to analyze bubbles generated by the gas-liquid two-phase flow in chemical production processes, which inspired the research in this paper.

Liu *et al* [11] proposed an online bubble size distribution monitoring scheme through a fully convolutional network with multi-scale deblurring and multi-stage jumping feature fusion to identify the health status of the flotation froth process. Peng *et al* [12] proposed a watershed algorithm with an optimized mark and edge constraint for accurate segmentation of flotation froth images. Haas *et al* [13] proposed a faster region-based convolutional neural network (CNN) detector to locate bubbles and a shape regression CNN to approximate bubble shapes in gas-liquid multi-phase flows. Qaddoori *et al* [14] used the traditional Hough circle detection algorithm to locate tiny bubbles in water flow and designed a shallow CNN network for bubble size classification. Cerqueira and Paladino

[15] proposed a method to reconstruct bubble morphology based on anchor points and boxes, and trained a CNN network to extract bubbles that approximate ellipses. Wang *et al* [16] proposed a fast 3D reconstruction method for dilute bubbly flow based on DIF-LeNet by combining bubble information in light field images. Wang *et al* [17] proposed a framework combining a deep edge-aware network and a marker-controlled watershed algorithm to extract bubble parameters from hysteroscopic images. However, since spirits bubbles are too dense, it is difficult to extract individual bubbles for processing by image segmentation or object detection methods. Therefore, this paper adopts the image classification method to realize the classification of spirit bubbles.

Based on machine vision technology and expert knowledge, this paper proposes a complete solution for the quality classification of spirits in the actual production process. For the first time, the use of machine vision technology to replace artificial spirits quality classification in the actual production process has been successfully implemented, thereby improving the final quality and production efficiency of spirits. The main contributions of this paper are: (a) a bubble region of interest (ROI) selection algorithm based on gray-level co-occurrence matrix (GLCM) and non-maximum suppression (NMS) is proposed, which can effectively improve the spirits quality classification accuracy of the CNN; (b) a post-processing method based on expert knowledge is proposed to improve the spirits quality classification accuracy; (c) by combining machine vision technology and expert knowledge, a high-precision spirits quality classification method is proposed, and the best classification accuracy reaches 99.82%; (d) a spirits quality classification solution is proposed and applied in actual production, which effectively improves the final quality and production efficiency of spirits.

The rest of this paper is organized as follows. Section 2 introduces the principle of spirit quality classification and the spirits distillation process. The spirits quality classification algorithm based on the CNN model and bubble ROI selection algorithm is proposed in section 3. The preprocessing and post-processing methods are also proposed in the same section. In section 4, a spirits quality classification dataset is established, and different CNN models are tested to verify the effectiveness of proposed methods. The practical application tests are also carried out in this section. Finally, the conclusions are provided in section 5.

2. Spirits distillation process and classification principle

2.1. Spirits distillation processes

Ethanol and water are the main components of distilled spirits, accounting for about 98%. Spirits also contain more than 1700 trace substances [18], including alcohols (such as isobutanol and n-pentanol), esters (such as ethyl caproate, ethyl lactate, and ethyl acetate), organic acids (such as acetic acid and lactic acid), and aldehydes (such as furfural, acetaldehyde, and acetal), etc.

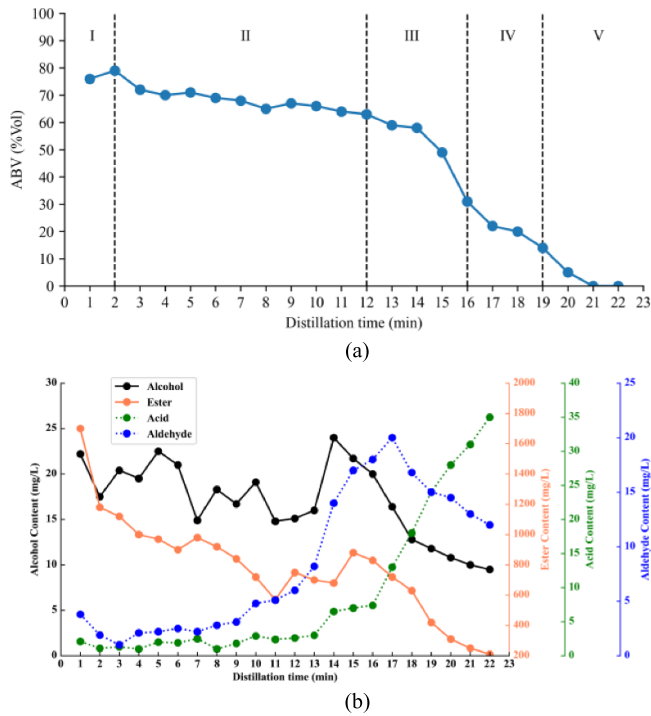


Figure 1. Changes in the content of different substances during distillation. (a) ABV changes with distillation time. (b) The alcohol, ester, acid, and aldehyde content changes with distillation time.

Take the distillation process of strong-flavor spirits as an example, which can be divided into five stages from I to V, namely the initial distillate (stage I), the second distillate (stage II), the third distillate (stage III), the last distillate (stage IV), and the tailwater (stage V). As the distillation time increases, the alcohol by volume (ABV) decreases, and the flavor gradually deteriorates, which means that the quality of spirits decreases. Figure 1 shows the effect of distillation time on the content of alcohol, ester, acid, and aldehyde in strong-flavor spirits [19].

The entire distillation lasts about 20 min. Figure 1(a) shows the division of five stages. In stage I, the ABV in the distillate is very high and contains more alcohol-soluble esters. The spirits taste in stage I is fragrant and miscellaneous. The content of ABV and ethyl hexanoate in the distillate of stage II is still relatively high, and the substances such as esters, aldehydes, and acids are relatively balanced, resulting in the best flavor and quality of spirits in this stage. As shown in figure 1(b), ethyl lactate, organic acids, and aldehydes readily soluble in the water begin to increase in stage III due to the decrease of ABV in the distillate, resulting in a slightly sour taste and insufficient flavor of spirits. However, it can still be used as ordinary spirits after storage and blending. In stage IV, the ABV in the distillate dropped sharply, resulting in a significant increase in the content of ethyl lactate, organic acids, aldehydes, and some poorly soluble oils. At this stage, the distillate begins to settle, becomes sour, and becomes irritating and greasy. The ABV in the distillate of stage V drops to zero, and many ethanol-insoluble substances are distilled out.

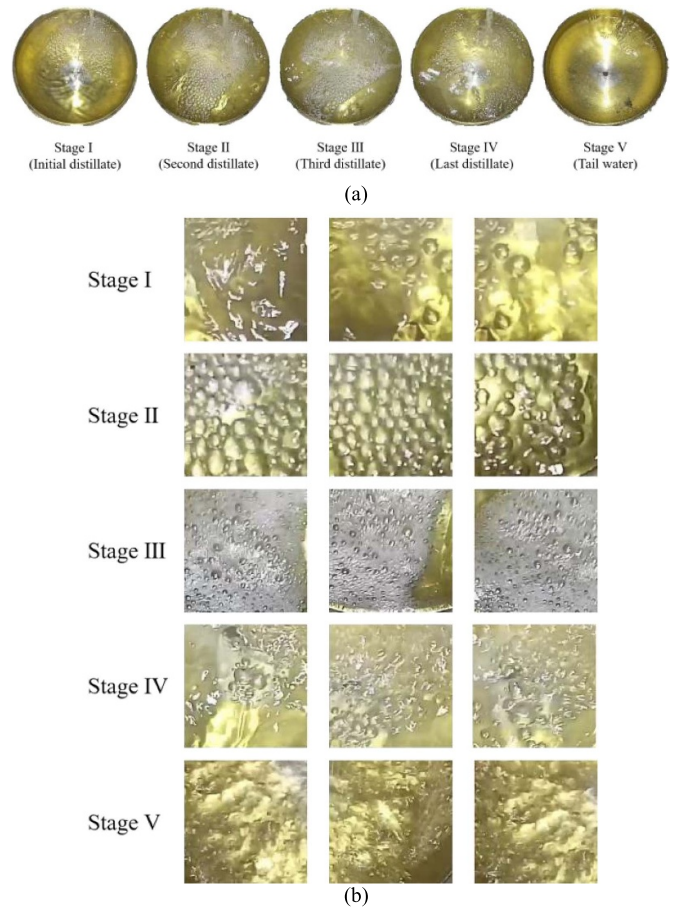


Figure 2. (a) Overall spirits bubble images in different distillation stages. (b) Details of spirits bubble images in different distillation stages.

2.2. Principles of spirits quality classification

The overall and partial details of distillate bubbles at different distillation stages are shown in figure 2. The golden bowl is used to collect the distillate to highlight the bubble morphology. Among the various substances in the distillate, the content of ethanol has the greatest influence on the visual morphology of bubbles. At 30 °C, the relationship between the surface tension of the ethanol solution σ and the ethanol concentration c is provided [20]:

$$\sigma = \frac{0.2243}{3.182 + c^{0.6899}} \quad (1)$$

As shown in figure 1(a), the alcohol content of the distillate decreases, and its surface tension increases as the distillation progress. The continued increase in surface tension leads to a decrease in bubble stability. Therefore, there are more bubbles at stages I, II, and III, as shown in figure 2(a). When entering stages IV and V, the sudden drop in alcohol content leads to a sharp increase in surface tension and a sharp deterioration in bubble stability, resulting in a sharp decrease in bubble number.

Alcohol also indirectly affects the visual characteristics of bubbles by affecting the content and solubility of trace substances in the distillate, as shown in figure 2(b). The distillate at stage I is high in alcohol content and low in water-soluble foam substances such as organic acids and ethyl lactate. At this stage, the amount of adsorbent molecules is sufficient to maintain the stability of bubbles only when the bubbles coalesce to a certain extent and the total surface area of bubbles is small. Therefore, the bubble diameter at this stage is relatively large and can last for a few seconds. In stage II, the alcohol content of the distillate begins to decrease, the content of water-soluble organic acids, ethyl lactate, and other substances increases, and the foaming effect is pronounced, resulting in fine and dense bubbles. During the last two distillation stages, the alcohol content suddenly drops, the content of the foaming substances is too high, and the bubble stability is greatly reduced. Moreover, the higher content of alcohols, ethyl oleate, and ethyl linoleate in the last two distillation stages negatively affects the formation of bubbles as esters and other poorly water-soluble liquids substances will precipitate and form droplets, thereby preventing the formation of bubbles.

In short, the ethanol content in the distillate directly affects the visual characteristics of bubbles by affecting the surface tension. The ethanol content also affects the content and solubility of various trace substances in the distillate, and indirectly affects the visual characteristics of bubbles.

3. The proposed classification method

3.1. Hardware system

An intelligent spirits quality classification system is developed, and the schematic diagram is shown in figure 3. The system is mainly composed of the high-speed camera (the model is HIKVISION DS-2CD7A47FWD-LZS/ZJ, the shutter speed is 1/2000 s, and the frame rate is 25 fps), industrial Ethernet, AI server (Model: Dell PowerEdge R740, CPU: 1x Intel Xeon Silver 4114 2.2 GHz 10Core, GPU: 1x NVIDIA Tesla P4 GPU), distributed control system (DCS) (Model: SUPCON ECS700), and segment switching devices (valves), as well as multiple industrial application software (OPC, Visual Field, etc). The system can obtain real-time images of spirits bubbles, judge the current spirits quality, and control valves to collect spirits from different distillation stages.

The working process of the whole distillation system is as follows: (a) steam first enters the distiller from the bottom, which drives the alcohol substances to be distilled from the distiller's grains in the form of steam, then enters the condenser. (b) In the condenser, the steam condenses into liquid spirits, which then flow out of the bottom of the condenser. (c) Bubbles are formed after the spirits flow into the golden bowl collector. The camera captures the bubble image in real-time and sends it to the server. The server runs the spirits quality classification algorithm and sends the result to the DCS. DCS controls the opening and closing of valves according to the classification result so that liquor of different quality flows to

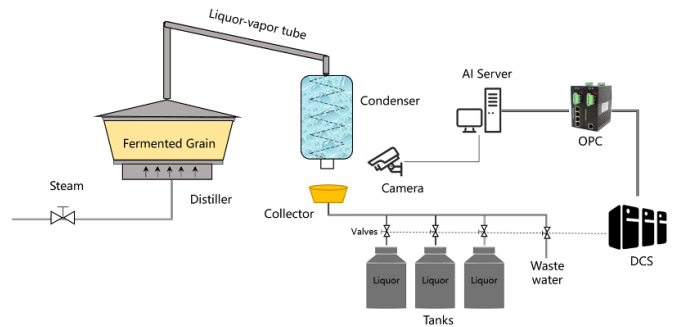


Figure 3. Schematic diagram of spirits distillation system.

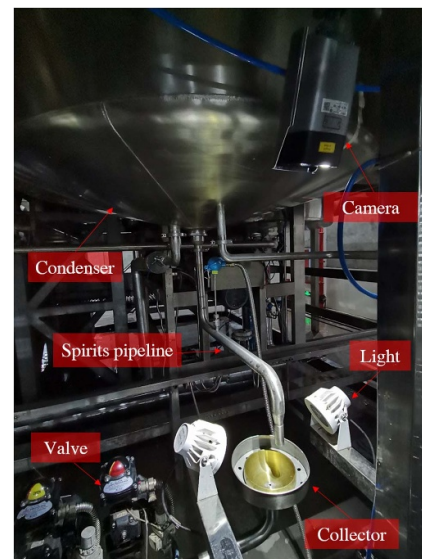


Figure 4. Detailed structure of spirits quality classification system.

the corresponding storage tank. The detailed structure of the spirits quality classification system is shown in figure 4.

3.2. Classification algorithm

The flowchart of the proposed CNN-based method for spirits quality classification is shown in figure 5, which includes three steps: image preprocessing, image classification, and post-processing. The image preprocessing algorithm uses traditional machine vision methods to quickly extract the foreground region of the collector from the original image obtained by the camera. The image classification method first extracts the bubble ROI, and then classifies the image based on the CNN. The post-processing algorithm detects and corrects the classification results of CNN based on expert knowledge to eliminate unreasonable stage switching, and then outputs the final classification results.

3.2.1. Prior knowledge constraints. The quality classification of spirits is closely related to the actual production process. Therefore, this section summarizes some prior empirical knowledge as auxiliary information for the classification

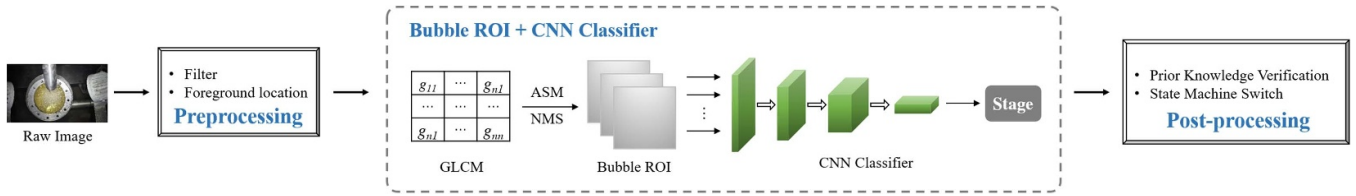


Figure 5. The proposed spirits quality classification algorithm.

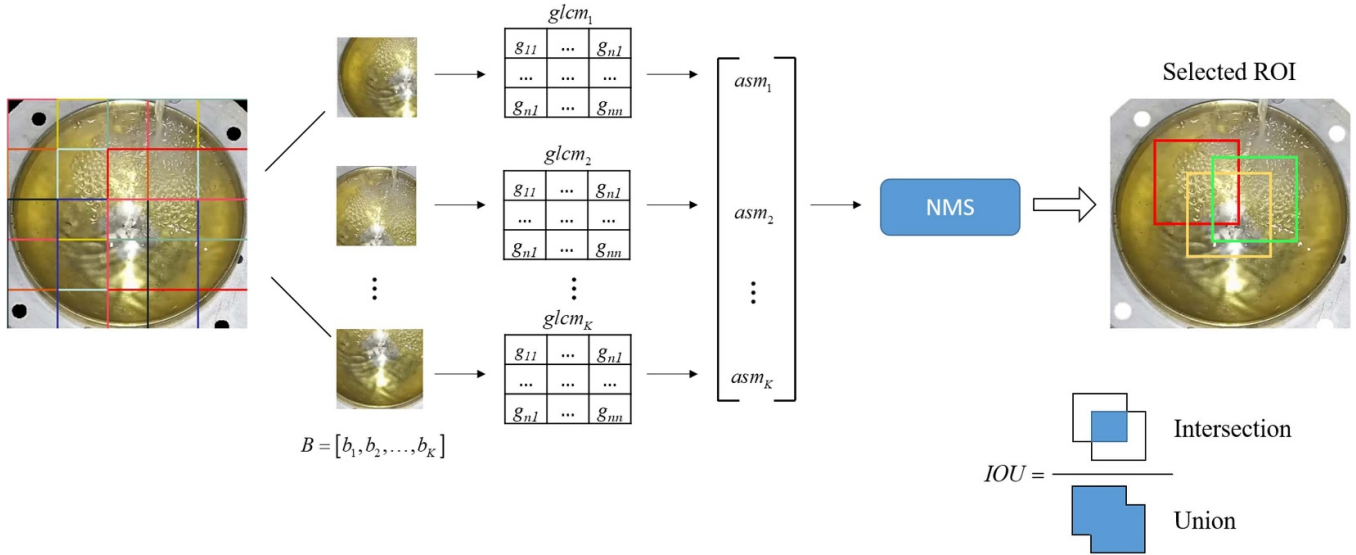


Figure 6. The proposed bubble ROI selection algorithm.

algorithm, as follows: (a) the sequence of the distillation process must be from stage I to stage V, and there will be no disorder; (b) stage I (initial distillate) usually lasts about 2 min, and the algorithm can identify this stage according to the distillation time; (c) due to the low quality of stage IV (last distillate), the distillation time of this stage is usually determined by the brewery, which means that the identification of stage V is also related to distillation time; (d) the classification algorithm in this study only needs to classify three stages of the distillation process: stages II, III, and IV.

3.2.2. Image preprocessing. The useful foreground is only the region where the golden collector is located in the original image. Finding the foreground region and discarding the irrelevant background can speed up the inference speed of the CNN and reduce the input of noise information. Besides, during the production process, the camera and collector may shake or move due to environmental vibration, and foreground position in the image will also change accordingly. Therefore, it is necessary to detect the foreground region of the collector in the image.

The Gaussian filter is used to remove noise in the original image. A golden collector is specially designed to enhance the contrast of the bubble image. Since the golden color of the collector is different from the background color (as shown in figure 6), the color space segmentation method is adopted. Specifically, the original image located in the RGB color space

is converted to the hue, saturation and value (HSV) color space [21]. Calculate the three color channels of the hue (h), saturation (s), and value (v) of the pixels inside and outside the collector. The conditions of threshold segmentation are

$$\begin{cases} h_{\min} \leq h \leq h_{\max} \\ s_{\min} \leq s \leq s_{\max} \\ v_{\min} \leq v \leq v_{\max} \end{cases} \quad (2)$$

where h_{\min} , s_{\min} , and v_{\min} are the lower thresholds of three color channels, and h_{\max} , s_{\max} , and v_{\max} are the upper thresholds. All thresholds are determined based on prior knowledge. Keep all pixels that meet the thresholds in equation (2) while setting other pixels to 0. The last step is to crop out all the zero value regions used for contour detection. Then input the foreground region of the collector into the CNN model for inference.

3.2.3. CNN models for bubble image classification. In this section, a total of 15 different CNN backbone models with different structures are utilized for classification, including five deep models (VGG16 [22], VGG19 [22], ResNet50 [23], InceptionV3 [24, 25], DenseNet121 [26], Xception [27, 28]) and nine lightweight models (SqueezeNet [29], ShuffleNet [30], NasNet_Mobile [31], MobileNet [30], NobileNetv2 [32], MobileNetv3_Small [33], MobileNetv3_Large [34], EfficientNetB0 [35], EfficientNetB3 [35]). All CNN models in

this paper are trained using pre-trained weights to speed up the training process. The pre-trained weights and architecture of all CNN models are derived from the CNN backbone models trained on the ImageNet dataset [36]. In this paper, the number of neurons in the final output layer is changed to 3, and other model structures are consistent with corresponding backbone models. The network structure information and training curves of each CNN model are also provided on the open source dataset webpage (<https://github.com/MengchiCMC/Chinese-Spirits-Bubble-Datasets.git>).

3.2.4. Bubble ROI selection algorithm. Since more convolution operations can extract deeper semantic features, deeper and larger CNN models can theoretically achieve higher image classification accuracy and have better generalization. However, a complex CNN model with many parameters requires better hardware equipment and more training data, which will undoubtedly increase the difficulty and cost of the classification algorithm. A bubble ROI selection algorithm based on GLCM and NMS is proposed to improve the accuracy and generalization of the lightweight CNN model. After combining the bubble ROI selection algorithm, the CNN models no longer need to learn global images but accept local bubble images as inputs. The proposed bubble ROI selection algorithm is shown in figure 6, which consists of three steps: Grid division, GLCM propriety calculation, and NMS. The detailed description of each step is as follows.

3.2.4.1. Grid division. First, assume that the global image resolution is $W \times H$, and the resolution of each bubble ROI is $w \times h$. Then, divide the global image into multiple grids at equal intervals in the row and column directions. The division step is s , and the list of candidate grid boxes is $B = [b_1, b_2, \dots, b_K]$. The element $b_i (i = 1, \dots, K)$ is the mark of each grid, and K is the total number of grids, which can be calculated as

$$K = \left(\frac{(W-w)}{s} + 1 \right) \times \left(\frac{(H-h)}{s} + 1 \right). \quad (3)$$

3.2.4.2. GLCM propriety calculation. The GLCM is a matrix that describes the gray-level relationship between pixels and can be used for image texture analysis [37]. The calculation of GLCM needs to set three parameters, namely gray quantization level l , direction A , and distance D .

First, assume that $A = [a_1, a_2, \dots, a_m]$ and $D = [d_1, d_2, \dots, d_n]$. Each element in A and D is the selected value of direction and distance, m and n are the elements of A and D , respectively. Then, calculate the GLCM with size $w \times h$, and K matrices of size $l \times l \times m \times n$ can be generated, which are the matrices labeled as $g_{lcm_1, \dots, K}$ in figure 6. In this paper, only one direction and one angle are selected to calculate the GLCM, and the angular second moment (ASM) is used as the texture feature index of each matrix. ASM can describe the uniformity of gray distribution and the thickness of texture. If the value of each element in the gray matrix fluctuates very little, then the value of index is small, otherwise it is large.

Therefore, the ASM of the bubble ROI will be larger than other regions without bubbles. ASM can be calculated as

$$ASM = \sum_i \sum_j \{P(i,j)\}^2 \quad (4)$$

where $P(i,j)$ is the element of GLCM with coordinates (i,j) . Every element in the evaluation parameter set $ASM = [asm_1, asm_2, \dots, asm_K]$ corresponds to the elements of B in figure 6.

3.2.4.3. NMS. The NMS is a type of algorithm utilized to select an entity (e.g. bounding box) from many overlapping entities. This paper uses the NMS to select the most suitable k bubble regions with the K candidate grids and their corresponding ASM indexes. The NMS selects the proposal regions with the highest index score, and the intersection over union (IOU) between selections should also be larger than the IOU threshold. Finally, k bubble ROIs are selected and marked as $R = [r_1, r_2, \dots, r_k]$.

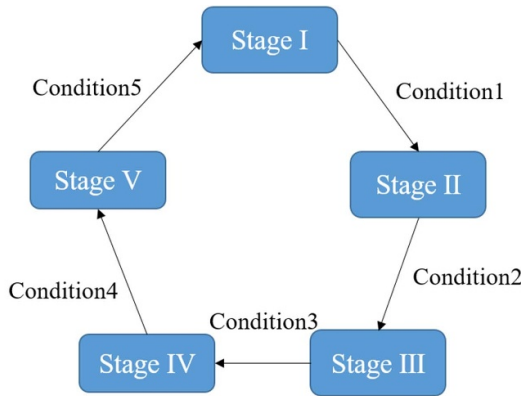
The following pseudo-code is the detailed bubble ROI selection process.

Algorithm 1. Bubble ROI selection

- Input: $B = [b_1, b_2, \dots, b_K]$, N_t
 B is the list of grid boxes
 N_t is the IOU threshold
 - Output: Selected bubble ROI $R = [r_1, r_2, \dots, r_k]$
 - Main loop:
 - 1: Calculate GLCM of each box in B ,
 GLCM = $[g_{lcm_1}, g_{lcm_2}, \dots, g_{lcm_K}]$
 - 2: Calculate ASM index of each GLACM,
 ASM = $[asm_1, asm_2, \dots, asm_K]$
 - 3: $R \leftarrow \emptyset$
 - 4: While $B \neq$ empty do
 - 5: $m \leftarrow \arg \max ASM$
 - 6: $R \leftarrow R \cup b_m$
 - 7: $B \leftarrow B - b_m$
 - 8: For b_i in B do
 - 9: If $IOU(b_m, b_i) > N_t$ then
 - 10: $B \leftarrow B - b_i$
 - 11: $ASM \leftarrow ASM - asm_i$
 - 12: End if
 - 13: End for
 - 14: End while
 - 15: Return R
 - End loop
-

3.2.4.4. CNN model inputs. Based on the bubble ROI selection algorithm, sub-images can be selected to maximize the extraction of bubble features while suppressing background noise. Unlike the classification method directly inputs the global image, the classification method combined with bubble ROI selection takes all the selected bubble sub-images as input. Due to the full features of the input bubble sub-images, there is less demand for the depth of the CNN model.

3.2.4.5. Post-processing. As shown in figure 4, the classification results of the proposed classification method must be



- Condition1: Duration reaches the set value (2 minutes) && CNN model outputs stage II continuously for n times
- Condition2: CNN model outputs stage III continuously for n times
- Condition3: CNN model outputs stage IV continuously for n times
- Condition4: Duration reaches the set value
- Condition5: New classification cycle

Figure 7. State machine switching rules for post-processing.

post-processed. When the distillation process and the distillate flow rate become unstable, the output result of the classification algorithm may be unreasonable. To further improve the performance of the classification algorithm, a post-processing method based on the state machine switching rule and expert knowledge is designed to correct the classification results, as shown in figure 7. In figure 7, there are five switching conditions. When the duration of stage I reaches the threshold based on expert knowledge (2 min in this paper) and the CNN classification results show that n ($n = 5$ in this paper) consecutive diagnosis results are stage II, it means that the distillation process of spirits has entered stage II. The switching rules between other stages are similar.

The post-processing step has two main functions. The first function is that the order of the classification results matches the actual production process. For example, stage III can only appear after stage II, not before stage II. The second function is to filter out incorrect classification results and improve the classification accuracy.

4. Test results and discussions

4.1. Datasets

Twenty-eight sets of spirits classification systems are deployed in a winery in Sichuan, China, as shown in figure 8. There are 28 spirits production lines in figure 8, including 28 distillation units, 28 cameras, 7 switches, 7 AI servers, 1 OPC server, and 1 DCS. Each production line has a separate distillation unit and camera. Each server is connected to four production line cameras and deploys four spirits quality classification algorithms, which operate independently. All servers communicate with the DCS via the OPC communication protocol.

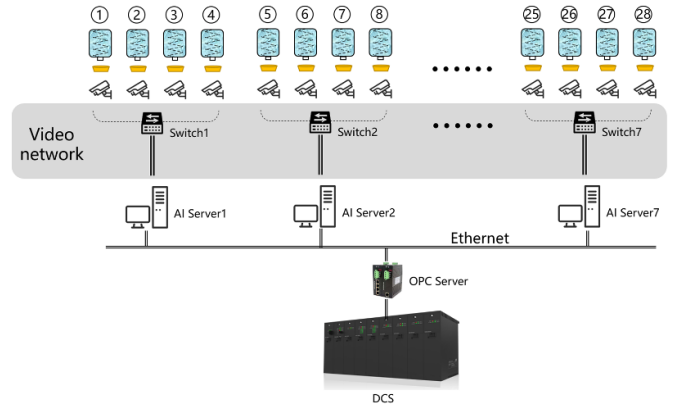


Figure 8. The deployed spirits classification systems.

Table 1. Image numbers of each distillation stage.

Stages	Training set	Test set
Stage II	70 471	4176
Stage III	31 199	4264
Stage IV	26 862	2227
Total	128 532	10 667

In this paper, videos of 13 different spirits production lines are used as datasets and labeled, among which 10 production lines are used as the training dataset and 3 production lines are used as the test dataset. Each video covers the complete spirits distillation process, which includes bubbles images from stage I to stage V. The resolution of videos is 2560×1440 , and the length of each video is about 1 h. The authors sample a frame from the video every 500 ms and manually label it as a dataset sample. Finally, a spirits quality classification dataset containing 139 199 images is established, including a training dataset of 128 532 images and a test dataset of 10 667 images. The image samples of different stages are shown in table 1. According to expert knowledge, stage I and stage V in the spirits production process can be directly identified and controlled by duration. Therefore, the proposed classification method is trained to classify stage II, stage III, and stage IV.

4.2. Preprocessing results

The original frame is an red, green and blue (RGB) image with a resolution of 2560×1440 , and the size of the foreground region is 900×900 , as shown in figure 9(a). The entire image is converted from RGB color space to HSV color space, and the pixel value ranges from 0 to 255. The entire image and foreground region's HSV histograms are shown in figures 9(b) and (c), respectively. Three graphs from left to right are the histogram distribution of hue channel, saturation channel, and value channel. The horizontal axis represents the pixel value of each color channel, and the longitudinal axis represents the normalized probability density.

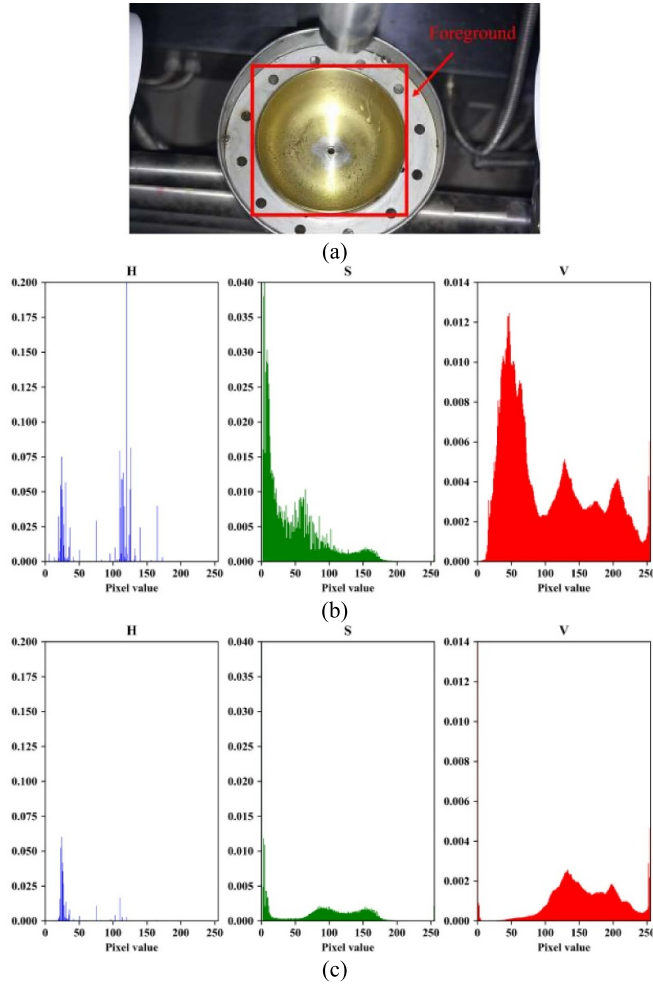


Figure 9. HSV histograms. (a) Entire image and foreground region. (b) HSV histograms of entire image. (c) HSV histograms of foreground region.

By comparing figures 9(b) and (c), it can be found that the hue (H) channel histogram of the entire image is mainly distributed in the range of [10, 40] and [100, 150], while the hue channel histogram of the foreground region only distributed in [15, 40], which means that pixels with values greater than 40 almost all represent the background. Therefore, the hue channel segmentation thresholds can be set to $h_{\min} = 15$ and $h_{\max} = 40$. The saturation (S) channel of the entire image is distributed in the range of [0, 200], with the smaller value region accounting for the most. Unlike the entire image, the middle pixel value occupies most of the foreground saturation histogram, which is almost equal to the obvious tail of the global saturation histogram. The saturation histograms of the entire image and the foreground region overlap from 50 to 200, and some pixels with smaller values also appear in the saturation histogram of the foreground region. After ignoring small value pixels, the saturation channel segmentation threshold can be roughly set to $s_{\min} = 70$ and $s_{\max} = 200$. Different from the hue and saturation histogram distribution, the value histogram of the entire image is distributed within the range of the complete coordinate axis, and there are four peak points located near 44, 127, 206, and 255. The value histogram of the foreground

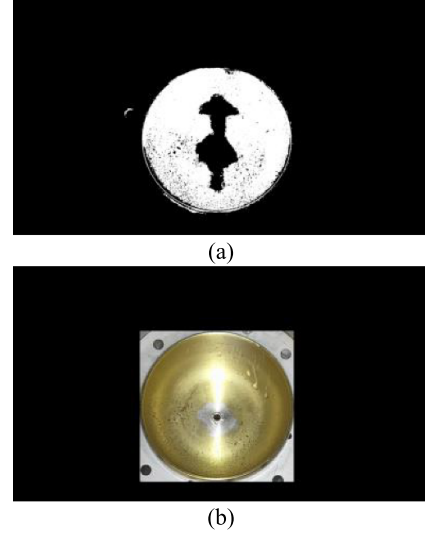


Figure 10. Foreground region extraction results. (a) HSV segmentation mask. (b) Extracted foreground region.

region mainly covers the range from 80 to 255, and there are three peak points around 130, 200, and 255, which are close to the peak points of the entire image. Therefore, the value channel segmentation thresholds can only be set to $v_{\min} = 100$ and $s_{\max} = 255$.

Figure 10(a) shows the mask for HSV color space segmentation of figure 9(a) using the thresholds set above, where the white pixels indicate that the thresholds are met and need to be retained, while the black pixels will be discarded. Due to the reflection of light, it can be found that some parts of the foreground region are no longer golden and are omitted. Therefore, some remedial measures are needed. The morphological opening operations are performed to filter small noise points outside the foreground region, and the closing operation fills the holes in the foreground region. Hough circle detection is used to realize the smallest circumscribed circle. The extraction result of the foreground region is shown in figure 10(b). The golden collector is accurately extracted and used as the input to the CNN model in the following classification step.

4.3. Bubble ROI selection results

For the bubble ROI selection algorithm, two parameters need to be determined when using the GLCM, namely, the pixel-pair distance offset and the pixel-pair angle. The grid shrinkage method is used to select appropriate parameters, and eight different values of $D = [0, 1, 2, 3, 4, 5, 6, 7]$ and $A = [0, \pi/4, \pi/2, 3\pi/4, \pi, 5\pi/4, 3\pi/2, 7\pi/4]$ are selected to calculate the GLCM and ASM indexes of randomly selected foreground images. The calculation results are shown in figure 11. When the distance is fixed, the ASM index increases monotonously as the angle increases. When the angle is fixed, the ASM index fluctuates in a small range as the distance increases. The ASM index is not a convex function or a concave function about the angle and distance, which means that the change of ASM with distance and angle can be predicted.

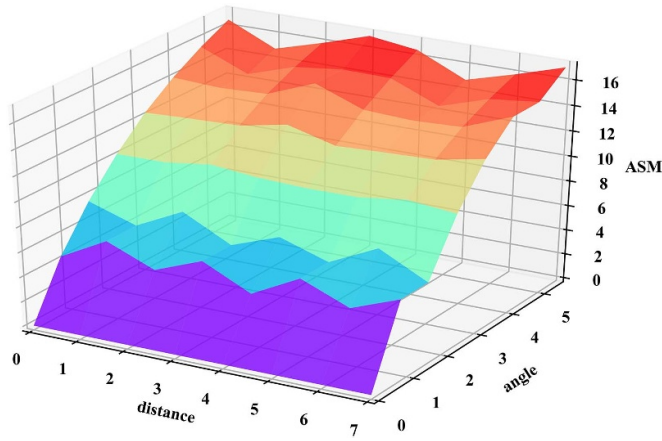


Figure 11. GLCM parameter selection.

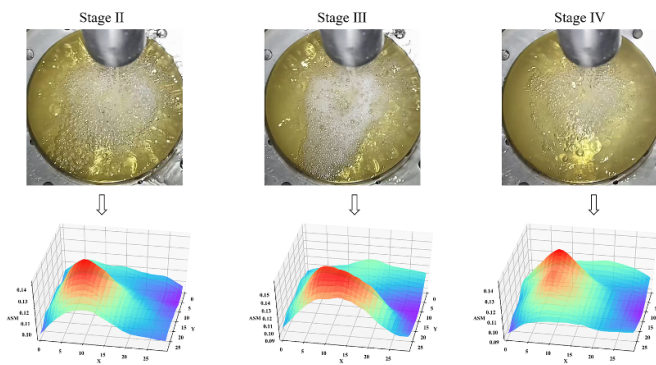


Figure 12. ASM indexes of foreground images at different distillation stages.

Therefore, it is only necessary to select a pair of distance and angle values to calculate the ASM index.

The bubble ROI resolution is set as 224×224 to fit the input size of the CNN model, and the grid division stride is set to $s = 20$. The number of candidate grid boxes, K , can be calculated by the formula (3). The distance parameter of GLCM is set as $D = [1]$, and the angle parameter is set as $A = [\pi/4]$. For three different distillation stages, the ASM index for each grid of the collector foreground image is shown in figure 12. It is worth noting that the horizontal and longitudinal axes refer to the grid number coordinates after division. It is easy to find that the grid with more bubbles has a higher ASM value. Then, NMS is applied to select the k most suitable grids as bubble ROI for CNN model training and testing. In this paper, N_t is set as $N_t = 0.5$, and k is set as $k = 5$.

4.4. Parameter settings

Figure 13 shows the bubble ROI selection corresponding to different value combinations of N_t and k . When the value of N_t is too small ($N_t \leq 0.3$), a large area without bubbles appears in some ROI images (such as ROI r_3 in figures 13(b), (f) and (j)). When the value of N_t is too large ($N_t \geq 0.7$), the ROI images will be severely overlapped, and the significance of selecting multiple ROI images will be lost (such as figures 13(d), (e), (h) and (i)). In order to select as many bubble areas as possible and

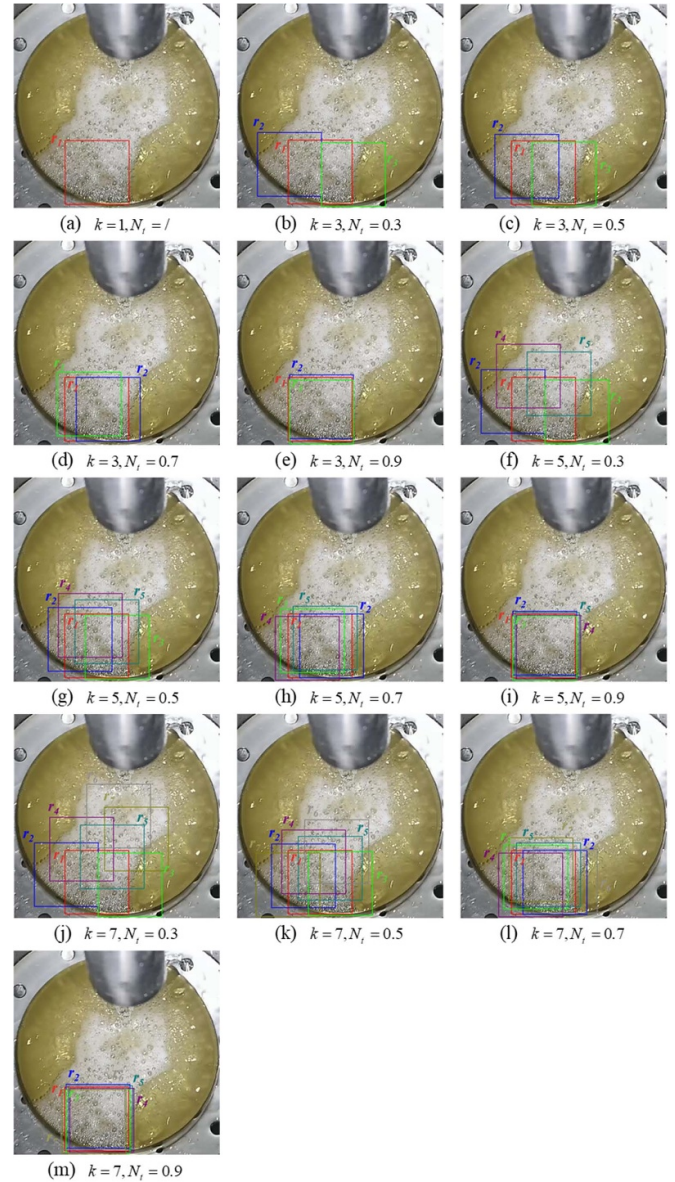


Figure 13. Different parameter value combinations of N_t and k .

at the same time ensure sufficient feature differences between different ROI images, $N_t = 0.5$ is set in this paper.

The setting of parameter k is similar to that of N_t . As shown in figure 13, when the value of k is too small ($k \leq 3$), the number of extracted ROI images is small, and the bubble features may not be rich enough. When the value of k is too large ($k \geq 7$), there are many repeated areas in ROI images, and it may also cover bubble-free area. Therefore, $k = 5$ is set in this paper.

4.5. Classification results

As mentioned above, only three stages (II, III, and IV) need to be classified by CNN, while stages I and V are based on expert knowledge. An NVIDIA TITAN XP GPU is utilized for training the CNN model, and hyper-parameters are set as: the input image size is $224 \times 224 \times 3$, the batch size is 64, the

Table 2. Classification results of different CNN models.

CNN models	Accuracy of global image input (1)	Accuracy of bubble ROI input (2)	Accuracy improvement (2) – (1)	Total parameters	FLOPs (G)	Model size (MB)	Inference time (ms)
VGG16	89.80%	95.28%	5.47%	134 272 835	30.95	1536.78	
VGG19	91.26%	96.59%	5.33%	139 582 531	39.28	1675.17	63.84
Resnet50	88.25%	96.77%	8.52%	23 888 771	7.75	273.96	75.96
InceptionV3	90.61%	96.90%	6.29%	21 956 387	5.69	252.37	88.47
DenseNet121	94.94%	96.70%	1.76%	7188 035	5.7	83.65	98.96
Xception	93.16%	98.15%	4.99%	21 162 539	9.13	242.52	74.26
SqueezeNet	75.25%	83.32%	8.07%	1251 427	1.78	14.57	63.97
ShuffleNet	80.72%	92.80%	12.08%	940 635	1.94	24.33	94.88
NasNet_Mobile	93.13%	96.66%	3.53%	4424 951	1.15	54.09	132.81
MobileNet	91.33%	92.10%	0.77%	3379 395	1.15	38.95	59.22
MobileNetv2	92.81%	93.75%	0.94%	2446 147	0.61	28.58	72.73
MobileNetv3_Small	95.01%	98.62%	3.60%	1533 043	0.12	18.46	69.22
MobileNetv3_Large	95.74%	98.15%	2.41%	4230 275	0.45	49.44	74.2
EfficientNetB0	85.28%	88.31%	3.03%	4237 734	0.79	49.3	76.34
EfficientNetB3	84.41%	86.96%	2.55%	11 244 338	1.97	127.14	97.34

Table 3. Overfitting test results.

Initialization	Standardization	L2 regularization	Dropout	Random flip	Random brightness and contrast	Accuracy
Random						89.72%
Pre-trained						98.62%
Pre-trained	✓					97.23%
Pre-trained		✓				97.03%
Pre-trained			✓			96.06%
Pre-trained				✓		96.04%
Pre-trained					✓	97.70%
Pre-trained		✓	✓			97.50%
Pre-trained	✓	✓	✓			97.47%
Pre-trained		✓	✓	✓		95.70%
Pre-trained		✓	✓	✓	✓	97.89%
Pre-trained	✓	✓	✓	✓		97.88%
Pre-trained	✓	✓	✓	✓	✓	97.20%

epoch is 5, the optimization algorithm is Adadelta [38], the learning rate is 1×10^{-2} , and the decay rate is 1×10^{-6} . The categorical cross-entropy loss function is selected, and test results are shown in table 2. The bold values in table 2 represent the results with the highest accuracy in different experiments.

As shown in table 2, the bubble ROI selection method can improve classification accuracy compared with the global image input. The deep model Dense121 and the lightweight model MobileNetv3_Large can achieve better classification results when using the global image as input. When using bubble ROI input, the classification accuracy of all CNN models is improved compared to the global image input. The deep model Xception and the lightweight model MobileNetv3_Small can achieve better classification results. Among the five deep models, Resnet50 has the lowest classification accuracy of 88.25% when using the global image input. However, the accuracy of the bubble ROI input reached 96.77%, which is an 8.51% improvement over the global image input. The meaning of 8.51% improvement is the difference obtained by subtracting the classification accuracy of the original method from the classification accuracy of the improved method.

Among the nine lightweight models, SqueezeNet has the lowest classification accuracy of 75.25% when using the global image input. Similarly, the accuracy of the bubble ROI input reaches 83.32%, which is an 8.07% improvement. Generally, the bubble ROI selection method can greatly improve the classification accuracy of the CNN model when the accuracy obtained by global image input is poor. Using bubble ROI input for CNN models with higher accuracy obtained by global image input can still improve the accuracy, but the improvement will be reduced. At the same time, the MobileNetv3_Small model also has advantages in terms of total parameters, floating point operations (FLOPs), model size, and inference time.

In addition, the overfitting problem of CNN training is also tested by using ablation experiments. The data augmentation (including normalization, random flipping, random brightness and contrast adjustment), regularization (L_2 regularization), and dropout methods are performed while training the MobileNetv3_Small model with the bubble ROI input, as shown in table 3. It can be found that the classification accuracy of the MobileNetv3_Small model using pre-trained weights (transfer learning) is 8.9% higher than that of the model using

Table 4. Post-processing results of different CNN models.

CNN models	Accuracy of global image input without post-processing (1)	Accuracy of global image input with post-processing (2)	Accuracy improvement (2) – (1)	Accuracy of bubble ROI input without post-processing (3)	Accuracy of bubble ROI input with post-processing (4)	Accuracy improvement (4) – (3)
VGG16	89.80%	95.20%	5.40%	95.28%	99.61%	4.33%
VGG19	91.26%	96.35%	5.09%	96.59%	99.71%	3.12%
Resnet50	88.25%	93.18%	4.93%	96.77%	99.66%	2.90%
InceptionV3	90.61%	96.82%	6.22%	96.90%	99.73%	2.83%
DenseNet121	94.94%	99.18%	4.24%	96.70%	99.78%	3.08%
Xception	93.16%	99.12%	5.96%	98.15%	99.78%	1.64%
SqueezeNet	75.25%	87.72%	12.47%	83.32%	99.45%	16.12%
ShuffleNet	80.72%	90.68%	9.96%	92.80%	99.63%	6.83%
NasNet_Mobile	93.13%	97.69%	4.57%	96.66%	99.75%	3.09%
MobileNet	91.33%	96.05%	4.72%	92.10%	99.79%	7.69%
MobileNetv2	92.81%	96.96%	4.16%	93.75%	99.80%	6.06%
MobileNetv3_Small	95.01%	97.86%	2.85%	98.62%	99.82%	1.20%
MobileNetv3_Large	95.74%	97.93%	2.18%	98.15%	99.82%	1.67%
EfficientNetB0	85.28%	94.24%	8.96%	88.31%	99.59%	11.27%
EfficientNetB3	84.41%	92.84%	8.43%	86.96%	99.54%	12.58%

the random weights (random initialization). Data augmentation, regularization, and dropout methods slightly reduce classification accuracy. The lowest accuracy (95.7%) is obtained by combining L_2 regularization, dropout and random flipping methods. After combining L_2 regularization, dropout, random flipping, and random brightness and contrast adjustment methods, the MobileNetv3_Small model has the least drop in accuracy (only 0.72%). The results of the ablation experiments in table 3 show that the accuracy of the MobileNetv3_Small model is not greatly reduced after using various methods to prevent overfitting, which verifies that there is no serious overfitting problem.

4.6. Post-processing results

The post-processing results of different CNN models are shown in table 4. The bold values in table 4 represent the results with the highest accuracy in different experiments. The comparison results show that the classification accuracy of CNN models can be effectively improved by using the post-processing method. After using post-processing, the classification accuracy obtained by CNN models with bubble ROI input is generally higher than CNN models with global image input. The smaller the classification accuracy of the CNN model before using post-processing, the more significant the improvement of accuracy after post-processing is used. After using post-processing, the classification accuracy of CNN models with Bubble ROI input exceeded 99%.

The post-processing results of the MobileNetv3_Small model with the bubble ROI input are shown in figure 14. It can be found that the misclassification results in the original stage III and stage IV are corrected after using post-processing.

As shown in table 2, compared with VGG16, VGG19 achieves 1.46% and 1.32% higher accuracy on global image input without post-processing and bubble ROI input without post-processing, respectively. Similarly, VGG19

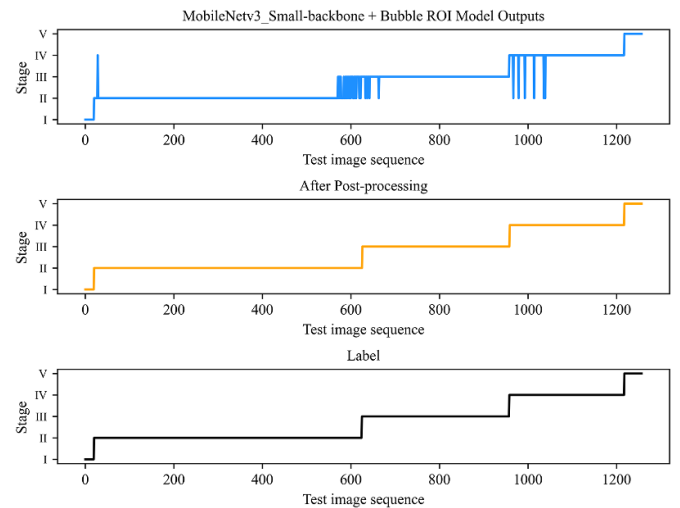


Figure 14. The post-processing results of the MobileNetv3 small model with the bubble ROI input.

has higher classification accuracy than VGG16 after using post-processing, as shown in table 4. VGG19 improves the accuracy of global image input with post-processing and bubble ROI input with post-processing by 1.15% and 0.10%, respectively. Although the classification accuracy of VGG19 is better than that of VGG16, its performance does not outperform CNN models such as Xception, NasNet Mobile and MobileNetv3_Small. Therefore, in this manuscript, VGG19 is not the preferred model for practical application deployment.

In general, the deeper the classification model based on the CNN architecture, the higher the image classification accuracy. However, there does not appear to be a similar relationship between model computational complexity (operations), model complexity (number of parameters), and image classification accuracy. Bianco *et al* [39] summarizes the relationship between image classification accuracy, model computational

Table 5. Resources consumed by the server to run the classification algorithm.

Time consumption	CPU usage	GPU usage	Memory usage	GPU memory usage	Network usage
110 ms	11%	4%	700 MB	279 MB	16 Mbps

Table 6. Resources consumed by the server to run the classification algorithm.

Samples	Stage I to stage II	Stage II to stage III	Stage III to stage IV	Stage IV to stage V
1	79.7	63.2	46.6	11.9
2	78.5	63.8	46.9	12.2
3	79.2	64.4	45.8	15.0
4	78.3	62.3	46.0	12.1
5	79.8	64.8	44.7	13.8
6	79.5	64.1	46.5	13.4
7	78.0	62.5	46.1	13.4
8	79.8	63.9	46.7	14.6
9	79.4	64.7	45.4	14.0
10	78.8	63.7	46.6	12.7
Mean	79.1	63.7	46.1	13.3
Standard deviation	0.6	0.8	0.6	1.0

complexity, and model complexity. The comparison results in [39] show that the model computational complexity, the number of model parameters, and the image classification accuracy are not simply proportional. For example, in [39], VGG-13 and ResNet-18 have similar accuracy but differ greatly in computational complexity and number of parameters. There is a similar comparison between SENet-154 and SE-ResNeXt-101 ($32 \times 4d$) in [39].

The main reason for a large number of parameters in the VGG model is that the last three cascaded fully connected layers (FC-4096/FC-4096/FC-1000) [22] bring a huge amount of parameters. For example, the parameters of the first fully connected layer (FC-4096) in VGG16 account for about 83% of the entire model parameters. Unlike VGG models, other models usually only have a small fully connected layer (FC-1000). Deeper convolutional layers improve image classification accuracy rather than more fully connected layers. Therefore, among all the models compared in this paper, although the VGG16 model has the largest number of parameters, it is not the model with the best accuracy, which is consistent with the conclusions of other literature. At the same time, the main reason for the shortest inference time of VGG16 is that the convolution layer is not deep, and the computational complexity is low. In conclusion, the fewer convolutional layers, the shorter the inference time and the lower the image classification accuracy. The more fully connected layers, the larger the number of parameters.

4.7. Practical application test results

In practical application tests, the MobileNetv3_Small model with bubble ROI selection and post-processing methods is deployed in spirits production lines. The running period of the classification algorithm is 500 ms. The parameters of the bubble ROI selection method are $s = 20$, $D = [1]$, $A = [\pi/4]$, $N_t = 0.5$, $k = 5$. For the post-processing method, the duration

of stage I and stage VI are 2 min and 30 min, respectively. Based on the above parameter settings, resources consumed by the server to run the classification algorithm are shown in table 5.

Ten spirits samples from different production lines are randomly selected for alcohol content test when the distillation stage is switched. Test results are shown in table 6. After applying the classification solution proposed in this paper, spirits switched from stage II to stage III has an average alcohol content of 63.7% VOL with a standard deviation of 0.8% VOL, and has an average alcohol content of 46.1% VOL with a standard deviation of 0.6% VOL when switching from stage II to stage III. Test results in TABLE VI verify that the proposed solution is accurate and stable in practical spirits production applications. One month after the deployment and operation of the spirits quality classification system, the output increased by 7.7%, and the production efficiency increased by 29.0%.

5. Conclusions

This paper proposes a spirits quality classification solution that combines machine vision technology and expert experience, which can classify spirits at different distillation stages by analyzing the bubble morphology. A spirits quality classification dataset is established and labeled based on the self-designed classification system and actual spirits production videos. A complete spirits quality classification algorithm from pre-processing to post-processing is proposed, and 15 different CNN models are trained and tested. Test results show that the proposed bubble ROI selection and post-processing method can effectively improve the classification accuracy of spirits quality. After using the bubble ROI selection method, the highest spirits quality classification accuracy can reach 98.62%. After adopting the post-processing method, the highest classification accuracy reached 99.82%. Furthermore, the practical application tests show that the solution proposed in this paper

can significantly improve spirits' final quality and production efficiency.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: <https://github.com/MengchiCMC/Chinese-Spirits-Bubble-Datasets>.

Acknowledgments

This work was supported by the National Key R&D Program of China (2018YFB1702200).

ORCID iD

Mengchi Chen  <https://orcid.org/0000-0001-5507-3141>

References

- [1] Lee J, Davari H, Singh J and Pandhare V 2018 Industrial artificial intelligence for industry 4.0-based manufacturing systems *Manuf. Lett.* **18** 20–23
- [2] Ding H, Gao R X, Isaksson A J, Landers R G, Parisini T and Yuan Y 2020 State of AI-based monitoring in smart manufacturing and introduction to focused section *IEEE/ASME Trans. Mechatronics* **25** 2143–54
- [3] Ai M, Xie Y, Tang Z, Zhang J and Gui W 2020 Two-stream deep feature-based froth flotation monitoring using visual attention clues *IEEE Trans. Instrum. Meas.* **70** 1–14
- [4] Fu Y and Liu Y 2018 3D bubble reconstruction using multiple cameras and space carving method *Meas. Sci. Technol.* **29** 075206
- [5] Knüpfer L and Heitkam S 2022 A machine learning approach to determine bubble sizes in foam at a transparent wall *Meas. Sci. Technol.* **33** 067001
- [6] Zheng X-W and Han B-Z 2016 Baijiu, chinese liquor: history, classification and manufacture *J. Ethn. Foods* **3** 19–25
- [7] Xu Y, Sun B, Fan G, Teng C, Xiong K, Zhu Y, Li J and Li X 2017 The brewing process and microbial diversity of strong flavour Chinese spirits: a review *J. Inst. Brew.* **123** 5–12
- [8] Mandrile L, Zeppa G, Giovannozzi A M and Rossi A M 2016 Controlling protected designation of origin of wine by Raman spectroscopy *Food Chem.* **211** 260–7
- [9] Wu Z, Xu E, Long J, Wang F, Xu X, Jin Z and Jiao A 2015 Measurement of fermentation parameters of Chinese rice wine using Raman spectroscopy combined with linear and non-linear regression methods *Food Control* **56** 95–102
- [10] Prez-Caballero G, Andrade J M, Olmos P, Molina Y, Jimnez I, Durn J, Fernandez-Lozano C and Miguel-Cruz F 2017 Authentication of tequilas using pattern recognition and supervised classification *TRAC Trends Anal. Chem.* **94** 117–29
- [11] Liu J, Gao Q, Tang Z, Xie Y, Gui W, Ma T and Niyoyita J P 2020 Online monitoring of flotation froth bubble-size distributions via multiscale deblurring and multistage jumping feature-fused full convolutional networks *IEEE Trans. Instrum. Meas.* **69** 9618–33
- [12] Peng C, Liu Y, Gui W, Tang Z and Chen Q 2022 Bubble image segmentation based on a novel watershed algorithm with an optimized mark and edge constraint *IEEE Trans. Instrum. Meas.* **71** 1–10
- [13] Haas T, Schubert C, Eickhoff M and Pfeifer H 2020 BubCNN: bubble detection using Faster RCNN and shape regression network *Chem. Eng. Sci.* **216** 115467
- [14] Qaddoori A S, Saud J H and Hamad F A 2021 A classifier design for micro bubble generators based on deep learning technique *Mater. Today Proc.* (<https://doi.org/10.1016/j.matpr.2021.07.013>)
- [15] Cerqueira R F L and Paladino E E 2021 Development of a deep learning-based image processing technique for bubble pattern recognition and shape reconstruction in dense bubbly flows *Chem. Eng. Sci.* **230** 116163
- [16] Wang H, Wang H, Zhu X, Song L, Guo Q and Dong F 2021 Three-dimensional reconstruction of dilute bubbly flow field with light-field images based on deep learning method *IEEE Sens. J.* **21** 13417–29
- [17] Wang D, Dai W and Tang D 2022 Deep learning approach for bubble segmentation from hysteroscopic images *Med. Biol. Eng. Comput.* **60** 1613–26
- [18] Wang L, Hu G, Lei L, Lin L, Wang D and Wu J 2016 Identification and aroma impact of volatile terpenes in Moutai liquor *Int. J. Food Prop.* **19** 1335–52
- [19] Li H, Huang W, Shen C and Yi B 2012 Optimization of the distillation process of Chinese liquor by comprehensive experimental investigation *Food Bioprod. Process.* **90** 392–8
- [20] Bardenhagen S, Brydon A and Guilkey J 2005 Insight into the physics of foam densification via numerical simulation *J. Mech. Phys. Solids* **53** 597–617
- [21] Chernov V, Alander J and Bochko V 2015 Integer-based accurate conversion between RGB and HSV color spaces *Comput. Electr. Eng.* **46** 328–37
- [22] Simonyan K and Zisserman A 2014 Very deep convolutional networks for large-scale image recognition (arXiv:1409.1556) pp 1–14
- [23] He K, Zhang X, Ren S and Sun J 2016 Deep residual learning for image recognition *Proc. IEEE Conf. on Computer Vision and Pattern Recognition* pp 770–8
- [24] Szegedy C, Vanhoucke V, Ioffe S, Shlens J and Wojna Z 2017 Rethinking the inception architecture for computer vision *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (Honolulu, HI, USA)* (<https://doi.org/10.48550/arXiv.1512.00567>)
- [25] Muhammad W and Supavadee A 2019 Multi-scale inception based super-resolution using deep learning approach *Electronics* **8** 892
- [26] Huang G, Liu Z, van der Maaten L and Weinberger K Q 2017 Densely connected convolutional networks *Proc. IEEE Conf. on Computer Vision and Pattern Recognition* pp 2261–9
- [27] Chollet F 2016 Xception: deep learning with depthwise separable convolutions *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (Las Vegas, NV, USA)* (<https://doi.org/10.1109/CVPR.2017.195>)
- [28] Muhammad W, Supavadee A and Takao O 2021 Multi-scale Xception based depthwise separable convolution for single image super-resolution *PLoS One* **16** e0249278
- [29] Iandola F N, Han S, Moskewicz M W, Ashraf K, Dally W J and Keutzer K 2016 SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and 0.5MB model size (arXiv:1602.07360)
- [30] Zhang X, Zhou X, Lin M and Sun J 2017 ShuffleNet: an extremely efficient convolutional neural network for mobile devices (arXiv:1707.01083)
- [31] Zoph B, Vasudevan V, Shlens J and Le Q V 2017 Learning transferable architectures for scalable image recognition (arXiv:1707.07012)
- [32] Howard A G, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M and Adam H 2017 MobileNets: efficient convolutional neural networks for mobile vision applications (arXiv:1704.04861)

- [33] Sandler M, Howard A, Zhu M, Zhmoginov A and Chen L C 2018 MobileNetV2: inverted residuals and linear bottlenecks (arXiv:[1801.04381](https://arxiv.org/abs/1801.04381))
- [34] Howard A, Sandler M, Chu G, Chen L C, Chen B, Tan M, Wang W, Zhu Y, Pang R and Vasudevan V 2019 Searching for MobileNetV3 (arXiv:[1905.02244](https://arxiv.org/abs/1905.02244))
- [35] Tan M and Le Q V 2019 EfficientNet: rethinking model scaling for convolutional neural networks (arXiv:[1905.11946](https://arxiv.org/abs/1905.11946))
- [36] Jia D, Wei D, Richard S, Li-Jia L, Kai L and Li F-F 2009 ImageNet: a large-scale hierarchical image database *Proc. IEEE Conf. on Computer Vision and Pattern Recognition* pp 248–55
- [37] Haralick R M, Shanmugam K and Dinstein I H 1973 Textural features for image classification *IEEE Trans. Syst. Man Cybern.* **SMC-3** 610–21
- [38] Zeiler M D 2012 Adadelta: an adaptive learning rate method (arXiv:[1212.5701](https://arxiv.org/abs/1212.5701))
- [39] Bianco S, Cadene R, Celona L and Napoletano P 2018 Benchmark analysis of representative deep neural network architectures *IEEE Access* **16** 64270–7