Online Defect Detection in LGA Crystallization Imaging Using MANet-Based Deep Learning Image Analysis

Yan Huo, Diyuan Guan * and Lingyan Dong

College of Information Engineering, Shenyang University, Shenyang 110044, China
* Correspondence: guandiyuan@syu.edu.cn

Abstract: In this paper, a MANet-based image detection approach is designed to inspect crystal defects during the cooling crystallization process, like that involving β-form L-glutamic acid (LGA), utilizing an online imaging device. The steps in the presented strategy encompass crystal image preprocessing, crystal image segmentation, and crystal classification. Firstly, the guided image filter is introduced to preprocess the collected crystallization images for offline training and online detection. Then, by using an image augmentation strategy to enlarge the number of crystal image samples for training, the MANet-based network is improved for crystal image segmentation. Accordingly, by defining some features, needle-like crystals can be categorized into four types with an efficient classifier for the detection of normal and defective crystals. The experimental results for the batch crystallization of β-form LGA are provided to illustrate the validity of the presented detection methodology.

Keywords: crystallization; deep learning; defect detection; LGA; image analysis

1. Introduction

The detection of crystal growth is necessary and important to control the crystal status during crystallization [1,2]. A lot of research efforts in the last two decades have been devoted to the detection of crystal sizes and morphology during batch crystallization processes, by using process analysis technologies (PATs). Specifically, the specialized instrument known as the focused beam reflectance measurement (FBRM) device is designed for the assessment of the chord length distribution (CLD) of crystals. Nevertheless, the margin of error in such measurements can prove to be unacceptable when dealing with particles that deviate significantly from a spherical shape.

Recently, in the past years, microscope cameras are increasingly being used for online monitoring of crystal growth, based on invasive and non-invasive imaging instruments [3]. Zhou et al. [4] proposed image analysis techniques employed to obtain valuable characteristics from micrographs, with particle vision and measurement (PVM) technology as an invasive imaging device. By comparison, an non-invasive imaging device placed outside a crystallizer was successfully used to monitor the morphology of the crystals [5,6]. Larsen et al. [7] effectively analyzed high-concentration crystal images and meticulously processed needle crystals, conducting an examination of various crystal characteristics to attain a comprehensive depiction of the crystals. Concerning the online measurement of crystallization by an imaging tool, several image detection approaches were investigated for crystallization, some for measuring the one-dimensional characteristic size, and others for two-dimensional (2D) or three-dimensional (3D) sizes [8,9]. For instance, Wu et al. [10] developed an efficient image processing method using neural networks to evaluate the particle size and morphology in different concentrations for batch crystallization processes. Gao et al. [11] employed deep learning technology to detect L-glutamic acid (LGA) crystals and accurately measure the two-dimensional sizes of the particles. Heisel et al. [12] used a novel two-step classifier via artificial neural networks for the task of discriminating between the crystalline shapes, including α- and β-form crystals. Zhang et al. [3] proposed an in
situ 3D measurement methodology to successfully assess 3D sizes of β-form LGA crystals with a binocular micro-vision system. Another 3D measurement technique [13], with the matching of the two sets of descriptor information, was further presented to assess the facet growth using the same binocular system. It is worth noting that this technique remains available for measuring or assessing crystal growth during crystallization. The size and shape of crystals traditionally may not fully reflect the quality of the crystals. It is interesting that the detection of crystal defects is carried out by using image analysis technologies.

Recently, deep learning strategies have been effectively employed in computer vision and various other domains [14]. The advent of the first deep network—AlexNet [15], in 2012, marked a significant milestone. This network used several advanced convolutional neural network (CNN) techniques, including parallel training, the ReLU function, and dropout technology. By stacking multiple convolutional layers and pooling layers, deeper network structures can be obtained, such as VGGNet [16], and ResNet [17], etc. In particular, based on these seminal neural network models, numerous experts and scholars have proposed improved network structures, thereby driving the rapid advancement of image segmentation. The CNN algorithm generates abstract deep features by merging shallow features and, subsequently, extracts distributed features for image semantic segmentation through pixel classification, resulting in an outstanding level of performance. Long et al. [18] designed the full convolutional network (FCN), using the convolution layer in place of the full connection layer, to solve the challenge of semantic image segmentation, realizing the end-to-end pixel classification of images. Badrinarayanan et al. [19] suggested a new semantic segmentation algorithm (SegNet) on the basis of the FCN. Groundtruth was used as the supervision information training network, and the image level classification was transformed into the pixel level classification, so as to realize semantic segmentation. Google proposed an extended semantic segmentation algorithm [20], which can expand the sensitivity field without increasing the number of parameters, and the representative algorithms include DeepLab-V1, DeepLab-V2, and DeepLab-V3 [21], etc. Ronneberger et al. [22] built a network of U-shaped architecture (U-Net) to solve the semantic segmentation issue, established based on the model architecture of the FCN.

Generally, the online detection of crystal quality in a suspension by utilizing an optical imaging system presents many difficulties. The major problem depends on the fact that there is no unified criterion for defect detection in needle-like crystals, subject to crystal agglomeration, crushing, and internal defects during crystallization. Nevertheless, there has been limited exploration of methods to address these challenges during batch crystallization processes in recent literature, particularly since the emergence of online microscopic image analysis technologies. To tackle the aforementioned issues, we develop a deep learning image analysis method for monitoring crystal defects during the cooling crystallization process, specifically focusing on β-form L-glutamic acid (LGA). Firstly, a guided image filter is introduced to preprocess images captured online of crystallization for offline training and online analysis, such that the impact of unbalanced illumination, solution turbulence in the stirred crystallizer, and particle movement could be effectively alleviated. Secondly, the multi-scale attention network (MANet) model [23] is adopted for deep learning analysis of crystal image segmentation, which does not need a large number of online crystal image samples. To overcome the problem of only having a very limited number of crystal images captured online to be used for training the deep learning network, an image augmentation strategy is proposed to improve the segmentation effect on the crystal images. Accordingly, four classes of β-form LGA crystals are recognized to complete the detection of crystal quality.

For clarity, the paper content is structured as follows. The experimental set-up is presented in Section 2. In Section 3, the suggested image detection process is recommended. Section 4 presents the experiment results to assess the performance of the method. Finally, some conclusions are presented in Section 5.
2. Experimental Set-Up

L-glutamic acid (LGA), C₅H₉NO₄, was chosen as the subject of this research. LGA crystals exist in two forms: the prismatic α-form and the needle-like β-form. In this study, β-form LGA crystals were crystallized to facilitate the detection of defective crystals. The non-invasive microscopic device (Pharma Vision, Qingdao, China) used to collect the crystallization images of β-form LGA is depicted in Figure 1. The experiments were conducted using a 1 L glass vessel, a four-paddle stirrer, and a temperature probe (PT100, JULABO, Seelbach, Germany). Temperature control was achieved using a thermostatic circulator (Julabo CF41, JULABO, Seelbach, Germany). The imaging system, equipped with two high-speed cameras (Pharma Vision, Qingdao, China), captured crystallization images during the crystallization processes. The camera (UI-2280SE-C-HQ, IDS, Obersulm, Germany) with a CCD sensor (IDS, Obersulm, Germany) and universal serial bus was manufactured by IDS Imaging Development Systems GmbH. The online images of β-form LGA crystals were extracted from the solution in the agitated reactor.

![Schematic of experimental set-up](image)

**Figure 1.** The schematic of the experimental set-up.

In this study, β-form LGA seeds were employed during the crystallization to suppress secondary nucleation and prevent the formation of little particles. It is worth noting that seed crystals were generated with β-form LGA crystals through milling, sieving, dissolving, and drying, etc. Initially, a 0.6 L LGA solution was poured into the vessel, and the stirring rate was set at 210 rpm. The solution was heated to 75 °C to dissolve all the LGA solutes. Subsequently, the temperature was gradually reduced to 45 °C at a rate of 0.5 °C/min. When the solution reached 45 °C, the crystal seeds were introduced into the crystallizer. Simultaneously, the online images were captured via the imaging system during the crystallization process, as shown in Figure 1. The solution was then cooled down to a temperature of 35 °C, at a cooling rate of 0.1 °C/min. During the experiment process, with one camera taking about six images per second, the non-invasive imaging device was able to collect 12 images per second.

3. Detection Methods

3.1. Overview of the Method

An overview of the presented detection strategy is described in Figure 2. As shown in Figure 2, the method involves two stages (i.e., offline training and online detection) employed to capture crystallization images in complicated situations (e.g., unbalanced illumination, solution movement, etc.). The first stage involves training the network parameters using the annotated images, followed by image augmentation. It is worth noting that the annotated image set includes the original images and the mask ones via manual labeling of the crystals. In the second stage, the online crystallization images, which are captured in real-time during the crystallization process, are segmented using the trained model after image preprocessing. The classification of the crystals can be completed for the segmented images.
3.2. Image Preprocessing

During online image collection and transmission, the collected images are often affected by noise pollution. The background of crystallization images may suffer from unevenness and low variation, while the reflective properties of the crystal plane due to variations in lighting and solution movement within the agitated reactor can significantly compromise the accuracy of crystal image segmentation. Therefore, preprocessing of the images is necessary. To mitigate the impact of noise on the subsequent operation, this study utilizes the guided filtering approach [24]. The guided image serves as a processed input image, and guided filtering acts as an edge-preserving filter, ensuring the conservation of information near the edges. The pixel $m_i$ in the input image and the pixel $q_i$ in the output one are defined by the following equation:

$$m_i = a_k q_i + \beta_k, \quad \forall i \in \omega_k$$  \hspace{1cm} (1)

where $k$ is the index of the local square window $\omega_k$, which is taken as $51 \times 51$ in the input image, and $(a_k, \beta_k)$ are the constants in $\omega_k$, which can be obtained by the following equation:

$$(a_k, \beta_k) = \arg \min_{a_i, \beta_i} \sum_{i \in \omega_k} \left( (a_k m_i + \beta_k - m_i)^2 + \varepsilon \alpha_k^2 \right)$$  \hspace{1cm} (2)

where $\varepsilon$ is the regularization parameter. And the solution is as follows:

$$a_k = \frac{1}{|\omega|} \sum_{i \in \omega_k} m_i^2 - \mu_k \mu_k}{\sigma_k^2 + \delta}$$  \hspace{1cm} (3)

$$\beta_k = \mu_k - a_k \mu_k$$ \hspace{1cm} (4)

where $\mu_k$ and $\sigma_k^2$ are the mean and variance of the input image in $\omega_k$, and $|\omega|$ is the total number of pixels in $\omega_k$, and

$$\mu_k = \frac{1}{|\omega|} \sum_{i \in \omega_k} m_i$$  \hspace{1cm} (5)

The output is calculated as:

$$q_i = \frac{1}{|\omega|} \sum_{i \in \omega_k} a_k m_i + \beta_k$$  \hspace{1cm} (6)

Furthermore, to maintain the image values within a consistent range, it is essential to normalize the denoising crystallization images. This normalization process involves dividing the pixel values by 255 to scale the image pixels between 0 and 1. While it is

---

**Figure 2.** Flow chart of the proposed detection strategy.
possible to achieve segmentation without normalization in a full CNN, normalizing the images can enhance segmentation accuracy and improve the convergence speed.

3.3. Image Segmentation

Although the deep learning network based on skipping joins can capture rich semantic features, it does not delineate the spatial and channel interdependencies among pixels, which are crucial for the segmentation of crystal images. In addition, the pixel equivalent is always defined by using clear crystal imaging, so fuzzy images of crystals can affect the extraction of clear crystals. For instance, overlapping or adhesion results are always misjudged, as shown in Figure 3. With the purpose of resolving the aforementioned problems, MANet is introduced, which needs a considerably smaller training set of crystallization image samples that have similar structures. An overview of the segmentation MANet model is shown in Figure 4. The model includes two parts, i.e., the encoder and decoder. In order to better obtain a large amount of useful multi-scale information in feature maps, MANet incorporates two attention layers known as the position-wise attention block (PAB) and the multi-scale fusion attention block (MFAB), which enhance the overall performance of the network. The PAB layer is designed to grab spatial dependencies among the pixels in a feature map using the attention mechanism. The MFAB layer is designed to grab channel dependencies across all feature maps using the attention mechanism. The MFAB layer not only obtains the channel dependencies in deep feature maps, but also obtains those in shallow ones.

Figure 3. The adhesion and overlapping image results: (a) adhesion; (b) overlapping.

Figure 4. The structure of the MANet-based model for crystal segmentation.

The feature maps are extracted by using the ResNet-50 model, pre-trained using ImageNet. The lowest feature is directly up-sampled by the CAPAFE++ method [25], and then the feature maps with different scales generated by the backbone network are input into the corresponding attention module for refinement. These refined feature maps are applied to the lower ones in the upper sample. The encoder has four layers of Resblocks, and each Resblock contains three $3 \times 3$ Conv blocks, a $1 \times 1$ Conv block, and one residual connection with group normalization. For controlling model overfitting, a
dropout operation [26] is applied in the convolution layer. The dropout ratio is 0.5. The channel of the PAB is set as 64. The fused features are then up-sampled in the decoder, by the corresponding up-sampling layer, using the CAPAFE++ method. This model has four skip connections. Finally, through the deconvolution operation, the output of the last layer is up-sampled to match the spatial resolution of the input, and input into the last convolution layer to obtain the predicted result.

To expedite the learning process for crystal images, the exponential linear unit (ELU) [27] is employed as an activation function \( f(t) \) with \( \lambda > 0 \) as:

\[
f(t) = \begin{cases} 
  t, & t > 0 \\
  \lambda(e^t - 1), & t \leq 0 
\end{cases}
\]  

To overcome pixel disproportion, the combined loss function is determined as:

\[ J_{\text{loss}} = C_{\text{loss}} + D_{\text{loss}} \]  

where \( C_{\text{loss}} \) is the cross-entropy loss, and \( D_{\text{loss}} \) is the dice coefficient loss.

In scenarios where only a limited number of training samples are available, image augmentation serves as a crucial strategy to augment the training dataset and enhance the generalization capability of the segmentation model. Specifically in the context of microscopic images, it is advisable to apply random rotation, scaling, reflection, and brightness jitter to the sparse set of annotated images to facilitate the training of a segmentation network. The random rotation degree spans the interval \((0, 360^\circ)\), while the random scale factor varies within the range of \((0.5, 1.5)\). Additionally, the brightness value varies within the range of \((-0.3, 0.1)\). As a result of the image augmentation process, four training samples are generated for each original image, as shown in Figure 5.

![Image augmentation results](image.png)

**Figure 5.** Image augmentation results: (a) original image; (b) rotation; (c) scaling; (d) reflection; (e) brightness jitter.

### 3.4. Image Classification

Generally, crystal adhesion, crushing, and dissolution, lead to irregular crystal defects during the cooling crystallization process. Therefore, it is possible that not all crystals in one batch are in the uniform \( \beta \)-shape. It is necessary that \( \beta \)-form LGA crystals are classified in order to inspect normal and defective particles. The semantic segmentation method, based on deep learning, may not address the classification issue for crystallization images on account of there being less unnormal samples and similar particle characteristics. In this study, \( \beta \)-form LGA crystals can be classified into four classes (i.e., normal class, incomplete class, adhesion class, and internal defect class), as shown in Table 1. Normal crystals with few defects are defined as Class 1. Incomplete crystals may be caused by particle collision and breakage, which are classified as Class 2. Adhesion crystals, which includes two or more crystals, are categorized as Class 3. Class 4 just encompasses defective crystals, with internal defects and irregular textures. As displayed in Table 1, Class 2 and Class 3 crystals can be distinguished from the other two classes by the convexity degree (CD), and the CD...
of Class 2 is much larger than that of Class 3. However, the CD of Class 1 and Class 4 are close to 1. The texture features between Class 1 and Class 4 are a bit different. Then, the number of defective crystals in one batch can be counted. Therefore, the CD, contrast, and entropy are selected as the features for classification. In addition, Fisher discriminant analysis [28], as a simple and efficient classifier, is used. Suppose the class of the training samples \( x_i \) \((i = 1, 2, \ldots, n)\) is \( k \) \((k = 4)\), and the \( j \)-th \((1 \leq j \leq 4)\) class \( C_j \) contains \( n_j \) samples. The inter-class divergence matrix is \( S_b \), and the intra-class divergence matrix is \( S_w \).

\[
S_b = \sum_{j=1}^{k} n_j (u_j - u)(u_j - u)^T
\]

\[
S_w = \sum_{j=1}^{k} \sum_{x_i \in C_j} (x_i - u_j)(x_i - u_j)^T
\]

where \( u_j = \frac{1}{n_j} \sum_{x_i \in C_j} x_i \) is the mean of the \( j \)-th class, \( u = \frac{1}{n} \sum_{i=1}^{n} x_i \) is the mean of the overall samples. The objective function with \( v \) is:

\[
T(v) = \arg \max_{v^T S_b v} \frac{v^T S_b v}{v^T S_w v}
\]

Then, the rate of the four types is easily obtained to further estimate the quality information combined with the CSD for a batch of crystals.

<table>
<thead>
<tr>
<th>Crystal Class</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1: Normal crystals</td>
<td><img src="image1" alt="Image" /></td>
</tr>
<tr>
<td>Class 2: Incomplete crystals</td>
<td><img src="image2" alt="Image" /></td>
</tr>
<tr>
<td>Class 3: Adhesion crystals</td>
<td><img src="image3" alt="Image" /></td>
</tr>
<tr>
<td>Class 4: Internal defective crystals</td>
<td><img src="image4" alt="Image" /></td>
</tr>
</tbody>
</table>
4. Experimental Results

The experiments were performed during the crystallization process of β-form LGA. The images used in the experiment were collected during the cooling crystallization process. The MANet-based model was improved and conducted via the open-source package [29]. In this study, about 80 images were randomly selected for the training of the segmentation model. The training images were annotated and augmented at first. The model was trained with a maximum epoch number of 180. The batch size for training was set as four. The Adam optimizer was employed during training. The initial learning rate was 0.005.

For segmentation testing, the improved segmentation method based on MANet for crystal extraction can take less than one second to process per image, and the reliable segmentation of clear crystals was realized. A general image collected during this study is shown in Figure 6a. Figure 6b shows the result of image preprocessing. It is observed that the noise can be removed, and the preprocessed image is of higher quality than the captured one. In Figure 6c, the crystal extraction result is obtained by using the improved segmentation method based on MANet applied to that of Figure 6b. We can see that the four particles are obvious in Figure 6c.

In addition, to indicate the superiority of the proposed image processing method for crystallization image analysis, the improved model, the thresholding segmentation method, the multi-scale segmentation method, and the original MANet model were used for comparison. The four segmentation methods are applied to the preprocessed crystal image shown in Figure 7a. Figure 7b shows the result with the improved MANet model. Figure 7c,d shows the result for the thresholding and multi-scale Canny segmentation methods, respectively. It is seen that the clear crystals are effectively extracted by the recommended technique for crystal classification, which can overcome the influence of crystal overlapping, while the two traditional techniques exhibit various shortcomings, such as the occurrence of fuzziness and the overlapping of crystals delineated by a red circle, as exhibited in Figure 7c,d. In addition, Figure 7e illustrates that the outcome generated by the original MANet model is less accurate when compared to that of the improved network. It is indicated that the improved MANet model can generally obtain clear particles to facilitate crystal classification for defect detection, as shown in Figure 7.

In this experiment, the LGA crystals can be classified into four classes (i.e., normal class, incomplete class, adhesion class, and internal defect class), as shown in Table 1. In order to verify the performance of the classification model, we randomly selected 250 particles with the feature dataset as the training set, and randomly selected 100 particles (not containing the training set) as the testing set. For the classification of LGA crystals, the three feature variables (i.e., convexity degree, contrast, and entropy) were extracted from the identified crystals as the input features of the classification model. The classification model was trained using the features of the training particles. Table 2 demonstrates that the classification model, with Fisher discriminant analysis, has high classification accuracy. The error rate was about 6% for crystal class determination. Particularly, the accuracy for Class 1
reaches 98%, as shown in Table 2. It is also seen from Table 2 that the defective particles in Class 3 can be identified with high accuracy. However, because the internal defects are not distinct, the accuracy for Class 4 is just 86%. Additionally, some particles in Class 2 may be misjudged as Class 3. During application, in order to improve the recognition rate, the features set requires full training, as much as possible. Accordingly, the rates of the normal (Class 1) and defective particles (Class 2–4) can be calculated to assess the crystallization quality of the batch.

Figure 7. The segmentation comparison results illustrated by LGA images: (a) the captured image; (b) the improved MANet result; (c) the thresholding segmentation result; (d) the multi-scale Canny detection result; (e) the MANet result.

Table 2. Classification accuracy.

<table>
<thead>
<tr>
<th>Class</th>
<th>Actual Number</th>
<th>Accurately Classified Number</th>
<th>Classification Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1</td>
<td>54</td>
<td>53</td>
<td>0.98</td>
</tr>
<tr>
<td>Class 2</td>
<td>13</td>
<td>11</td>
<td>0.85</td>
</tr>
<tr>
<td>Class 3</td>
<td>19</td>
<td>18</td>
<td>0.95</td>
</tr>
<tr>
<td>Class 4</td>
<td>14</td>
<td>12</td>
<td>0.86</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>94</td>
<td>0.94</td>
</tr>
</tbody>
</table>

5. Conclusions

In this work, an online detection method has been developed by utilizing a non-invasive imaging device, enabling us to assess crystal defects in a stirred reactor. The raw crystal images were captured during the crystallization process. Guided filtering is employed to enhance the image quality by effectively reducing image noise. The segmentation method based on MANet is actually utilized to obtain the particles from the images characterized by an uneven background, caused by the illumination light or the movement of particles and solution. Four β-form LGA crystals are categorized for the further evaluation of defective crystals. In conclusion, the experiment results on the crystallization process highlighted the effectiveness of the online defect detection method, including segmentation and classification, in facilitating crystallization control. In fact, crystal classification through 2D image analysis is still a bit inaccurate, especially in regard to internal defects. So, future work will focus on the detection of crystal defects by reconstructing the 3D shape with a binocular vision system.

Author Contributions: Conceptualization, L.D.; methodology, D.G.; software, Y.H.; investigation, L.D. and Y.H.; writing—original draft preparation, Y.H.; writing—review and editing, D.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported in part by the funding project of Northeast Geological S&T Innovation Center of the China Geological Survey, grant number, QCJJ2022-24; in part by the China Postdoctoral Science Foundation, grant number, 2020M680979; and in part by the Basic
Scientific Research Project of the Higher Education Institutions of Liaoning Province, grant number, JYTM20231165.

**Data Availability Statement:** The original contributions presented in the study are included in the article material, further inquiries can be directed to the corresponding author.

**Acknowledgments:** We acknowledge Ji Fan for his help during the crystallization experiments.

**Conflicts of Interest:** The authors declare no conflicts of interest.

**References**


**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.