Interpretability Analysis and Attention Mechanism of Deep Learning-Based Microscopic Vision

Zheng Xu *, Xinwei Zhao, Xiaodong Wang *, Yuchen Kong, Tongqun Ren, Yanqi Wang

Posted Date: 11 April 2024

doi: 10.20944/preprints202404.0823.v1

Keywords: microscopic vision; micro-assembly; convolutional neural network; attention mechanism; gradient-weighted class activation mapping

 doi:10.20944/preprints202404.0823.v1

Article

Interpretability Analysis and Attention Mechanism of Deep Learning-based Microscopic Vision

Zheng Xu [1,*], Xinwei Zhao [1], Xiaodong Wang [2,*], Yuchen Kong [1], Tongqun Ren [1] and Yanqi Wang [2]

[1] State Key Laboratory of High-Performance Precision Manufacturing, Dalian University of Technology
[2] Key Laboratory for Micro/Nano Technology and System of Liaoning Province, Dalian University of Technology
* Correspondence: xuzheng@dlut.edu.cn (Z.X.); xdwang@dlut.edu.cn (X.W.)

Abstract: Microscopic vision plays an important role in automated micro-assembly. However, some uncertain factors in the assembly process, such as occlusion and stains can lead to the mistakes of feature extraction. Herein, to solve the problem, the deep learning techniques are introduced into the feature recognition tasks, focusing on the attention mechanism and visualizing CNNs for DL-based microscopic vision. The main contributions are summarized as follows: The CBAM attention mechanism is combined with the YOLOv5 algorithm to improve the accuracy and robustness of feature extraction. The micropart feature occlusion experiment results show that at 70% occlusion degree, YOLOV5-CBAM can reach 97.9% mAP@0.5, which is 4.6% higher than the original one. Visualization analysis of DL-based model is conducted using Grad-CAM to make the decision result more transparent and avoid potential visual detection risks during assembly. The heatmap matching degree between GT area and high-light area is increased by 27.81% on average, which further verify the effectiveness of attention mechanism in micropart feature localization. Additionally, micropart surface stain and droplet quality classification models based on ResNet50 are trained to replace the manual sorting. The visual results are consistent with human eye discernment and judgement, confirming the reliability of parts and droplets sorting.

Keywords: microscopic vision; micro-assembly; convolutional neural network; attention mechanism; gradient-weighted class activation mapping

1. Introduction

Automated micro-assembly is a crucial technology for cost-effective manufacturing of complex micro-products, in which microscopic vision is primarily utilized to guide mechanical units in picking up, placing, and connecting parts. Feature extraction methods in vision typically rely on feature points or contours *via* gradient calculation with a single fixed differential operator or a combination of operators. In fact, various uncertainties exist during the assembly process, including the surface contamination of microparts, the variations from pre-processing such as dispensing, and the partial occlusions between parts and grippers which affect the integrity and clarity of the feature points and contours, thereby leading to the mistakes of feature extraction, as shown in Figure 1. To prevent further harmful actions with the mistakes, the auxiliary constraints to mechanisms or algorithms are commonly indispensable. However, if the assembly process requires changes, numerous adjustments may be necessary with these approaches. Therefore, there is an urgent need to improve both the universality and robustness of vision techniques in the micro-assembly field. [1–4]

 doi:10.20944/preprints202404.0823.v1

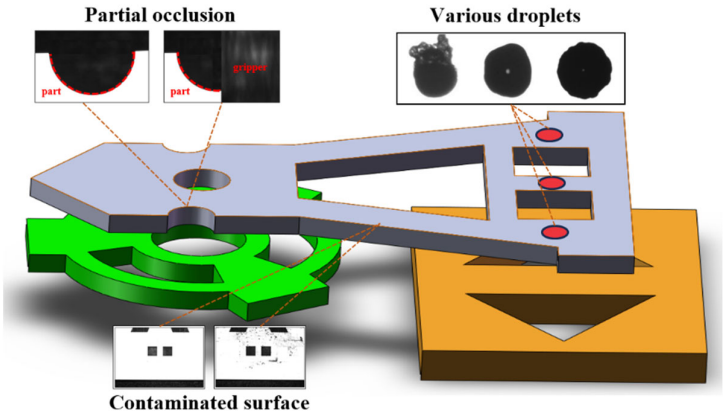

Figure 1. Some examples of various uncertainties in assembly of MEMS sensor.

In recent decades, the rapid development of artificial intelligence, particularly deep learning (DL) and reinforcement learning, has presented new opportunities for advancing micro-assembly. Deep networks, being the most powerful and practical machine learning models, are commonly encountered and significantly enhance the performance of various recognition tasks[5–8].

In terms of feature detection, DL-based algorithms can be broadly categorized into single-stage algorithms and two-stage algorithms. Two-stage algorithms, such as the R-CNN series, firstly generate candidate regions and then perform classification and regression on these regions. While two-stage algorithms exhibit high detection accuracy, they involve a complex two-stage computation process. In contrast, single-stage algorithms, such as the YOLO series, merge region proposal, classification, and regression into a single stage, significantly enhancing detection efficiency. Currently, such algorithms had been successfully applied to the recognition of mechanical fasteners for aerospace assembly and rough positioning of target workpieces for gear assembly, resulting in significant improvements of both efficiency and accuracy[9–13].

In practical tasks of micro-assembly, the detection aim is often to focus on key features that dominantly affect the assembly result, even if they only occupy a small area in scene. However, all parts in the scene are equally emphasized for primary deep neural network. Fortunately, the attention mechanism can provide a feasible solution, in which the system automatically chooses and emphasizes these significant regions or features based on the task requirements and image labels. Thus, the attention mechanism can obviously reduce the interference from irrelevant information and enhances the expression of important clues. Currently, there are several attention mechanisms available, including SENet, ECANet, and CBAM, all of which consist of either Channel Attention Module (CAM) or Spatial Attention Module (SAM). The CAM module learns global channel correlations, while the SAM module focuses on local structural positions[14,15]. Attention mechanisms have been utilized by Liu[16], Wu[17], Luan[18], et al. to inspect component defects, weld seams. Their results show improvements in model detection mean Average Precision etc.

Another issue worth noting is the interpretability. For many applications in precision manufacturing field, the requirements for error compliance and process traceability are very rigorous. However, these machine learning methods are generally considered as black box containing billions of parameters that takes an input vector and returns an output vector. Unfortunately, no definitive answer exists for what exactly happen in the black box. Therefore, the existential risks could result in serious failures of assembly, and it is also difficult to exactly give out remedial tactics to ensure that it will not occur again. To solve the problem, the explainable AI had been presented to intrinsically assess the black box, in which the visualization plays a key role. Gradient-weighted Class Activation Mapping (Grad-CAM) generates a coarse localization map by using the gradients of the target concept in the last layer, thereby highlighting important areas of the image that are used to predict the target concept. Recently, it has been frequently used to produce "visual explanations" for decisions from class of CNN-based models[19]. For example, Noh[20] and Lin[21] use the Grad-CAM

 doi:10.20944/preprints202404.0823.v1

visualization algorithm to evaluate the quality of anti-loosening coating on bolts and detect the bearing status of machine tools. The results show that this approach makes the decision results of their model more transparent and explainable.

Herein, in order to improve the performance of micro-assembly, we focus on the attention mechanism and visualizing CNNs for DL-based microscopic vision. The attention mechanism is combined with the single-stage detection algorithm to improve the sensitivity of detection algorithm to the target features. Visualization analysis of DL-based model is conducted using Grad-CAM to reveal the black box roperty and avoid potential risks during micro-assembly. The experiments about three typical assembly scenarios are performed to verify the learning effect and the detection accuracy.

2. Principle and Method

Figure 2 illustrates the overall framework and the work process. **Firstly**, sample images are collected *via* the microscope installed on home-made device as shown in Figure 4 and annotated with the software of LabelImg. Then the constructed datasets are inputted into the object detection model and classification model for training. **Secondly**, after model deployment, the home-made device performs visual detection tasks through two main processes. On one hand, it utilizes the microscope to capture images of microparts and inputs them into the object detection model to achieve feature recognition and localization, guiding the manipulator to carry out various operations. On the other hand, it captures images of micropart surfaces and droplets, feeding them into a Resnet50-based classification model for categorization.

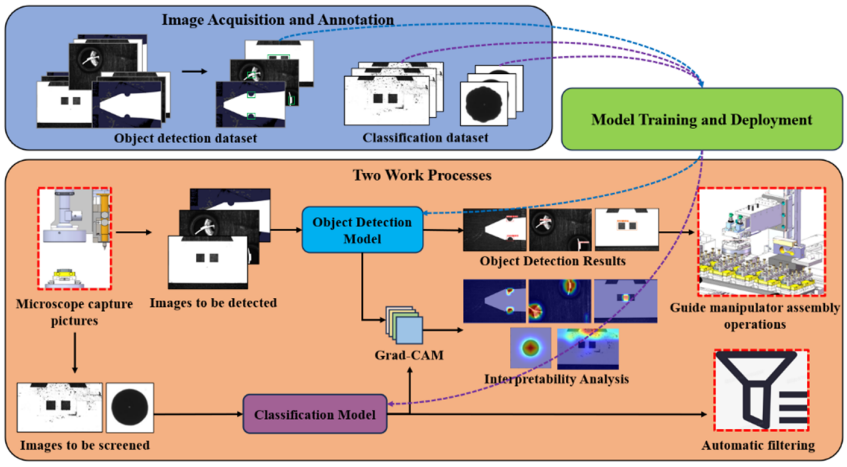

Figure 2. Overall framework and visual detection task.

Finally, the interpretability algorithms are utilized to generate heatmaps for visual analysis and to confirm whether there are potential visual detection risks during micro-assembly.

2.1. Enhanced YOLOv5s Model with Attention Mechanism

Here the single-stage YOLOv5s algorithm model as shown in Figure 3 is adopted. This algorithm utilizes the CSPDarknet53 backbone network, incorporating FPN feature pyramids and its detection head structure. Furthermore, the introduction of Adaptive Anchor Boxes effectively enhances detection performance and training speed.

 doi:10.20944/preprints202404.0823.v1

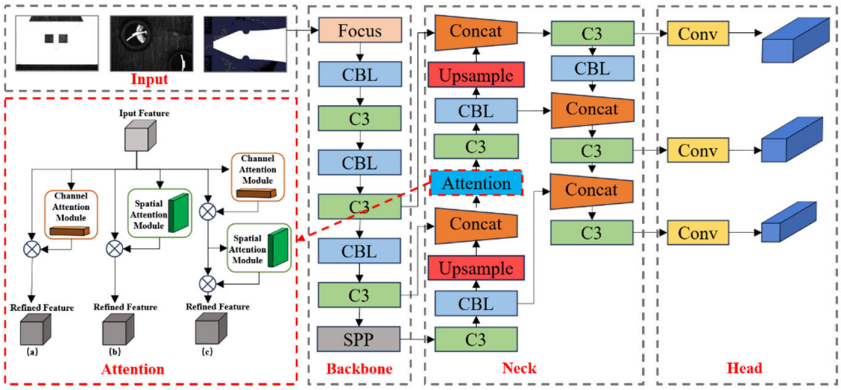

Figure 3. Structure of the YOLOv5s enhanced by attention mechanism.

Attention mechanisms, as pluggable modules, can be integrated into neural networks, enhancing the model's detection accuracy with minimal increase in computational complexity. Here we embed SENet, ECANet, CBAM, SimAM, and ShuffleAttention attention mechanisms into the neck of YOLOv5s model. This incorporation aims to enable the model to capture crucial features more accurately, thereby improving its generalization and accuracy. A comparative analysis will be conducted to select the optimal attention mechanism.

2.2. Implementation of Deep Learning Visualization Methods

Grad-CAM can evaluate the importance of each pixel in image for a specific class by computing the gradients of the model's last convolutional layer. These gradients are treated as weights for feature maps, which are used to weight the sum of the last convolutional layer's feature maps, resulting in a heatmap of class activations. The formula for Grad-CAM algorithm is described as Equations (1) and (2) [17].

$$L^{c}_{Grad\text{-}CAM} = ReLU\left(\sum_i \alpha_i^c A^i\right) \tag{1}$$

$$\alpha_i^c = \frac{1}{Z}\sum_{k=1}\sum_{j=1}\frac{\partial S_c}{\partial A^i_{kj}} \tag{2}$$

where S_c represents the predicted score by the network for class c, A^i_{kj} represents the data at position (k,j) in channel k of feature layer A, a_i^crepresents the weight for A^i, Z represents the width × height of the feature map.

Guided Grad-CAM is obtained by element-wise multiplication of the heat map generated by Grad-CAM and the gradient results ($\frac{\partial y^c}{\partial x}$) produced by Guided Backpropagation, resulting in a more detailed visualization effect. The calculation formula is described as Equation (3) [19].

$$R^{c}_{Guided\ Grad\text{-}CAM} = L^{c}_{Grad\text{-}CAM} \odot \frac{\partial y^c}{\partial x} \tag{3}$$

Where $\odot$ represents element-by-element multiplication operation.

In the heatmap, the red and highlighted areas indicate regions where the model has a significant impact on the classification prediction results, suggesting that the model relies more on the features extracted from these regions.

3. Experimental Setup

3.1. Experimental Platform

As shown in Figure 4, the experimental platform consists of microscope, dispensing module, 3-DOF manipulator, feeding robot, rotating platform, etc. The microscope consists of one CCD (MER-630-60U3/C-L, Resolution: 3088 × 2064, Pixel size: 2.40 μm) and one sleeve lens. For the microscope,

 doi:10.20944/preprints202404.0823.v1

the working distance is 65 mm, and its maximum visual field is 7.5 mm. The dispensing module is composed of a dispenser and a pneumatic slide table. To adjust these observed microparts, the rotating platform is constructed with a rotary stage and two linear stages. The 3-DOF manipulator is utilized for tasks such as picking, aligning, and assembling microparts.

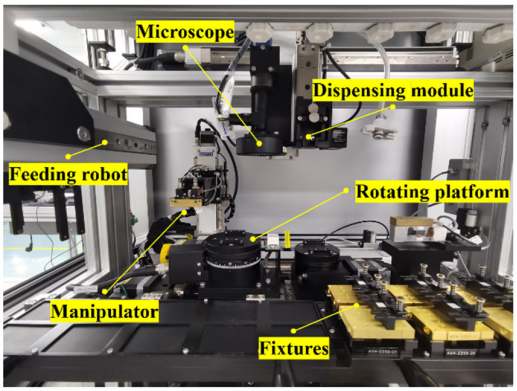

Figure 4. Experimental platform.

3.2. Dataset Preparation

Preparation of Object Detection Dataset. As shown in Figure 5, three specific regions on microparts for MEMS sensor are selected as interest regions for target recognition based on the formulated automated assembly strategy. With the assistance of experimental platform, the datasets comprising a total of 1538 images are collected. There are 1123 images in the training set, 230 in the validation set and 185 in the testing set. Approximately 30% of the images exhibit varying degrees of occlusion in the feature region due to the involvement of the manipulator.

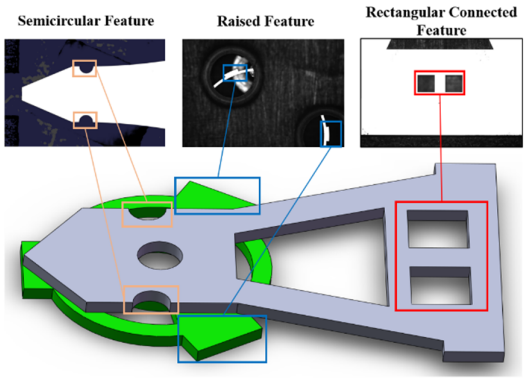

Figure 5. Feature regions of MEMS sensor.

Preparation of Droplet Classification Dataset. Adhesive droplets are dispensed on the surface of 300 sets of microparts using needles with an inner diameter of 0.2 mm. The parameters of the dispenser are set to pressure 150 kPa and dispensing time 100~700 ms. A dataset, which comprises 2462 droplet images, is selected from a large number of droplet samples, including 817 images of circular droplets, 740 images of serrated droplets, 544 images of gourd-shaped droplets, and 361 images of irregular droplets. The four types of droplets are shown in Figure 6.

 doi:10.20944/preprints202404.0823.v1

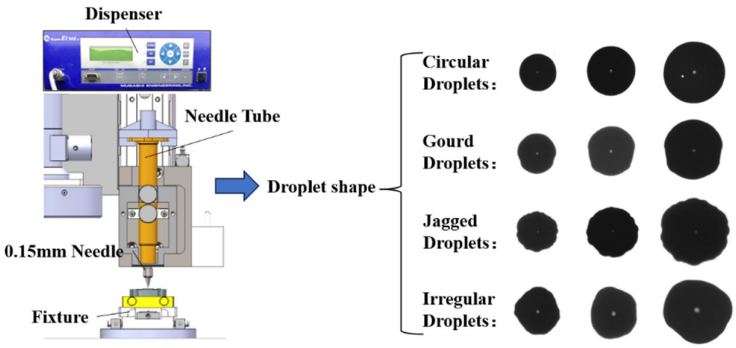

Figure 6. Various adhesive droplets.

Circular droplets exhibit good spreading properties, while gourd-shaped and Jagged droplets are caused by incomplete cleaning of the surfaces between microparts. Irregular-shaped droplets are often caused from bubbles and particles in adhesive. Generally, these imperfect droplets may lead to the poor connection strength or the additional stress. Thus, it is crucial to promptly identify and address those adhesive droplets before the contact connection of parts.

Preparation of Surface Classification Dataset. A total of 1286 images of micropart surfaces are collected for model training in this experiment. Among them, there are 655 images of contaminated micropart surfaces and 631 images of clean micropart surfaces.

Figure 7. Images of micropart surface. (a) Clean surface. (b) Contaminated surface.

3.3. Training Details

All experiments are conducted based on the Pytorch deep learning framework with Python. The hardware is configured with a NVIDIA GeForce GTX 3080Ti GPU and an Ubuntu operating system. A batch size of 16 is employed during training to expedite the process. Regarding the training epochs, all data are iterated over 80 times and the best model is saved.

For the object detection task, the training results of YOLOv5-SENet, YOLOv5-ECANet, YOLOv5-CBAM, YOLOv5-SimAM, and YOLOv5-ShuffleAttention models are shown in Figure 8.

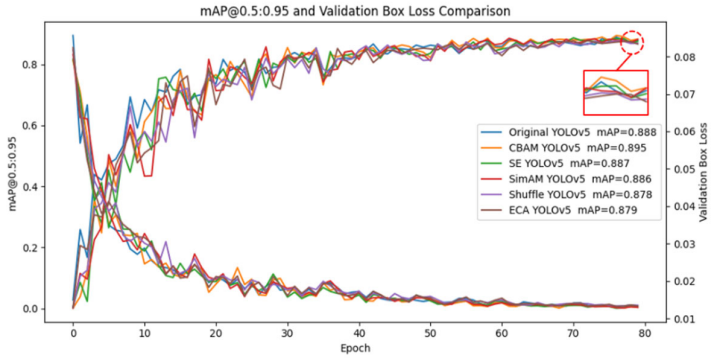

Figure 8. Training results of models.

 doi:10.20944/preprints202404.0823.v1

By the 25th epoch, the loss gradually stabilizes, and after 30 epochs, the loss decreases to around 0.015. Comparing to other models, it is observed that the YOLOv5-CBAM method achieves an mAP@0.5:0.95 metric improvement of approximately 2%. Therefore, it is chosen for subsequent testing of micropart target feature occlusion.

Precision (P), Recall (R), Average Precision (AP), and mean Average Precision (mAP) are used to comprehensively evaluate the models[16].

Precision in Equation (4) refers to the percentage of correctly predicted targets, with high precision indicating a low false detection probability.

$$P = \frac{TP}{TP+FP} \tag{4}$$

where TP represents correctly predicted positive targets, FP represents incorrectly predicted positive targets, and FN represents incorrectly predicted negative targets.

Recall in Equation (5) represents the percentage of correctly predicted targets among all targets.

$$R = \frac{TP}{TP+FN} \tag{5}$$

The average precision (AP) value in Equation (6) is the area under the precision-recall curve.

$$AP = \sum_n \{(r_{n+1} - r_n) p(r_{n+1})\} \tag{6}$$

nd mAP in Equation (7) is the average of AP values for all classes.

$$mAP = \frac{1}{C} \sum_{i=1}^{C} AP_i \tag{7}$$

where $mAP\_0.5$ represents the average precision (AP) when the IOU threshold is 0.5, and $mAP\_0.5{:}\,0.95$ calculates the mAP value by averaging the AP values calculated for IOU thresholds from 0.5 to 0.95 with an interval of 0.05. This metric focuses more on the precision of target localization.

4. Results and Discussion

4.1. Visualization Analysis of YOLOv5-CBAM

The visualization results of YOLOv5-CBAM and YOLOv5 model with Grad-CAM are shown in Figure 9. The original YOLOv5 model struggles to filter background information and tends to disperse attention across regions, resulting in poor visualization outcomes. In contrast, the YOLOv5-CBAM model can better focus attention on places around the target regions.

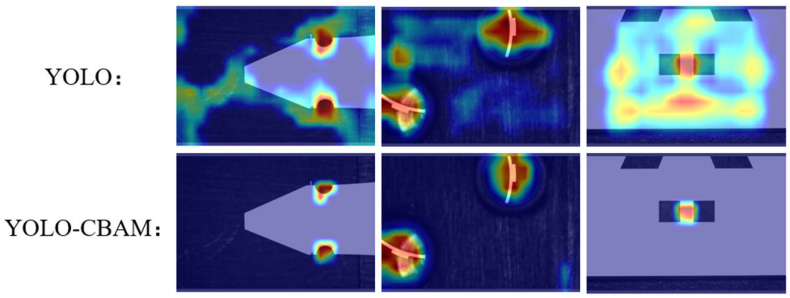

Figure 9. Visualizations of models with Grad-CAM.

To quantitatively evaluate its effectiveness, we utilized the Intersection over Union (IOU) metric to assess the degree of alignment between the highlighted regions in the heatmap and the ground truth (GT) regions. A higher IOU value indicates a superior match. IOU is defined as the ratio of the number of intersecting pixels between the actual target region and the highlighted region to the total number of pixels in their union as shown in Figure 10. The formula is described as Equation (8).

 doi:10.20944/preprints202404.0823.v1

$$Pixels\ IOU = \frac{GT\ Area \cap Highlight\ Area}{GT\ Area \cup Highlight\ Area} \times 100\% \quad (8)$$

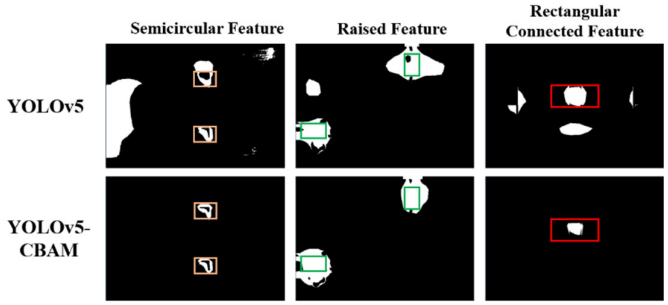

Figure 10. Segmentation images of highlighted regions.

From the results in Table 1, the YOLOv5-CBAM model achieves significantly higher IOU values on all the three types of target features compared to the original model, with an average improvement of 27.81%. Therefore, the YOLOv5 model with integrated CBAM can more proficiently filter out the useless background and accurately reflect the location of the target regions when generating heatmaps.

Table 1. Calculation results of IOU between GT regions and high-light regions.

Types	IOU(YOLOv5)	IOU(YOLOv5-CBAM)	Improvement Ratio
Semicircular Feature	17.16%	44.86%	**27.72%**
Raised Feature	10.48%	42.79%	**32.31%**
Rectangular Feature	10.63%	34.04%	**23.41%**

4.2. Occlusion Experiment of YOLOv5-CBAM

Firstly, the part features in the test set are subjected to four different degrees of occlusion processing, as illustrated in Figure 11a. Subsequently, the YOLO-CBAM model and YOLO model are selected to perform object detection on them. The respective detection performance metrics are presented in Table 2. Some sample detection results are shown in Figure 11b.

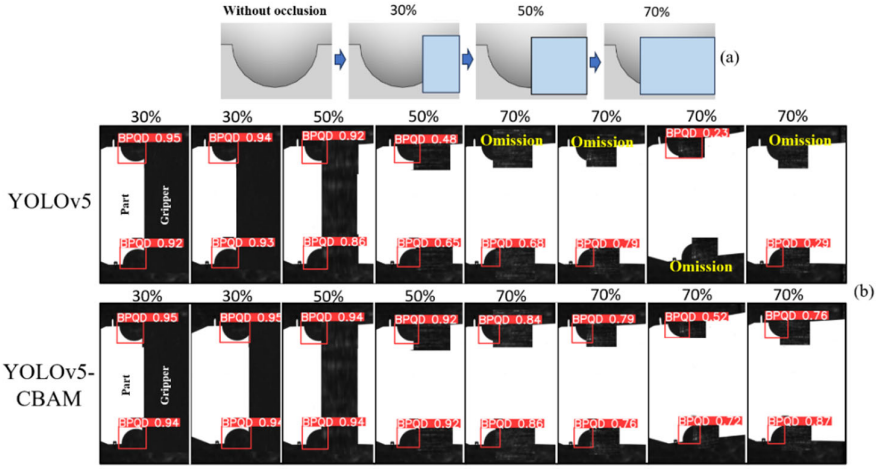

Figure 11. Object detection results of the model under various occlusions.

 doi:10.20944/preprints202404.0823.v1

Table 2. Performance metrics of model under various occlusions.

Occlusion rate	Model	P	R	mAP@0.5	mAP@0.5-0.95
Without Occlusion	YOLOv5	0.999	1	0.995	0.785
	YOLOv5-CBAM	0.999	1	0.995	0.792
30%	YOLOv5	0.997	1	0.995	0.737
	YOLOv5-CBAM	0.999	1	0.995	0.771
50%	YOLOv5	0.997	0.993	0.995	0.447
	YOLOv5-CBAM	0.998	1	0.995	0.461
70%	YOLOv5	0.907	0.751	0.933	0.258
	YOLOv5-CBAM	0.970	0.972	**0.979**	**0.317**

At no occlusion and 30% occlusion levels, the YOLOv5-CBAM model exhibits performance comparable to or even better than YOLOv5 across various metrics. However, as the occlusion level increases to 50% and 70%, the YOLOv5-CBAM model outperforms the YOLOv5 model in all metrics. Moreover, it can still accurately identify and outline the target objects with high confidence even with increased occlusion. In contrast, the original YOLOv5 model has a missing detection phenomenon in most images, and the recall rate is only 0.751. This suggests that the performance of the original YOLOv5 model significantly deteriorates when facing visual occlusion pressure.

4.3. Visualization Analysis of Droplet Classification Model

The classification accuracy of 91.43% on the test set is achieved by the trained droplet classification model. And the visualization results using Grad-CAM and Guided Grad-CAM are shown in Figure 12.

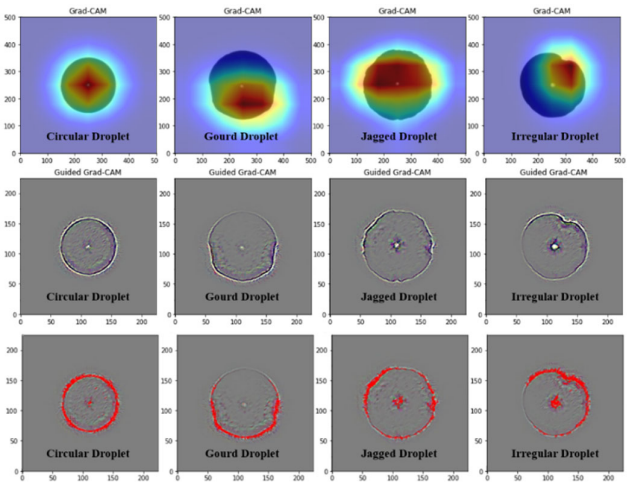

Figure 12. Grad-CAM and Guided Grad-CAM images of adhesive droplets.

 doi:10.20944/preprints202404.0823.v1

We collect the pixel coordinates of the highlighted and dark regions in all Guided Grad-CAM charts (225×225) by category and calculate the mean and standard deviation using the formulas Equations (9) and (10). The results are shown in Table 3.

$$\mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i, \mu_y = \frac{1}{N}\sum_{i=1}^{N} y_i \tag{9}$$

$$\sigma_x = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_x)^2},\ \sigma_y = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \mu_y)^2}, \tag{10}$$

where (x_i, y_i) represents the coordinates of each pixel, N is the total number of pixels, and $\mu(\mu_x, \mu_y)$ is the mean of the pixel coordinates.

Table 3. Mean and standard deviation of the high-light pixels distribution.

Type	Mean(μ_x, μ_y)	Standard Deviation(σ_x, σ_y)
Circular Droplet	(113.19,110.69)	(22.84, 22.38)
Gourd Droplet	(113.25,137.66)	(42.90, 24.35)
Jagged Droplet	(108.37,104.25)	(41.62, 33.16)
Irregular Droplet	(126.11,92.33)	(28.25,31.13)

Obviously, the droplet quality classification model in this study tends to focus on the edge and central features of the droplets during the task. For the circular droplets, their smooth and regular edges are the main features, resulting in a regular distribution of attention areas in the heatmaps, with the mean of the highlighted areas close to the center of the image and minimal standard deviation. For the gourd-shaped and serrated droplets, the distinctive features lie in the waist depression and serrated patterns, leading to a sinking mean center for gourd-shaped droplets and the largest standard deviation for serrated droplets. As for irregular-shaped droplets, due to their variable morphology and lack of specific features, the model needs to analyze the droplets as a whole, with particular attention to their irregular edges.

Therefore, the droplet classification model trained in this study is a reliable model with high classification accuracy and it can effectively identify the prominent features of various-shaped droplets, which are consistent with human visual discernment and judgment.

4.4. Analysis of Surface Classification Model

With the trained model, the classification accuracy of 98.45% is achieved on the test set. The Grad-CAM hierarchical visualization results are shown in Figure 13. In layer1, the heatmap is relatively scattered and blurry, without any obvious concentrated areas, indicating that the model does not have a clear judgment on stains or black spots. In layer2 and layer3, some distinct green and bright yellow areas start to appear, indicating localized attention areas and preliminary recognition of stain features. In layer4, the model's attention scale reaches its maximum, accurately covering the appearance and potential locations of stains and black spots. This process demonstrates the gradual improvement of model in feature extraction, ultimately achieving precise localization and classification recognition of stains. It also validates the effectiveness and the interpretability of surface classification model.

 doi:10.20944/preprints202404.0823.v1

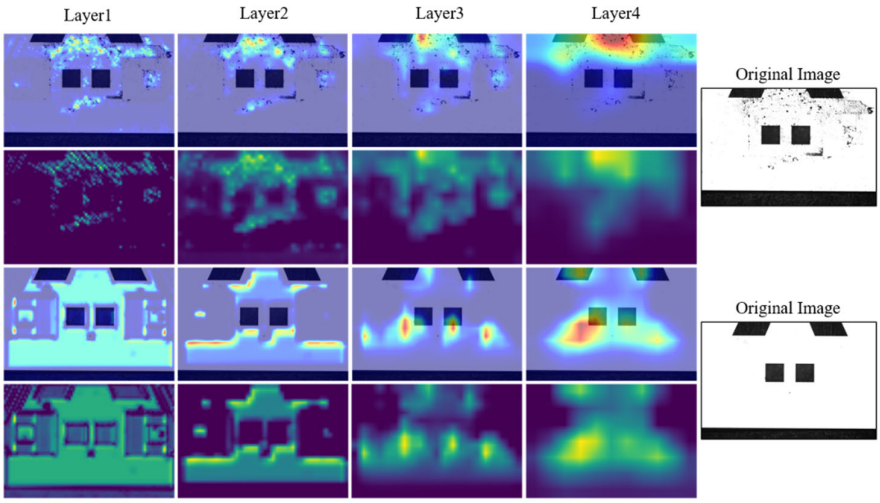

Figure 13. Visualization of different layers of the model.

5. Conclusions

To improve the microscopic-related tasks, we focus on the attention mechanism and visualizing CNNs for DL-based microscopic vision. The main results are as follows:

(1) The attention mechanism is combined with the YOLOv5 algorithm. With this hybrid algorithm, the robustness of the algorithm in scenes with occluded feature has been improved. The results demonstrate that even under 70% occlusion, the proposed algorithm has shown promising results with a mAP@0.5 of 97.9%, surpassing the original model by 4.6%.

(2) The visualization effect of YOLOv5-CBAM model is evaluated with Grad-CAM, which make the decision result more transparent, and the quantitative analysis results further verify the effectiveness of the attention mechanism in micropart feature localization.

(3) The trained micropart surface stain and droplet classification models both exhibit accuracies exceeding 90% in the experiments. And the results of the visual analysis align with human eye discrimination, validating the reliability of parts and droplets sorting.

Author Contributions: Conceptualization, Zheng Xu; Formal analysis, Yuchen Kong and Tongqun Ren; Investigation, Yanqi Wang; Methodology, Xinwei Zhao; Project administration, Xiaodong Wang; Software, Xinwei Zhao; Writing – original draft, Xinwei Zhao; Writing – review & editing, Zheng Xu and Xinwei Zhao.

Funding: This work was supported by the Defense Industrial Technology Development Program JCKY2022203B006.

Conflicts of Interest: All authors declare that there are no conflict of interests regarding the publication of this article.

References

1. M.B. Cohn, K.F. Bohringer, J.M. Noworolski, A. Singh, C.G. Keller, K.Y. Goldberg, R.T. Howe, Microassembly technologies for MEMS, Micromachining and Microfabrication Process Technology IV, 3511 (1998) 2-16, https://doi.org/10.1117/12.324300.
2. A.V. Kudryavtsev, G.J. Laurent, C. Clevy, B. Tamadazte, P. Lutz, Stereovision-based Control for Automated MOEMS Assembly, 2015 Ieee/Rsj International Conference On Intelligent Robots and Systems (Iros), (2015) 1391-1396
3. J. Liu, Y. Wang, Assembly planning and assembly sequences combination using assembly feature interference, Proceedings of 2007 10Th Ieee International Conference On Computer Aided Design and Computer Graphics, (2007) 545-548
4. M. Santochi, G. Fantoni, I. Fassi, Assembly of microproducts : State of the art and new solutions, Amst '05: Advanced Manufacturing Systems and Technology, Proceedings, (2005) 99-115
5. K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, 2016 Ieee Conference On Computer Vision and Pattern Recognition (Cvpr), (2016) 770-778, https://doi.org/10.1109/CVPR.2016.90.

 doi:10.20944/preprints202404.0823.v1

6. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going Deeper with Convolutions, 2015 Ieee Conference On Computer Vision and Pattern Recognition (Cvpr), (2015) 1-9, https://doi.org/10.1109/cvpr.2015.7298594.
7. S. Lei, F. He, Y. Yuan, D. Tao, Understanding Deep Learning via Decision Boundary, Ieee T Neur Net Lear, (2023), https://doi.org/10.1109/TNNLS.2023.3326654.
8. N. Adnan, F. Umer, Understanding deep learning - challenges and prospects, J Pak Med Assoc, 72 (2022) S66-S70, https://doi.org/10.47391/JPMA.AKU-12.
9. C. Liu, Y. Wu, J. Liu, Z. Sun, H. Xu, Insulator Faults Detection in Aerial Images from High-Voltage Transmission Lines Based on Deep Learning Model, Applied Sciences, 11 (2021) 4647, https://doi.org/10.3390/app11104647.
10. J. Ming, D. Bargmann, H. Cao, M. Caccamo, Flexible Gear Assembly with Visual Servoing and Force Feedback, 2023 Ieee/Rsj International Conference On Intelligent Robots and Systems (Iros), (2023) 8276-8282, https://doi.org/10.1109/IROS55552.2023.10341833.
11. F. Mushtaq, K. Ramesh, S. Deshmukh, T. Ray, C. Parimi, P. Tandon, P.K. Jha, Nuts&bolts: YOLO-v5 and image processing based component identification system, Eng Appl Artif Intel, 118 (2023) 105665, https://doi.org/10.1016/j.engappai.2022.105665.
12. Y. Wang, Z. Xu, Y. Yang, X. Wang, J. He, T. Ren, J. Liu, Deblurring microscopic image by integrated convolutional neural network, Precis Eng, 82 (2023) 44-51, https://doi.org/10.1016/j.precisioneng.2023.03.005.
13. Z. Xu, G. Han, H. Du, X. Wang, Y. Wang, J. Liu, Y. Yang, A Generic Algorithm for Position-Orientation Estimation With Microscopic Vision, Ieee T Instrum Meas, 71 (2022), https://doi.org/10.1109/TIM.2022.3176893.
14. S. Woo, J. Park, J. Lee, I.S. Kweon, CBAM: Convolutional Block Attention Module, Computer Vision - Eccv 2018, Pt VII, 11211 (2018) 3-19, https://doi.org/10.1007/978-3-030-01234-2_1.
15. M. Yao, Z. Min, Summary of Fine-Grained Image Recognition Based on Attention Mechanism, Thirteenth International Conference On Graphics and Image Processing (Icgip 2021), 12083 (2022), https://doi.org/10.1117/12.2623383.
16. J. Liu, X. Zhu, X. Zhou, S. Qian, J. Yu, Defect Detection for Metal Base of TO-Can Packaged Laser Diode Based on Improved YOLO Algorithm, Electronics-Switz, 11 (2022) 1561, https://doi.org/10.3390/electronics11101561.
17. L. Wu, L. Chen, Q. Zhou, J. Shi, M. Wang, A Lightweight Assembly Part Identification and Positioning Method From a Robotic Arm Perspective, Ieee Access, 11 (2023) 104866-104878, https://doi.org/10.1109/ACCESS.2023.3318016.
18. S. Luan, C. Li, P. Xu, Y. Huang, X. Wang, MI-YOLO: more information based YOLO for insulator defect detection, J Electron Imaging, 32 (2023) 043014-043014, https://doi.org/10.1117/1.JEI.32.4.043014.
19. R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization, 2017 Ieee International Conference On Computer Vision (Iccv), (2017) 618-626, https://doi.org/10.1109/ICCV.2017.74.
20. E. Noh, S. Hong, Automatic Screening of Bolts with Anti-Loosening Coating Using Grad-CAM and Transfer Learning with Deep Convolutional Neural Networks, Applied Sciences, 12 (2022) 2029, https://doi.org/10.3390/app12042029.
21. C. Lin, J. Jhang, Bearing Fault Diagnosis Using a Grad-CAM-Based Convolutional Neuro-Fuzzy Network, Mathematics-Basel, 9 (2021) 1502, https://doi.org/10.3390/math9131502.