

Article

Protein Crystal Instance Segmentation Based on Mask R-CNN

Jiangping Qin ^{1,2} , Yan Zhang ², Huan Zhou ³, Feng Yu ³, Bo Sun ^{3,*} and Qisheng Wang ^{1,3,*}¹ Shanghai Institute of Applied Physics, Chinese Academy of Sciences, Shanghai 201210, China; qinjiangping@stu.shmtu.edu.cn² Logistics Engineering College, Shanghai Maritime University, Shanghai 201306, China; zhangyan@shmtu.edu.cn³ Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai 201210, China; zhouhuan@sari.ac.cn (H.Z.); yufeng@sari.ac.cn (F.Y.)

* Correspondence: sunb@sari.ac.cn (B.S.); wangqs@sari.ac.cn (Q.W.)

Abstract: Protein crystallization is the bottleneck in macromolecular crystallography, and crystal recognition is a very important step in the experiment. To improve the recognition accuracy by image classification algorithms further, the Mask R-CNN model is introduced for the detection of protein crystals in this paper. Because the protein crystal image is greatly affected by backlight and precipitate, the contrast limit adaptive histogram equalization (CLAHE) is applied with Mask R-CNN. Meanwhile, the Transfer Learning method is used to optimize the parameters in Mask R-CNN. Through the comparison experiments between this combined algorithm and the original algorithm, it shows that the improved algorithm can effectively improve the accuracy of segmentation.

Keywords: protein crystal; Mask R-CNN; instance segmentation; transfer learning



Citation: Qin, J.; Zhang, Y.; Zhou, H.; Yu, F.; Sun, B.; Wang, Q. Protein Crystal Instance Segmentation Based on Mask R-CNN. *Crystals* **2021**, *11*, 157. <https://doi.org/10.3390/cryst11020157>

Academic Editor: Rocco Caliandro

Received: 8 January 2021

Accepted: 1 February 2021

Published: 4 February 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Protein crystallography is an important subject for studying structure biology. The three-dimensional structural characterization of biological macromolecules is very important in order to understand their mechanism of action. The crystallography method is widely used in the drug discovery also, especially in the fragment-based drug screening [1–3].

At present, 169,436 protein structures have been deposited in the Protein Data Bank (PDB), and more than 88% of them are resolved by the X-ray crystal diffraction method (<http://www1.rcsb.org> (accessed on 8 January 2021)). To crystallize the protein, in most situations, it is still necessary for researchers to observe the samples through a microscope and to determine whether the crystallization process is completed. The observation would cost time and labor. Therefore, to design an automated system for protein crystallography, from protein purification to crystal growth, becomes an urgent requirement in the field of life sciences [4–7].

Some work has been done to classify the images of protein crystals [8,9]. Because they relied on a special imaging system, which was relatively rare at that time, it could not achieve good classification results. The image of protein crystal is seriously affected by the performance of backlight and the focusing of the microscope. Therefore, the quality of the classification algorithm based on a traditional machine learning method was totally dependent on the design of a feature vector, and the classification task could not be well realized. Bruno et al. [10] proposed a classification algorithm based on the deep convolutional neural network to classify protein crystallization results in 2018, which can achieve about 94% of the classification effect. However, it is only able to show whether there are crystals in the droplet. If we wanted to know where crystals are in the droplets from image, the classification task would not be able to solve this problem. Meanwhile, some commercial devices have been developed to satisfy the requirement from academic centers and pharmaceutical companies. The fully automatic crystallization imaging system,

Rockimager1500, was designed by the Formulatrix [11] company. However, it was only able to analyze the experimental drop and not crystals in the drop.

Other optical methods also have been adopted for more trials. To reduce the influence of optical scattering, the second-order nonlinear optical imaging method was used to identify protein crystals [7]. The second harmonic signal that was generated by light and materials was used to search crystals. Meanwhile, it was able to help observe smaller crystals because a second harmonic generation (SHG) signal can frequently be observed from structures that are approximately the same size as or even smaller than the lateral resolution of the microscope. However, the method is more suitable for drug crystals with chiral. Therefore, the scope of application is limited. Compared with the above mentioned studies, none of them could well meet the requirements of researchers. Image segmentation can determine pixels which belong to the object or background in the image. Therefore, this paper attempts to use the instance segmentation algorithm to better identify protein crystals in the drop.

Mask R-CNN was proposed by the Facebook Researcher KM He [12], which integrates target detection and instance segmentation. In this paper, based on the Mask R-CNN method, the collected protein crystal image is marked as a suitable format that is in accordance with the format of the Microsoft Common Objects in Context (MS COCO) dataset for network training. The self-built protein crystal dataset was trained using the pre-trained network weight in the way of data transfer, and the Mask R-CNN network was fine-tuned.

2. Algorithm Design

2.1. Network Introduction

As protein crystal images are highly affected by the light and lens focus during collection process, a pre-processing module was added before the frame of Mask R-CNN to process the input images, which could better highlight the features of protein crystals in the image. The improved Mask R-CNN structure is shown in Figure 1.

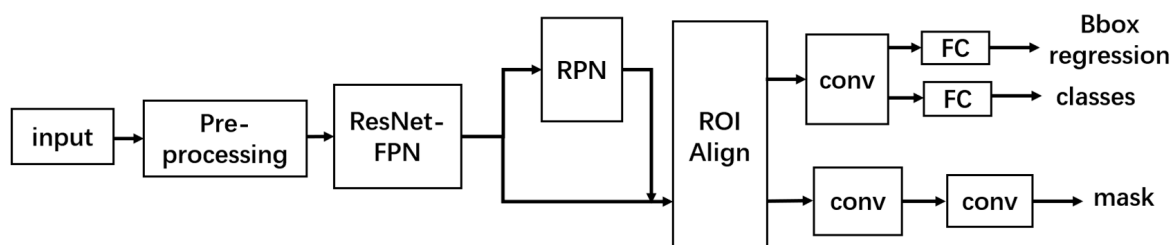


Figure 1. The network structure of improved Mask R-CNN.

The output of Mask R-CNN is divided into three parts: the prediction box regression, the image classification, and the mask branch. Among them, the prediction box regression and the image classification belong to the target detection part, while the mask branch belongs to the instance segmentation part.

In Mask R-CNN structure, the protein crystal image is input into the network, and then different feature maps are output by means of a series of convolution and pooling in feature pyramid networks (FPN). After that, different feature maps are delivered into the region proposal networks (RPN) so as to extract the region of interest (ROI). Then the ROI is input to the ROI Align to perform pixel correction on the feature map for subsequent target classification and bounding box regression. In the mask branch, the original images are cropped using the corrected bounding box, and then the images in ROI are performed by mask prediction. Therefore, the object in the bounding box belongs to the two-class classification problem (0: background, 1: object). This can avoid inter-class competition and

the final result belongs to instance segmentation. The total loss function of Mask R-CNN is defined as

$$L_{total} = L_{cls} + L_{box} + L_{mask}, \quad (1)$$

where L_{cls} is classification loss; L_{box} is regression loss of bounding box; L_{mask} is semantic segmentation loss.

2.2. Pre-Processing Module

It is necessary to use image enhancement technology to enhance the contrast of protein crystal images and to highlight the features of protein crystals in the image. Histogram equalization (HE) is a common contrast enhancement method in gray space. Firstly, the frequency of each pixel level is counted by means of histogram equalization. Then, cumulative distribution function (CDF) is used in the acquired frequency as a result that the pixels of original image are mapped to new pixels through CDF, and this transformation process is a nonlinear transformation. Because CDF is a monotonic increasing function, after transformation, the brighter areas in the original image are still brighter. Histogram equalization is a nonlinear mapping method which is performed over the entire gray image instead of focusing on the local features of the image. Although the pixels are stretched in the more concentrated area of the gray image to make the dynamic range expand, noise information may be enhanced by HE in images, and for those images that contain obviously brighter or darker areas, it often fails to achieve a significant enhancement effect.

In order to better deal with local features, this paper uses the contrast limited adaptive histogram equalization (CLAHE) algorithm to preprocess images [13]. The input images are divided into $m \times n$ areas in CLAHE, and these areas are processed separately. Firstly, the gray histogram of each area is calculated. Next, the parts of the histogram above the threshold are cropped and then are accumulated. Then, the accumulated result is averagely distributed to each pixel level. The slope of the cumulative distribution function could be effectively limited by a cropping operation. Enhancement of neighborhood noise around the pixel is mainly caused by the slope of transformation function, and the noise that is around the pixel is proportional to the cumulative distribution function of the neighborhood. Therefore, once the slope of the cumulative distribution function is limited, the noise could be effectively limited. Then, the limited gray histogram is equalized to obtain the pixel mapping relationship. The pixel information of the edge between regions is discontinuous. As a result, a block effect occurs. Therefore, bilinear interpolation is used to fix the block effect in images. Meanwhile, the bilinear interpolation can also improve the computational efficiency. The effect of CLAHE is shown in Figure 2.

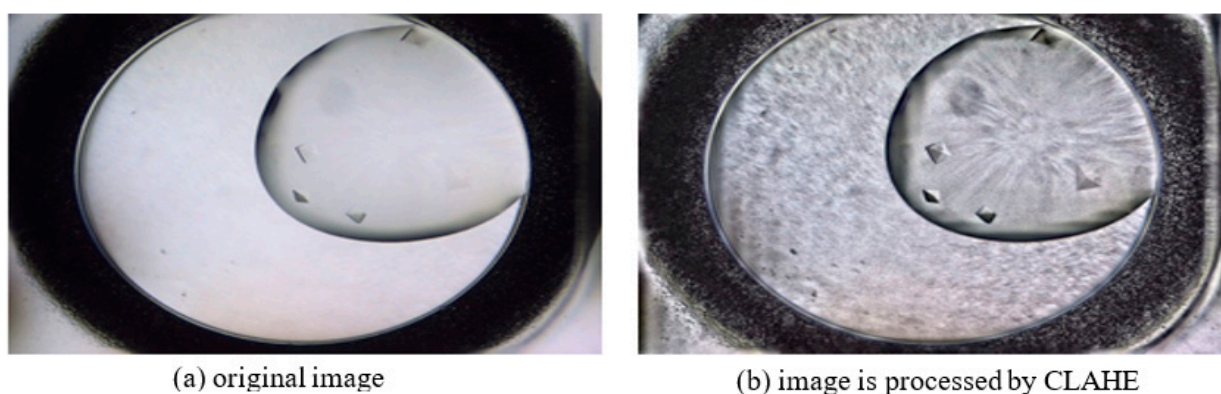


Figure 2. Comparison before and after processed by contrast limit adaptive histogram equalization (CLAHE). (b) is the image processed by CLAHE, and it can be clearly seen that the features of the crystal edge in (b) are more obvious than in (a).

2.3. FPN Module

FPN is a multi-scale feature fusion network structure which was proposed by the team of KM He in 2017 [14]. FPN is different from the traditional image pyramid structure. It is divided into three parts: bottom-up, top-down, and horizontal-connection. The structure is shown in Figure 3.

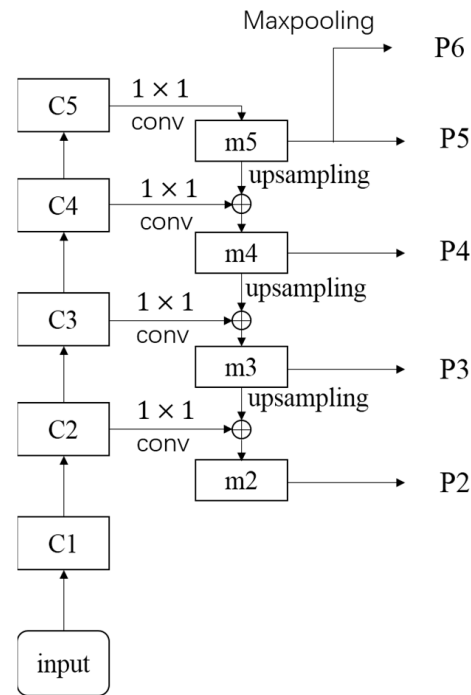


Figure 3. Feature pyramid networks (FPN) structure.

ResNet101 was adopted as a feature extraction network for obtaining different feature maps $[C_1, C_2, C_3, C_4, C_5]$ in the bottom-up structure. The scale relative to the original image is $[\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{32}]$. Up-sampling operation is continuously performed from the C_5 to C_3 layers in the top-down structure. The up-sampling is performed by means of nearest neighbor up-sampling in the top-down structure, and the purpose of up-sampling is to double the scale of the upper-layer feature map. Through the horizontal connection, the up-sampled high-level features can be fused with the low-level features, which can better integrate the semantic information and location information, and can also use different scale features more effectively. After fusion, the 3×3 convolution kernel will be used to process the fused features in order to eliminate the aliasing effect of up-sampling.

Finally, the $[P_2, P_3, P_4, P_5, P_6]$ which are generated in FPN are sent to the RPN for generating the region proposal, and the region proposal are performed for target detection in RPN. There are three priori boxes with aspect ratios $[2 : 1, 1 : 1, 1 : 2]$ generated on each pixel from different feature maps in RPN. The scale of the priori box increases as the scale of the feature map decreases. At the same time, $[P_2, P_3, P_4, P_5]$ are sent to Fast RCNN and are combined with the region proposal that is output by RPN to perform regression and classification on the detection frame and the recognized object. The RPN loss function is

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*). \quad (2)$$

2.4. ROI Align Module

The ROI Align method was proposed in Mask R-CNN. Because it does not perform the quantization and rounding of coordinates of the ROI area, the problem of mis-alignment between the feature map and the original image in ROI pooling was solved by ROI Align.

The structure of ROI Align is shown in Figure 4. The region with dotted line represents the generated feature maps, and the rectangle region surrounded by a solid line represents the ROI that has been adjusted. The ROI is divided into 5×5 cells. If the number of samples in each cell is 4, each cell will be averaged divided into four bins, and the center of each bin is the sampling point. Since the coordinates of the ROI are floating-point numbers, the coordinates of the sampling points are usually also floating-point numbers. Therefore, bilinear interpolation is adopted for each sampling point pixel, as shown by the arrow in Figure 3. This operation can be used to obtain the pixel value of the sampling point, and then four sampling points are performed max pooling on each cell. Finally, the ROI Align output are obtained.

2	32	14	2	3	5	6	7	2	3
4	5	6	25	14	7	8	9	10	12
13	14	15	16	48	12	85	22	12	11
23	25	45	76	88	96	12	25	22	29
45	44	75	73	72	71	56	92	12	35
55	22	44	11	12	32	95	14	18	23
25	23	45	75	78	46	47	25	56	66
77	88	92	47	15	63	66	1	21	4
7	5	8	9	5	45	1	12	12	25
12	55	12	23	3	5	7	5	9	14

Figure 4. ROI Align structure.

3. Results and Analysis

3.1. Experiment Platform

The software environment for the experiment platform is based on Windows 10. The experiment framework is keras 2.2.4 and tensorflow 1.13. The CPU is AMD R5 3600. The memory is 16G, and the graphics processing unit (GPU) is NVIDIA RTX2060. In order to effectively utilize the GPU resources, the scale of the original image is adjusted to 512×512 before training. The area with no image is filled with black edges, and then, adjusted images are input into the network for training.

3.2. Experiment Dataset

The COCO dataset is a dataset which is provided by the Microsoft company. It can be used for image segmentation or target detection. In this paper, the weight model is obtained by pre-training on this dataset, and the crystal images are downloaded from Machine Recognition of Crystallization Outcomes (MARCO) (<https://marco.ccr.buffalo.edu/> (accessed on 8 January 2021)). Because the MARCO dataset is only used for classification, the ground truth of downloaded crystal images is annotated by colleagues with the background of protein crystallography. The Labelme software (<https://github.com/wkentaro/labelme/tree/v3.11.2> (accessed on 8 January 2021), version is 3.16.2) is used to annotate the crystals as masks in the image, and according to the format of the COCO dataset, these annotated images are designed as a crystal dataset. Corresponding json files, yaml files, and mask files are generated in the crystal dataset. The labeled mask image is shown in Figure 5.

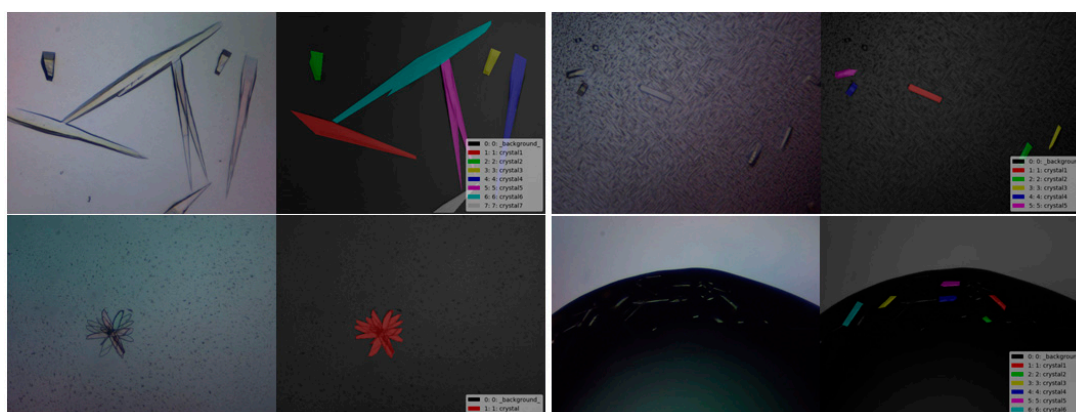


Figure 5. Four labeled mask images.

3.3. Experiment Results and Analysis

There are two important evaluation indicators for the performance of the classification problem. One is precision, which is used to evaluate how many objects are correctly identified in the result of classification. The other is Recall, which is used to evaluate how many positive examples are predicted correctly in the total positive samples. The calculation formulas for Precision and Recall are (3) and (4), respectively.

$$P = \frac{TP}{TP + FP}, \quad (3)$$

$$R = \frac{TP}{TP + FN}, \quad (4)$$

where TP means that the positive class is predicted to be positive; FP means that the negative class is predicted to be positive; FN means that the positive class is predicted to be negative.

For the target detection network, there is a very important concept, intersection over union (IOU). The degree of overlap of two regions is expressed by IOU. When it is adopted to test the accuracy of the network prediction, IOU expresses the overlap between the prediction box and the labeled box. The calculation formula is as follows:

$$IOU = \frac{A \cap B}{A \cup B}. \quad (5)$$

Firstly, the result of experiment is evaluated by mAP (IOU = 0.50) in this paper, and 10 images are selected randomly from the validation set to calculate mAP values. Secondly, 100 images are randomly selected from the validation set to calculate mAP (IOU = 0.65). According to mAP values, precision of network prediction can be verified after adding the CLAHE algorithm. The results are shown in Table 1. We can see that the precision of network prediction is improved by means of adding a preprocessing module.

Table 1. The mAP values of network prediction before and after adding CLAHE.

	ResNet101	CLAHE-ResNet101	IOU
10 images	0.668	0.703	0.50
100 images	0.302	0.430	0.65

The instance segmentation results of the dataset are shown in Figure 6. Even with many precipitations, most protein crystals are identified by the network after adding the CLAHE. The results of instance segmentation are more conforming with the shape of protein crystals.

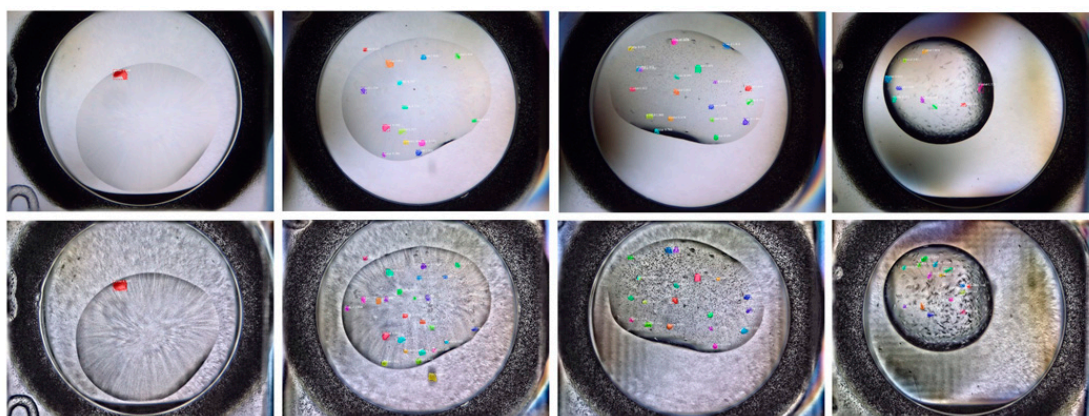


Figure 6. Result of protein crystal instance segmentation. The cases of instance segmentation without CLAHE are listed in the first row. The cases of instance segmentation with CLAHE are listed in the second row.

4. Conclusions

Mask R-CNN is introduced in this paper and the CLAHE algorithm is tried as an image pre-processing module. The two parts are combined to realize the instance segmentation task of protein crystal. From the perspective of mAP quantitative and qualitative analysis, the test accuracy of the network improved by 42%, from 30.2% to 43.0% after addition of CLAHE in IOU = 0.65. Even with just 10 images and an IOU of 0.5, the test accuracy of the network also improved by 5% (from 66.8% to 70.3%). It is proved that the performance of the network can be improved by image preprocessing, even when the model structure has not been changed too much.

However, according to the results of instance segmentation, the spots outside the droplet that are not crystals may be misidentified as crystals by the network. Therefore, the accuracy of network prediction will be improved in subsequent experiments from two aspects: increasing the segmentation accuracy or preprocessing the droplet. At the same time, the following work will also reduce the amount of computation and parameters in order to improve the network computing speed.

Author Contributions: Data curation, J.Q.; funding acquisition, Q.W.; investigation, J.Q.; methodology, J.Q., Y.Z., and B.S.; resources, Q.W.; supervision, Y.Z., H.Z., F.Y., and B.S.; validation, J.Q.; writing—original draft, J.Q.; writing—review & editing, B.S. and Q.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by a grant from the National Key Research and Development Program of China (Grant No. 2017YFA0504901).

Acknowledgments: We thank the staff of beamline BL17U1 and the experimental auxiliary system at Shanghai Synchrotron Radiation Facility (SSRF), beamlines BL18U1 at National Facility for Protein Sciences (NFPS) for images support and technology guidance.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Blundell, T.L. Protein crystallography and drug discovery: Recollections of knowledge exchange between academia and industry. *IUCrJ* **2017**, *4*, 308–321. [[CrossRef](#)] [[PubMed](#)]
2. Spiliopoulou, M.; Valmas, A.; Triandafillidis, D.-P.; Kosinas, C.; Fitch, A.N.; Karavassili, F.; Margiolaki, I. Applications of X-ray Powder Diffraction in Protein Crystallography and Drug Screening. *Crystals* **2020**, *10*, 54. [[CrossRef](#)]
3. Brink, A.; Helliwell, J. Why is interoperability between the two fields of chemical crystallography and protein crystallography so difficult? *IUCrJ* **2019**, *6*, 788–793. [[CrossRef](#)] [[PubMed](#)]
4. Theveneau, P.; Baker, P.; Barrett, R.; Beteva, A.; Bowler, M.W.; Carpentier, P.; Caserotto, H.; Sanctis, D.; Dobias, F.; Flot, D.; et al. The Upgrade Programme for the Structural Biology beamlines at the European Synchrotron Radiation Facility—High throughput sample evaluation and automation. In Proceedings of the 11th International Conference on Synchrotron Radiation Instrumentation, Lyon, France, 9–13 July 2012. [[CrossRef](#)]

5. Ng, J.T.; Dekker, C.; Reardon, P.; Von Delft, F. Lessons from ten years of crystallization experiments at the SGC. *Acta Crystallogr. Sect. D Struct. Biol.* **2016**, *72*, 224–235. [[CrossRef](#)] [[PubMed](#)]
6. Zheng, B.; Tice, J.D.; Roach, L.S.; Ismagilov, R.F. A Droplet-Based, Composite PDMS/Glass Capillary Microfluidic System for Evaluating Protein Crystallization Conditions by Microbatch and Vapor-Diffusion Methods with On-Chip X-ray Diffraction. *Angew. Chem. Int. Ed.* **2004**, *43*, 2508–2511. [[CrossRef](#)]
7. Kissick, D.J.; Wanapun, D.; Simpson, G.J. Second-Order Nonlinear Optical Imaging of Chiral Crystals. *Annu. Rev. Anal. Chem.* **2011**, *4*, 419–437. [[CrossRef](#)] [[PubMed](#)]
8. Spraggon, G.; Lesley, S.A.; Kreusch, A.; Priestle, J.P. Computational analysis of crystallization trials. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2002**, *58*, 1915–1923. [[CrossRef](#)] [[PubMed](#)]
9. Snell, E.H.; Luft, J.R.; Potter, S.A.; Lauricella, A.M.; Gulde, S.M.; Malkowski, M.G.; Koszelak-Rosenblum, M.; Said, M.I.; Smith, J.L.; Veatch, C.K.; et al. Establishing a training set through the visual analysis of crystallization trials. Part I: ~150 000 images. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2008**, *64*, 1123–1130. [[CrossRef](#)] [[PubMed](#)]
10. Bruno, A.E.; Charbonneau, P.; Newman, J.; Snell, E.H.; So, D.R.; Vanhoucke, V.; Watkins, C.J.; Williams, S.; Wilson, J.C. Classification of crystallization outcomes using deep convolutional neural networks. *PLoS ONE* **2018**, *13*, e0198883. [[CrossRef](#)] [[PubMed](#)]
11. Jones, H.G.; Wrapp, D.; Gilman, M.S.A.; Battles, M.B.; Wang, N.; Sacerdote, S.; Chuang, G.-Y.; Kwong, P.D.; McLellan, J.S. Iterative screen optimization maximizes the efficiency of macromolecular crystallization. *Acta Crystallogr. Sect. F Struct. Biol. Commun.* **2019**, *75*, 123–131. [[CrossRef](#)] [[PubMed](#)]
12. He, K.; Gkioxari, G.; Dollar, P.; Girshick, R.B. Mask R-CNN. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 386–397. [[CrossRef](#)] [[PubMed](#)]
13. Reza, A.M. Realization of the Contrast Limited Adaptive Histogram Equalization (CLAHE) for Real-Time Image Enhancement. *J. VLSI Signal Process. Syst. Signal Image Video Technol.* **2004**, *38*, 35–44. [[CrossRef](#)]
14. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2016, Las Vegas, NV, USA, 27–30 June 2016; pp. 937–944.