ROYAL SOCIETY OF CHEMISTRY

PAPER

Check for updates

Lightweight target detection for large-field ddPCR images based on improved YOLOv5†

Xingyu Jin,[a] Jing Yang,[b] Xiaorui Jiang,[c] Zhenqing Li, [ID] *[a] Jinrong Shen,[d] Zhiheng Yu,[a] Cunliang Yang,[a] Fengli Huang,[e] Dunlu Peng,[a] Yoshinori Yamaguchi[f] and Jijun Feng [ID] *[a]

The large-field rapid nucleic acid concentration measurement system is capable of achieving one-time gene chip imaging with high resolution. However, it encounters challenges in the precise detection of positive microchambers, which is caused by factors such as reagent residue, uneven lighting, and environmental noise. Herein we proposed an improved, lightweight algorithm based on You Only Look Once (YOLOv5) for detecting the positive microchambers. We determined appropriate detection scales based on the target size distribution and utilized the bidirectional feature pyramid network (BiFPN) for efficient multi-scale feature fusion. To reduce model size without sacrificing performance, GhostConv, C3Ghost, and a simple, parameter-free attention module (SimAM) were integrated into the network, followed by network pruning. The improved YOLOv5 model was trained on a self-built dataset, and employed a partitioned fusion prediction strategy to detect large-field ddPCR images by self-developed software. In contrast to single-stage lightweight object detection algorithms, our model features a mere 1.5MB size while achieving 99.5% precision, 99.5% recall, and a 78.1% mAP(0.5 : 0.95), significantly reducing the system's demand for computing resources without compromising efficiency and accuracy.

Introduction

Droplet digital polymerase chain reaction (ddPCR) is a precise method for quantifying nucleic acids, widely used in clinical diagnosis,[1] environmental monitoring,[2] and bio-pharmaceuticals.[3] It partitions DNA samples into thousands of droplets or microchambers for individual PCR reactions, counts the positive microchambers after amplification, and utilizes Poisson distribution for absolute quantification.[4,5] The large-field rapid nucleic acid concentration measurement system was an innovative ddPCR method.[6] It integrated a microfluidic chip with 20 000 cylindrical microchambers, a white laser to improve dye excitation efficiency, and a large-field optical system for rapid imaging of the chip within 15 seconds,

eliminating the need for image stitching technology. The system effectively resolved the contradiction between target template resolution and field of view range, achieving one-time gene chip imaging with high resolution. However, when processing images using ImageJ software,[7] accurately counting low-quality positive microchambers remains a challenge due to factors such as residual reagents, uneven lighting, and environmental noise. Current methods depend on the thresholds of the pixels in an image, which imitates the distribution of fluorescence based on the grayscal.[8] Although multiple heuristic optimization techniques have been applied to refine threshold determination,[9] but minor threshold adjustment can lead to significant variations in counting the "positive" microchambers, especially when dealing with large numbers of microchambers. Therefore, improving the algorithm is crucial for an accurate and efficient large-field nucleic acid concentration measurement system.

In recent years, advancements in deep learning have spurred continuous innovation in ddPCR image processing algorithm. The Mask R-CNN had been employed to perform instance segmentation and boundary fitting of droplets, resulting in highly precise determination of microfluidic droplet size distribution.[10] Improved methods based on the combination of the Hough transform and deep learning techniques, such as the Attention DeepLabV3+ model[11] and convolutional neural networks,[12] have been utilized to accurately segment droplets of various sizes, even in low fluorescence intensity and blurred

[a]*Shanghai Key Laboratory of Modern Optical System, Engineering Research Center of Optical Instrument and System (Ministry of Education), School of Optical- Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China. E-mail: fjijun@usst.edu.cn; zhenqingli@163.com; Tel: +86 13917961036*

[b]*Faculty of Engineering, Anhui Sanlian University, Hefei, 230000, China*

[c]*National Key Laboratory of Electromagnetic Space Security, Tianjin 300308, China*

[d]*State Key Laboratory of ASIC and System, Fudan University, Shanghai 200433, China*

[e]*The Key Laboratory of Medical Electronics and Digital Health of Zhejiang Province, Jiaxing University, Jiaxing 314001, China*

[f]*Comprehensive Research Organization, Waseda University, Tokyo 162-0041, Japan*

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d5dd00006h

images. Additionally, Wei *et al.* first applied the Segment Anything Model for zero-shot microchamber segmentation in nucleic acid detection.[13] These methods achieved pixel-by-pixel differentiation between positive microchambers and background through image segmentation.[14] Conversely, object detection prioritizes the detection and localization of positive microchambers.[15] Recent research had utilized the YOLOv5s model to achieve binary classification and accurate identification of negative and positive droplets against complex backgrounds.[16] Another study had optimized the YOLOv5m model using region proposal network, enabling real-time automatic detection and classification of fluorescent images with good generalization performance.[17] These algorithms demonstrated the significant advantages of deep learning in processing ddPCR images. However, when dealing with higher-resolution images over a larger field of view, existing models become inefficient and complex. Although cloud servers or local high-performance workstations can provide powerful computing resources, they fail to meet the requirements for portability and independence in detection. Edge devices, on the other hand, are limited by computational, storage, and power constraints, making it challenging to support the efficient processing of complex models.

To address these challenges, we proposed an improved lightweight algorithm based on YOLOv5. By adjusting the feature fusion method and introducing GhostConv, C3Ghost, and SimAM models, alongside employing network pruning and partitioned fusion prediction strategy, this method significantly reduced reliance on computing resources, thereby enabling lightweight detection of high-resolution ddPCR images on edge devices. The deep learning model was trained using just two large-field images and combined with X-Anylabeling software

and the Albumentations library for dataset labeling and augmentation. To assess the efficacy of each enhancement, ablation experiments were conducted, and the improved algorithm was comprehensively compared with five mainstream single-stage object detection algorithms, demonstrating its outstanding performance.

Materials and methods

Preparation of dataset

We conducted a comprehensive large-field ddPCR experiment, which involved synthesizing primers, preparing PCR mixes, constructing ddPCR chips, amplifying PCR, configuring the system, and imaging.[6] There are 20 000 microchambers in the microfluidic chip (28 mm × 18 mm), which was imaged by CMOS camera (MARS-3140-3GM-P-03). Each chamber can hold 0.81 nL droplet with 270 pixels. Positive droplets typically appear as clear fluorescence signals in the images, with a radius of approximately 9 pixels, fluctuating by about 3 pixels. To aid in model training, the effective areas of images were cropped to 3900 × 2400 and segmented into 140 images (320 × 320).

To efficiently annotate images, X-Anylabeling software was employed for semi-supervised labeling.[18] After manually annotating a small subset of images to fine-tune the base model, we utilized the software's auto-labeling function to complete the remaining annotations. Finally, detailed manual proofreading was performed to ensure the accuracy and consistency of the labels. YOLO format files with category and coordinates were directly generated, and a corresponding VOC format dataset was compiled using a conversion program. Moreover, data diversity was enhanced with the Albumentations library's localized augmentation, which incorporated flipping,
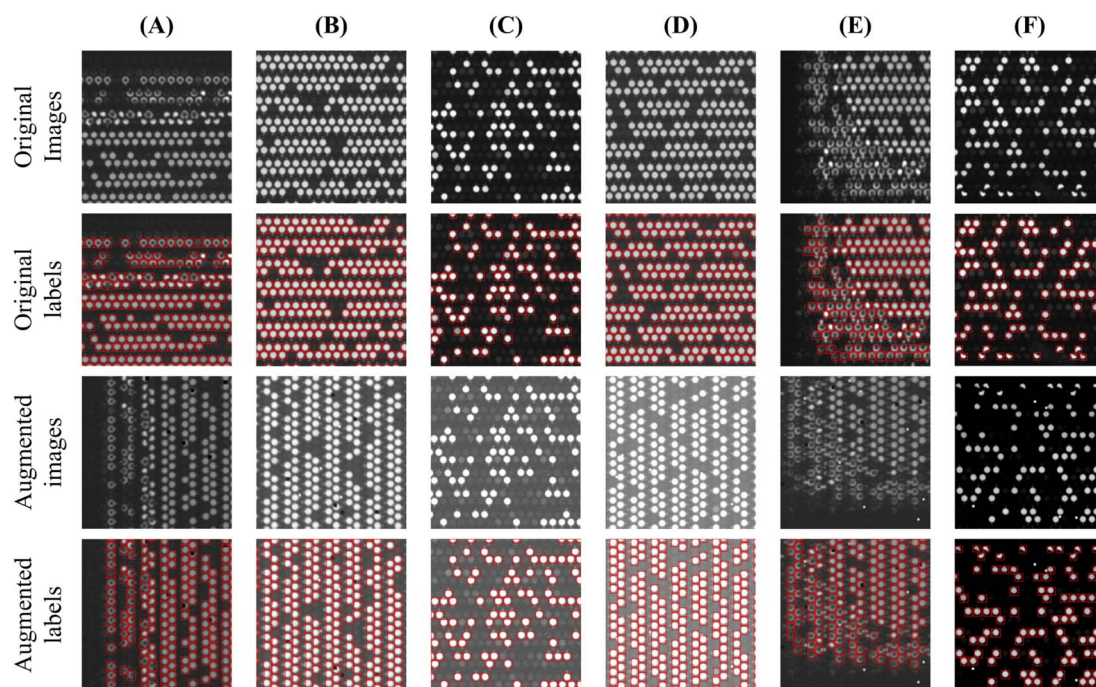

Fig. 1 Sample images from the self-built datasets. (A)−(F) represents six examples randomly selected from the self-built dataset.

transposing, foreign object occlusion, camera noise, random brightness and contrast, and motion blur, expanding the dataset to 280 images. Although the dataset size remains smaller than typical deep learning benchmarks, these techniques mimic realistic experimental variations—such as reagent distribution irregularities, chip impurities, lighting inconsistencies, as well as camera defocus and jitter—thereby enhancing the model's ability to generalize to unknown conditions. Although our experimental results on the current dataset are promising, continuous data collection, periodic model retraining, and rigorous cross-validation will further help ensure the model's robustness and reliability across various real-world scenarios. For model training and evaluation, we randomly split the dataset into training, validation, and test sets in a 7 : 2 : 1 ratio. Fig. 1 randomly showcases six examples from the self-built dataset.

Experimental settings and evaluation indicators

The improved YOLOv5 network was trained and tested on a Dell G15 laptop with the hardware configuration and operating environment outlined in Table S1.† To ensure reproducibility and comparability, we standardized the parameters: 600 epochs, 16 batch size, 640 × 640 image input size, 0.01 learning rate, and 0.937 momentum. Additionally, we adopted the Adam optimizer, a gradient-based optimization algorithm with adaptive learning rate capabilities,[19] to achieve faster convergence and avoid overfitting.

We assessed the network's performance using metrics such as precision ($P = \text{TP}/(\text{TP} + \text{FP})$), recall ($R = \text{TP}/(\text{TP} + \text{FN})$), and mean average precision at intersection over union (IoU) thresholds from 0.5 to 0.95 ($(\text{mAP}(0.5 : 0.95) = \sum_{i=1}^{C} \text{AP}_i/C)$). Precision ($P$) measures the accurate detection rate of positive microchambers, recall (R) captures the proportion of actual positives detected, whereas mAP(0.5 : 0.95) reflects the model's overall performance across a range of IoU thresholds, where true positives (TP) represent the count of correct detections, false positives (FP) denote the number of incorrect detections, false negatives (FN) are the undetected actual objects, AP_i is the average precision at each IoU threshold and C is the number of IoU thresholds. In addition, the number of parameters and model size directly affect the potential applications of the model within limited storage space. Floating-point operations (FLOPs) influence the inference speed and energy efficiency. These comprehensive evaluation metrics have provided valuable references for subsequent optimizations.

Improved YOLOv5 model and lightweight

YOLOv5 is a single-stage deep learning model for object detection, comprises a backbone, neck, and head network[20] (Fig. S1†). Its modular design allows independent component customization. The YOLOv5 (v7.0) network was chosen as the baseline due to the optimal balance between accuracy and speed. It features five models—n, s, m, l, x—with uniform architecture but varying depths and widths, enhancing performance and size as the model scales.[21] To accommodate the detection and analysis requirements of large-field ddPCR

images, the YOLOv5s model was selected, which employs a multi-scale detection strategy, utilizing feature maps of varying sizes to accommodate different object dimensions. It defaults to three detection layers—P3, P4, and P5—with feature map sizes of 80 × 80 (for targets ≥8 × 8 pixels), 40 × 40 (for targets ≥16 × 16 pixels), and 20 × 20 (for targets ≥32 × 32 pixels) respectively, based on a 640 × 640 input image.

The microchamber imaging of large-field ddPCR can reach up to 270 pixels, and the distribution of annotation box size is shown in Fig. S2.† In order to better match the targets, the model's backbone was optimized by removing the Conv module at layer 7 and the C3 module at layer 8, which excluded the P5 feature map computation. Subsequently, the neck network's output channels were reduced by half, and feature maps at layers (2,14), (4,10), (8,20), and (12,17) were strategically interconnected. Object detection was conducted by the head network on resized feature maps of 160 × 160, 80 × 80, and 40 × 40, matching the P2, P3, and P4 scales. The bidirectional feature pyramid network (BiFPN)[22] was also incorporated for superior feature fusion (Fig. 2).

The traditional method for feature extraction uses multiple convolutional kernels to convolve across all input feature map channels, demanding extensive parameters and computational power, particularly with large images or numerous kernels.[23] To improve its efficiency, we incorporated the GhostConv and C3Ghost modules from Ghostnet,[24] a lightweight deep learning model framework, followed by the network pruning.[25] The computational efficiency of GhostConv surpasses traditional convolution by approximately X times, where X represents the number of linear operations involved in the ghost convolution.[26] The GhostBottleneck, a lightweight bottleneck structure within the GhostConv, supersedes the cross stage partial module's bottleneck, leading to the novel C3Ghost module (Fig. 2B). This improvement is accompanied by the maintenance of the output feature map count and a notable reduction in model size. These modules, which are part of the widely used GhostNet architecture, are becoming increasingly popular in resource-constrained environments. The significant reduction in the number of parameters also reduces the potential risk of overfitting. However, the reduction in parameters often causes a decline in performance. To maintain the accuracy, a simple, parameter-free attention module (SimAM)[27] was introduced (Fig. 2D), because it performs better than the most representative SE and CBAM attention modules in most cases.[28] SimAM assesses the importance of neurons by optimizing an energy function and generates attention weights based on local self-similarity, thereby enhancing critical features and suppressing irrelevant ones. By incorporating the SimAM attention mechanism within the GhostConv framework (Fig. 2C), the resulting SimGhostConv module inherits the lightweight nature of GhostConv while gaining enhanced sensitivity and abstraction capabilities.

Partitioned fusion prediction strategy

Images with high resolution often leads to high computational complexity during both training and inference, thereby
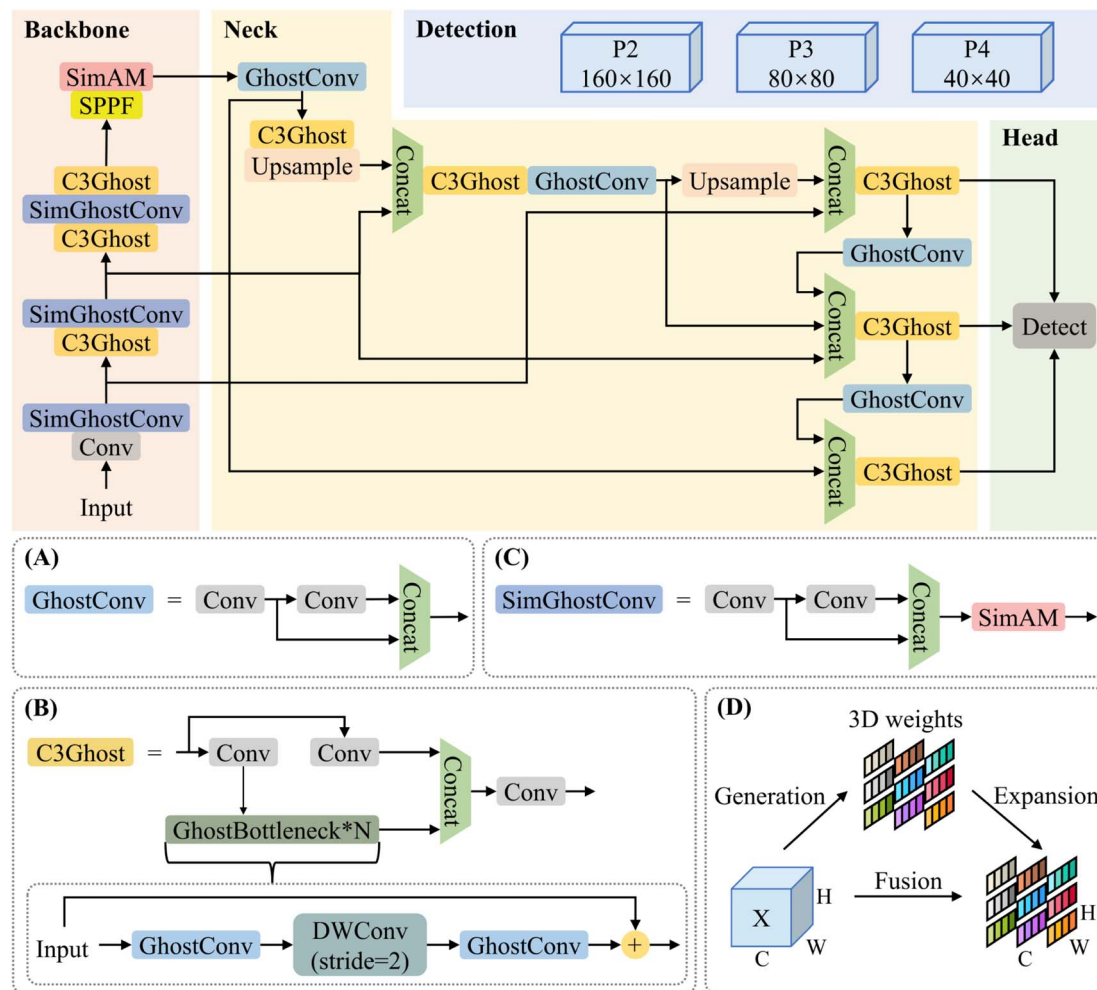
Fig. 2 Improved YOLOv5 architecture. It consists of the backbone, neck, and head, with detection showcasing the feature map sizes for P2, P3, and P4. (A) Structure of GhostConv module. (B) Structure of C3Ghost module, DWConv means that the depthwise convolution. (C)Structure of the SimGhostConv module, a specialized structure developed by integrating the SimAM attention mechanism into the GhostConv module. (D) SimAM module structure, with 'X' as the input feature map and 'C', 'H', 'W' indicating its channels, height, and width, respectively; 3D weights signify the attention dimensions.

affecting efficiency and deployment on edge devices. Thus, we employed a partitioned fusion prediction strategy to address it (Fig. 3). The essence of this strategy lies in dividing high-resolution images into several sub-regions with overlap.[29] The purpose of overlapping cropping is to detect the edges of each sub-image, thereby addressing the issue of inaccurate detection caused by cutting off microchambers at the edges.

When predicting a gene chip image of 3900×2400 resolution, sub-images are initially sized at 640×640, with a sliding cropping stride of 600, resulting in 28 sub-images. Each sub-image undergoes individual prediction to obtain detection results. Subsequently, the non-maximum suppression (NMS) algorithm[30] merges these results to eliminate duplicate counts in overlapping areas. The merged result represents object detection for the entire high-resolution image. Finally, the number of target pre-selection boxes represents the counting result, followed by quantitative analysis of target molecule concentration using the Poisson distribution. This strategy enables the model to flexibly utilize existing computational

resources during training and inference, significantly reducing the demands on system performance while maintaining operational efficiency and prediction accuracy.

Results and discussion

Comparison of multi-scale feature fusion methods

The study examined four detection scales provided by the YOLOv5s model: P2, P3, P4, and P5. Considering the distribution of target sizes (Fig. S2†), experiments focused exclusively on three scale sets: (P3, P4, P5), (P3, P4), and (P2, P3, P4). Subsequently, the performance of the optimal scale combination was combined with three feature fusion methodologies: feature pyramid network (FPN), path aggregation network (PANet), and BiFPN. Among the three experimental sets (Table S2†), the combination of (P2, P3, P4) demonstrated the highest performance, achieving 78.3% mAP(0.5 : 0.95), while reducing parameters to 1 701 526. Notably, the P5 detection scale exhibited the weakest correlation with positive microchambers,
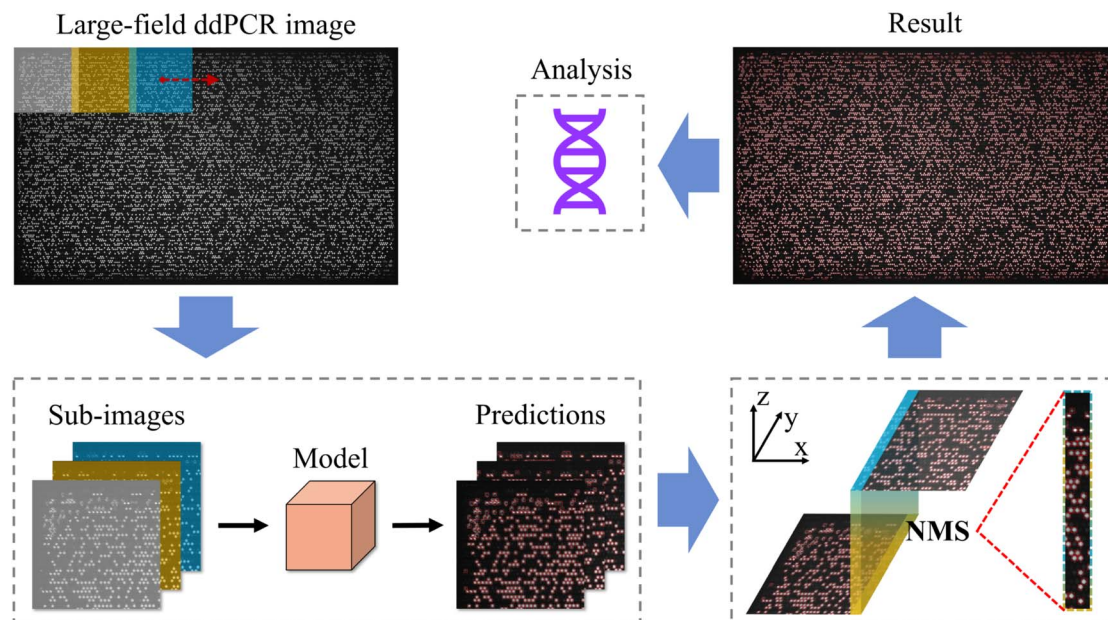
Fig. 3 The flowchart of partitioned fusion prediction strategy. The large-field ddPCR image serves as the input. Each colored slider represents a sub-image generated from each cropping operation. Each red detection box on the predictions corresponds to a detected droplet or microchamber. The blue-yellow gradient region represents the fusion process of the NMS algorithm. Finally, the number of positive droplets is counted based on the detection boxes in the result for further quantitative analysis.

while P2 displayed the strongest correlation, consistent with the distribution of target sizes in the self-built dataset. Table S3† presents the performance comparison among FPN, PANet, and BiFPN. While the introduction of BiFPN increased model parameters by 0.95%, it enriched semantic and detailed information across scales. As a result, the mAP(0.5 : 0.95) reached 78.5%, surpassing other methods.

Evaluation of model lightweighting

In the model lightweighting, we conducted the following experiments: introducing GhostConv and C3Ghost modules into backbone network, neck network, and backbone and neck networks, respectively. Subsequently, these experiments were compared with and without the insertion of SimAM (Table S4†). When SimAM was inserted separately, the model's performance significantly improved, which achieved 99.4% precision, 99.7% recall, and 78.7% mAP(0.5 : 0.95). When GhostConv and C3Ghost modules were simultaneously introduced in both the backbone and neck networks, the model achieved 99.5% precision, 99.3% recall, and 78.2% mAP(0.5 : 0.95). At this point, the model size was reduced to 2.9MB, reaching the optimal value among the previous three sets of experiments. When combined with SimAM attention mechanism, this experiment achieved 78.6% mAP(0.5 : 0.95), and the model parameters, size, and FLOPs remain unchanged, achieving the best overall performance. The results indicate that the inclusion of the SimAM enables the model to extract more focused features, significantly enhancing its ability to detect small target objects. After network pruning, the model maintains a high mAP(0.5 : 0.95) of 78.1%, with precision, recall, parameters, size, and FLOPs at 99.5%, 99.5%, 262 047, 1.5MB, and 2.6G, respectively.

The results of the ablation experiments, as shown in Fig. 4, highlight the consistent performance improvements across different model variants. Detailed data can be found in Table S5.† Additionally, the confusion matrices for the original and the final improved models are shown in Fig. S3.†

Analysis of partitioned fusion prediction and model comparison

We conducted partitioned fusion prediction on two additional gene chip images with a resolution of 4800 × 3000 to demonstrate the generalization and flexibility of our approach. Fig. 5 depicts the detection results of four different methods. Fig. 5A shows the raw images, Fig. 5B applies the traditional thresholding algorithm, Fig. 5C uses the original model for direct prediction, Fig. 5D employs our model for direct prediction, and Fig. 5E adopts our model for partitioned fusion prediction. Results demonstrate that the traditional detection method and the original YOLOv5s model exhibits significant missed detections and false positives when facing complex images with variable imaging environments. In contrast, the partitioned fusion prediction method not only accurately locates the positions of positive microchambers in high-resolution images, but also successfully addresses issues such as irregular shapes of residual reagents within microchannels, uneven lighting, and environmental noise. Furthermore, we had independently developed a detection software (see Movie S1 in the ESI†) using the PyQt5 module in Python, which provides valuable insights for detecting large-field ddPCR images and even higher-resolution images.

To comprehensively assess the improved model, we conducted comparative experiments on several representative
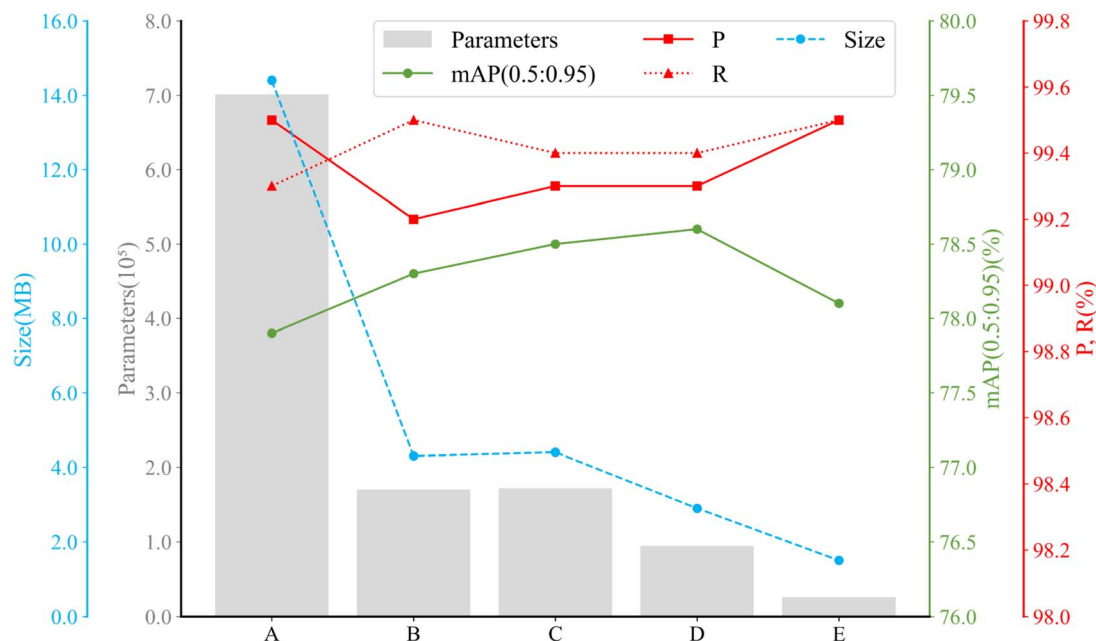
Fig. 4 Results of ablation experiment. Scheme A: original model without modifications. Scheme B: improved with scales P2, P3, P4. Scheme C: improved with BiFPN. Scheme D: improved with GhostConv, C3Ghost, and SimAM. Scheme E: improved with network pruning.
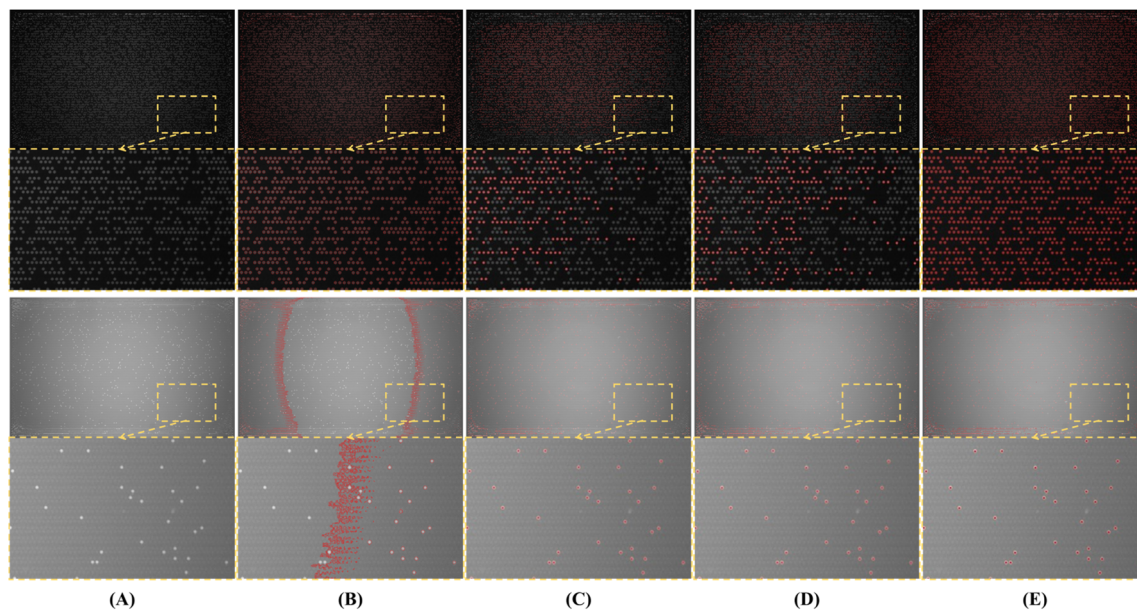

Fig. 5 Detection results of four different methods. (A) Raw. (B) Traditional thresholding algorithm, where red enclosed shapes (connected domains) represent positive detection results. (C) YOLOv5s-based direct prediction. (D) Improved YOLOv5s-based direct prediction. (E) Improved YOLOv5s-based partitioned fusion prediction. In (C–E), positive detections are marked with small red dots.

single-stage detection algorithms: SSD,[31] YOLOv3-tiny, YOLOv5s-ShuffleNetv2,[32] YOLOv5s-MobileNetv3,[33] and YOLOv5n. As shown in Fig. 6 and Table S6,† our model consistently demonstrates superior performance across various metrics when compared to these algorithms. Additionally, compared to YOLOv5-based methods such as YOLOv5s[16] and YOLOv5m,[17] our approach achieves higher detection accuracy with a smaller model size and is applicable to high-resolution

ddPCR images. Notably, although our experiments were based on a large-field ddPCR image dataset, the proposed method has broad applicability in other biomedical imaging domains. For example, the method has the potential to assist in rare cell identification in low-resolution fluorescence microscopy images by leveraging multi-scale detection and lightweight design. It may also contribute to cancer region detection in high-resolution whole-slide histopathology images through
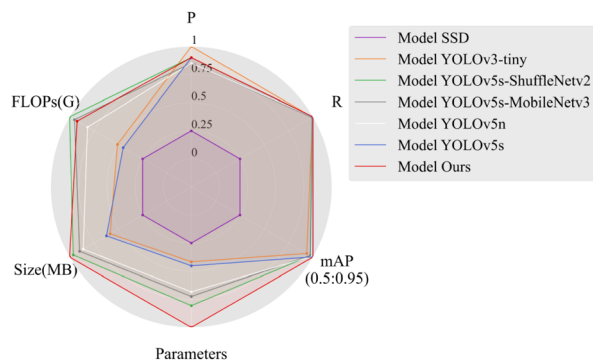
Fig. 6 Comparison of mainstream single-stage object detection algorithms. Through normalizing the data using logarithmic transformation and min–max scaling. In this process, the ranges of the three dimensions, parameters, size (MB), and FLOPs (G), are set to decrease gradually.

a partitioned fusion strategy. Additionally, it could be applied to microorganism counting in large-field water samples, potentially achieving high precision with low computational cost. Our core innovations are versatile and can be customized for various imaging conditions.

Conclusions

The paper presents an improved, lightweight algorithm based on YOLOv5 model for detecting and counting positive microchambers. We selected detection scales based on target size distribution and employ BiFPN for multi-scale feature fusion. To reduce model size while preserving performance, we integrated GhostConv, C3Ghost, and SimAM into the network, followed by pruning. We employed a partitioned fusion prediction strategy to detect large-field ddPCR images by self-developed software. The experimental results show that compared with the traditional YOLOv5s, there is a significant improvement, with parameters, model size, and FLOP reduced by 96.26%, 89.85%, and 83.54%, respectively, while achieving 99.5% precision, 99.5% recall, and 78.1% mAP(0.5 : 0.95). Compared to the existing models, it not only achieves higher detection accuracy but also lower complexity, exhibiting superior performance and efficiency. This is significant for the efficient conduct of nucleic acid concentration measurement and related research in fields such as medicine and bioscience. Future research will focus on further improving real-time detection performance and exploring the possibility of applying it to the counting of flowing fluorescent targets.

Data availability

The data supporting this article have been included as part of the ESI.† The dataset, model weights, and source code supporting this study are publicly available under the CC BY 4.0 license and can be accessed at: **https://data.mendeley.com/datasets/xw3zjwbw2w/2**. The corresponding DOI is the following: **https://doi.org/10.17632/xw3zjwbw2w.2**. Additional

details on data usage and implementation can be found in the README file within the repository.

Author contributions

Xingyu Jin: conceptualization, methodology, investigation, validation, formal analysis, writing–original draft; Jing Yang: conceptualization, investigation, writing; Jijun Feng: conceptualization, supervision, writing–review & editing, project management, funding acquisition; Xiaorui Jiang: methodology, investigation, validation, and editing; Zhenqing Li: conceptualization, supervision, review & editing; Jinrong Shen: conceptualization, curated; Zhiheng Yu: methodology, investigation, validation, and editing; Cunliang Yang: conceptualization, supervision, writing–review & editing; Fengli Huang: conceptualization, supervision, writing–review & editing; Dunlu Peng: conceptualization, supervision, writing–review & editing; Yoshinori Yamaguchi: conceptualization, supervision, writing-review & editing; all authors have given approval to the final version of the manuscript.

Conflicts of interest

The authors declare no conflict of interest.

Acknowledgements

References

1 M. Li, F. Yin, L. Song, X. Mao, F. Li, C. Fan, *et al.*, Nucleic Acid Tests for Clinical Translation, *Chem. Rev.*, 2021, **121**(17), 10469–10558.

2 H.-G. Lee, H. M. Kim, J. Min, C. Park, H. J. Jeong, K. Lee, *et al.*, Quantification of the paralytic shellfish poisoning dinoflagellate Alexandrium species using a digital PCR, *Harmful Algae*, 2020, **92**, 101726.

3 B. F. McCarthy Riley, H. T. Mai and T. H. Linz, Microfluidic Digital Quantitative PCR to Measure Internal Cargo of Individual Liposomes, *Anal. Chem.*, 2022, **94**(20), 7433–7441.

4 Y. Hou, S. Chen, Y. Zheng, X. Zheng and J.-M. Lin, Droplet-based digital PCR (ddPCR) and its applications, *TrAC, Trends Anal. Chem.*, 2023, **158**, 116897.

5 D. Xu, W. Zhang, H. Li, N. Li and J.-M. Lin, Advances in droplet digital polymerase chain reaction on microfluidic chips, *Lab Chip*, 2023, **23**(5), 1258–1278.

6 J. Shen, J. Zheng, Z. Li, Y. Liu, F. Jing, X. Wan, *et al.*, A rapid nucleic acid concentration measurement system with large field of view for a droplet digital PCR microfluidic chip, *Lab Chip*, 2021, **21**(19), 3742–3747.

7 A. B. Schroeder, E. T. Dobson, C. T. Rueden, P. Tomancak, F. Jug and K. W. Eliceiri, The ImageJ ecosystem: Open-source software for image visualization, processing, and analysis, *Protein Sci.*, 2021, **30**(1), 234–249.

8 M. M. Saeid, Z. B. Nossair and M. A. Saleh, A Fully Automated Spot Detection Approach for cDNA Microarray Images Using Adaptive Thresholds and Multi-Resolution Analysis, *IEEE Access*, 2019, **7**, 80380–80389.

9 S. Pare, A. Kumar, G. K. Singh and V. Bajaj, Image Segmentation Using Multilevel Thresholding: A Research Review, *Iran. J. Sci. Technol., Trans. Electr. Eng.*, 2020, **44**(1), 1–29.

10 S. Zhang, X. Liang, X. Huang, K. Wang and T. Qiu, Precise and fast microdroplet size distribution measurement using deep learning, *Chem. Eng. Sci.*, 2022, **247**, 116926.

11 Y. Song, S. Lim, Y. T. Kim, Y. M. Park, D. A. Jo, N. H. Bae, *et al.*, Deep learning enables accurate analysis of images generated from droplet-based digital polymerase chain reaction (dPCR), *Sens. Actuators, B*, 2023, **379**, 133241.

12 H. Yang, J. Yu, L. Jin, Y. Zhao, Q. Gao, C. Shi, *et al.*, A deep learning based method for automatic analysis of high-throughput droplet digital PCR images, *Analyst*, 2023, **148**(2), 239–247.

13 Y. Wei, S. Luo, C. Xu, Y. Fu, Q. Dong, Y. Zhang, *et al.*, SAM-dPCR: Real-Time and High-throughput Absolute Quantification of Biological Samples Using Zero-Shot Segment Anything Model, *arXiv*, 2024, preprint, arXiv:2403.18826, DOI: **10.48550/arXiv.2403.18826**.

14 S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz and D. Terzopoulos, Image Segmentation Using Deep Learning: A Survey, *IEEE Trans. Pattern Anal. Mach. Intell.*, 2022, **44**(7), 3523–3542.

15 Y. Xiao, Z. Tian, J. Yu, Y. Zhang, S. Liu, S. Du, *et al.*, A review of object detection based on deep learning, *Multimed. Tool. Appl.*, 2020, **79**(33), 23729–23791.

16 Y. Yao, S. Zhao, Y. Liang, F. Hu and N. Peng, A one-stage deep learning based method for automatic analysis of droplet-based digital PCR images, *Analyst*, 2023, **148**(13), 3065–3073.

17 Y. Wei, S. M. T. Abbasi, N. Mehmood, L. Li, F. Qu, G. Cheng, *et al.*, Deep-qGFP: A Generalist Deep Learning Assisted Pipeline for Accurate Quantification of Green Fluorescent Protein Labeled Biological Samples in Microreactors, 2024, **8**, 3, 2301293.

18 W. Wang, Advanced Auto Labeling Solution with Added Features, *Github*, 2023, [cited 2023 Oct 26], Available from: **https://github.com/CVHub520/X-AnyLabeling**.

19 S. Li, S. Zhang, J. Xue and H. Sun, Lightweight target detection for the field flat jujube based on improved YOLOv5, *Comput. Electron. Agric.*, 2022, **202**, 107391.

20 J. Redmon, You only look once: Unified, real-time object detection, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

21 O. L. García-Navarrete, O. Santamaria, P. Martín-Ramos, M. Á. Valenzuela-Mahecha and L. M. Navas-Gracia, Development of a Detection System for Types of Weeds in Maize (Zea mays L.) under Greenhouse Conditions Using the YOLOv5 v7.0 Model, *Agriculture*, 2024, **14**(2), 286.

22 M. Tan, R. Pang and Q. V. Le, Efficientdet: Scalable and efficient object detection, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 10781–10790.

23 F. Kong, K. Hu, Y. Li, D. Li and S. Zhao, Spectral–Spatial Feature Partitioned Extraction Based on CNN for Multispectral Image Compression, *Remote Sens.*, 2021, **13**(1), 9.

24 K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu and C. Xu, Ghostnet: More features from cheap operations, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 1580–1589.

25 A. Kumar, A. M. Shaikh, Y. Li, H. Bilal and B. Yin, Pruning filters with L1-norm and capped L1-norm for CNN compression, *Appl. Intell.*, 2021, **51**(2), 1152–1160.

26 S. Chen, Y. Liao, F. Lin and B. Huang, An Improved Lightweight YOLOv5 Algorithm for Detecting Strawberry Diseases, *IEEE Access*, 2023, **11**, 54080–54092.

27 L. Yang, R.-Y. Zhang, L. Li and X. Xie, SimAM: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks, *Proceedings of the 38th International Conference on Machine Learning*, 2021, 139, pp. 11863–11874.

28 H. You, Y. Lu and H. Tang, Plant Disease Classification and Adversarial Attack Using SimAM-EfficientNet and GP-MI-FGSM, *Sustainability*, 2023, **15**(2), 1233.

29 Q. Li, W. Yang, W. Liu, Y. Yu and S. He, From contexts to locality: Ultra-high resolution image segmentation via locality-aware contextual correlation, *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 7252–7261.

30 R. Rahman, Z. Bin Azad and M. Bakhtiar Hasan, Densely-Populated Traffic Detection Using YOLOv5 and Non-maximum Suppression Ensembling, *Proceedings of the International Conference on Big Data, IoT, and Machine Learning*, 2022, pp. 567–578.

31 W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, *et al.*, SSD: Single Shot MultiBox Detector, *Computer Vision – ECCV*, 2016, vol. 2016, 21–37.

32 N. Ma, X. Zhang, H.-T. Zheng and J. Sun, Shufflenet v2: Practical guidelines for efficient conn architecture design, *Proceedings of the European conference on computer vision (ECCV)*, 2018, 116–131.

33 A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, *et al.*, Searching for mobilenetv3, *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 1314–1324.