YOLORG: A multi-scale intestinal Organoid detection algorithm

Zhipeng Shang[1,2], Xun Deng[1], Tianyu Sun[3], Feng Tan[4], Lun Hu[1], Xi Zhou[1]✉, Pengwei Hu[1]✉

1 Xinjiang Technical Institute of Physics and Chemistry, Chinese Academy of Sciences, Urumqi, China
2 *The School of Software, Xinjiang University, Urumqi, China*
3 *The First Affiliated Hospital of Chongqing Medical University, Chongqing, China*
4 *AI and Quantum Lab, Merck KGaA, Darmstadt, Germany*

Abstract. Intestinal organoids exhibit valuable research potential in drug screening and disease modeling, yet their morphological diversity poses challenges for accurate morphological information acquisition. Traditional fluorescent labeling assays risk compromising organoid integrity while bounding box detection methods fail to capture intricate organoid details. The large size of gut organoid image datasets and the limitations of manual classification hinder research efficiency. While deep learning has progressed in image processing, complex structures like organoids, with varying shapes and sizes, remain challenging. The paper introduces YOLORG, an intestinal organoid detection method that utilizes a multi-scale feature extraction module to fuse organoid attributes. This approach mitigates background interference and image noise, enhancing organoid detection accuracy and robustness.

Keywords: Intestinal Organoids, Detection, Multi-scale.

1 Introduction

Organoids are simple models made using cells grown in a dish that mimic the structure and functions of real tissues[1]. Their spatial structure resembles that of specific organs, enabling them to replicate physiological properties [2,3] partially. Thus, organoids find extensive applications in modeling organ development, studying disease mechanisms, basic science research, drug discovery, and regenerative medicine [4]. With technological innovation, digital therapy (DTx) is moving toward intelligence, and its personalized treatment not only broadens the medical field, but also injects new vitality into organoid research[5].

To comprehend their growth characteristics and developmental processes, acquiring accurate information is vital. Brightfield images, as organoid culture methods advance, provide invaluable insights into growth, development, morphology, and structure [6]. Analyzing these images offers an intuitive understanding of organoid growth status, supporting culture condition optimization and enhancing culture efficiency.

Image data analysis was often limited by organoid heterogeneity, with issues like

defocus, overlaps, and occlusions reducing recognition accuracy [7,8]. Morphological differences during organoid growth also complicated analysis [9]. Traditional cell fluorescence staining methods were inapplicable [10, 11], and manual labeling in bright field images was inefficient and prone to bias [12]. Therefore, a stable, high-speed, automated detection platform was needed for large-scale organoid analysis [13].

This paper introduces YOLORG, a novel model designed to address challenges in detecting intestinal organoids. YOLORG utilizes a combination of techniques to enhance accuracy and efficiency in organoid detection. This work's contributions are summarized as follows:

—Multi-scale feature extraction enhanced organoid recognition for varying sizes and morphologies. This marks the first time Res2Net optimized the YOLOv5 backbone, reducing misdetections by improving background processing.

—We designed the Bineck module to dynamically allocate attention to organoid transparency and bud counts. By fusing and processing multi-scale feature maps, it improves detection accuracy.

—To enhance intestinal organoid sample accuracy, we propose a new detection model structure with a decoupling head, decomposing detection into regression and classification.

2 Related Work

2.1 Organoid Detection

The lining of the intestines has been studied a lot to learn about how stem cells work and to understand important diseases like colorectal cancer. In the past, it was thought that in gut organoids, stem cells changed as they moved into the upper part of the crypt[14]. This process was driven by cells multiplying at the bottom of the crypt, and then moving upward towards the villus region. The villus region is where most differentiated cells are located, and they ultimately migrate into the lumen [15]. In the past, Domeѭech-Moreno et al. [25] created an annotated image dataset of intestinal organoids. These organoids were manually categorized into four types based on their morphology. The morphological differences between cyst organoids and sphere organoids are subtle, making traditional manual classification challenging. Fig. 1 shows the morphology of the four intestinal organoids. Deep learning networks are making progress in the single task of intelligent healthcare, and medical diagnosis requires multi-modal data fusion. Organoid research involves complex biological processes, and multimodal analysis has great potential in organoid field[16].

Deep learning methods greatly enhanced organoid analysis accuracy and speed compared to manual methods, covering classification, detection, segmentation, and tracking [17]. LiViT-Net [18] efficiently addresses the challenges of retinal vessel segmentation through the integration of MobileViT+ and local representations, demonstrating real-time performance and strong robustness.Deep-Orga [20], based on Yolox[19], automated feature extraction but required high computational resources. D-CryptO[21] using XCeption model showed auto-extraction and classification but was

sensitive to hyperparameters. CAMPEOD[22] automated organoid segmentation via cross-attention. MACPNet[23] stands out for its proficiency in capturing multi-scale data, aligning perfectly with the inherent diversity of distance dependencies observed in organoid images. MacrOrga[24] proficiently captures long-range pixel relationships within organoid images, thereby enhancing the accuracy of organoid segmentation. Tellu[25], based on YOLOv5, could classify organoids but accuracy was limited. Traditional methods relied on manual tuning, while CNN-based methods struggled with small objects. We proposed YOLORG, a multi-scale detection method, to address challenges in small object detection and noise interference, significantly improving accuracy and robustness.

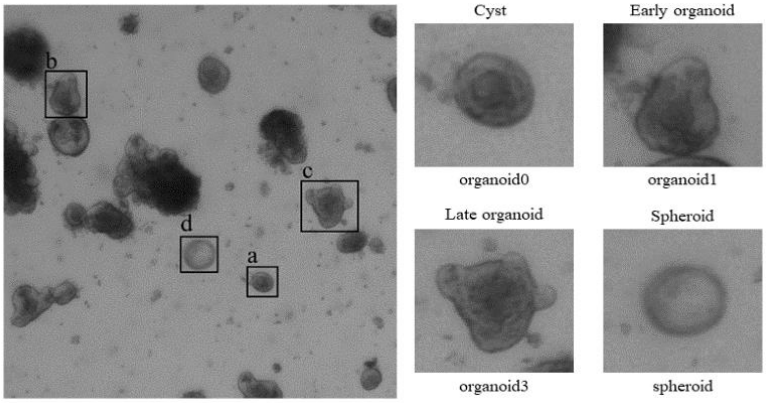

Fig. 1. Sorting out intestinal organoids. Very young organoids (organoid0) are small and do not have any growths yet. Early organoids (organoid1) are slightly older and have 1-2 small growths. Old organoid3 with 3 or more growths. Under specific circumstances, separate crypts won't grow into bud organoids, but take on a round shape (spheroids). Describable as having big, circular, and thin walls.

2.2 YOLOv5

With the rise of organoid technology, automated organoid detection has become a critical research area. YOLOv5, an advanced target detection algorithm, showed promise in this field. The introduction of OrgaQuant [7] established the foundation for deep learning in organoid image analysis. Du, X et al. [26] combined YOLOv5 and u-net to achieve accurate organoid detection and analysis, including crypt counting and skeleton extraction. Kshirsagar, V. et al. [27] utilized the Yolo algorithm with RANSAC to improve organoid detection accuracy and stability by detecting and removing outliers.

Liu, C et al. [28] tailored YOLOv5 for organoid detection by introducing an improved SE module. This enhancement improved the network's feature discrimination ability, leading to better leukocyte detection performance. This approach provides insights for similar tasks.

Cui, C et al. [29] integrated the Swin Transformer into the YOLOv5 backbone,

achieving superior blood cell detection results. This approach offers a new perspective for the advancement of organoid detection technology. Haq, I et al. [30] combined ResNet with the Yolo architecture for cancer cell detection. The optimized detection performance provides a valuable tool for cancer research. Liu, W et al. [31] built a framework based on YOLOv5 that incorporates a differentiable image processing module and a small CNN. This framework improves the detection of small objects, making it highly valuable for tasks involving numerous small targets, such as organoid detection.

Compared with previous tasks, organoid detection in bright-field intestinal organoid images possessed unique challenges. The large number of class organoids, their small size, and the presence of impurities all made the detection task more complex. To address these characteristics, we previously proposed the YOLORG model. This model achieved robust detection of class organoids by constructing hierarchical connections and introducing a multi-branch structure.

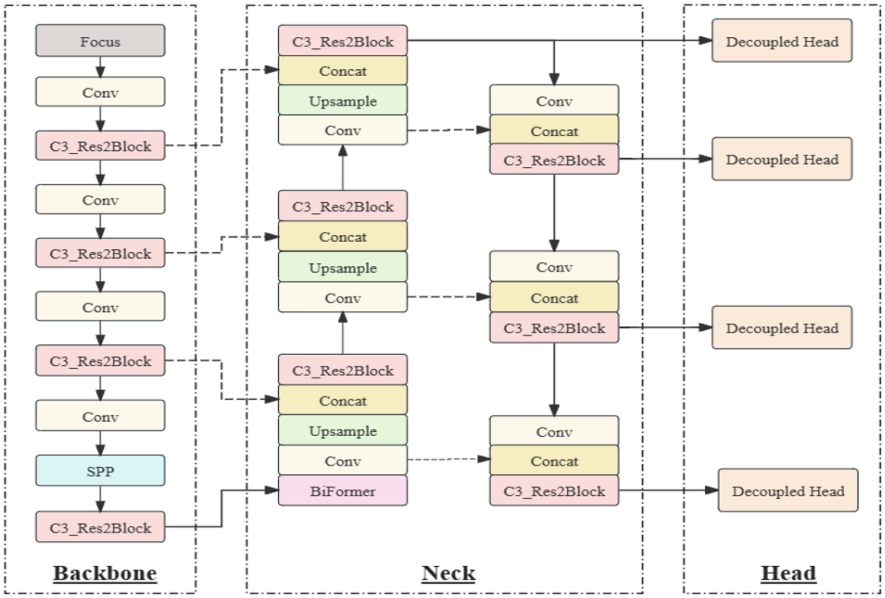

Fig. 2. Proposed YOLORG Structure: It extracts features from organoid images, and enhances them with a two-layer routing attention module to focus on small organoids. Then, multi-scale feature maps are fused to utilize both low- and high-level information for small target detection. The network ends with a decoupling head for target detection.

3 YOLORG

The figure shows a summary of YOLORG. YOLORG is made up of three main parts: backbone module, Bineck module, and decoupling head module. If you give a picture

of an organoid, the backbone will find its features with different levels of detail. The organoid picture is put into the backbone network to get the feature.

The text in Section 3.1 talks about using information and a module to process features at different sizes to make the model better at understanding different features. In Section 3.2, it uses a two-way attention method to better understand the connections between different parts of the sequence.In Section 3.3, the decoupling header is processed in different parts of the network to make the detection process better.

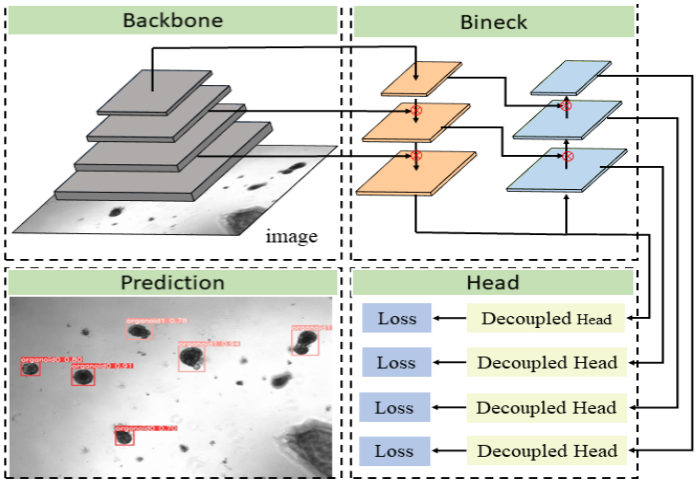

Fig. 3 Characteristic pyramid structure

3.1 Multi-modal Feature Extraction

Res2Net[32] is a type of deep neural network that can capture details at different sizes because of its special feature pyramid structure. We take a smaller group of filter banks with w channels each and use them instead of a larger 3x3 filter with n channels. We can divide x into 4 parts and use s=4 for simplicity. This is shown in Fig 4.

Filter sets are linked together in a step-by-step way to make the final picture look more detailed. One group of filters removes things from a group of pictures. The pictures from before are sent to another set of filters with more pictures. This is done many times until all the input maps are processed.

Groups are combined and sent to another filter to fully merge the information. From the input maps to the output maps, there are many possible paths.

Maps increase in size as they pass through the 3x3 filters, and this results in getting different feature sizes. The Res2Net formula is as follows:

$$Split(X) = \{x_1, x_2, \dots, x_s\} \tag{1}$$

$$Conv1(X) = Conv_{1\times1}(x, C_{width} \times s) \tag{2}$$

$$Conv3(X_i) = Conv_{3\times3}(x_i, C_{width}) \quad (3)$$

$$Concat(Conv3(X_i)) = Concat(Conv3(X_1), Conv3(X_2), \dots, Conv3(X_s)) \quad (4)$$

$$y = \begin{cases} Conv_{1\times1}\big(Concat\big(Conv3(X_i)\big), C_{width} \times s, C_{out}\big) + X, & if\ short\ is\ True \\ Conv_{1\times1}\big(Concat\big(Conv3(X_i)\big), C_{width} \times s, C_{out}\big), & if\ short\ is\ False \end{cases} \quad (5)$$

where x is the tensor, dividing the tensor into s parts, C_{width} is the number of output channels, short is the residual mapping, and y is the final output of the residual module.

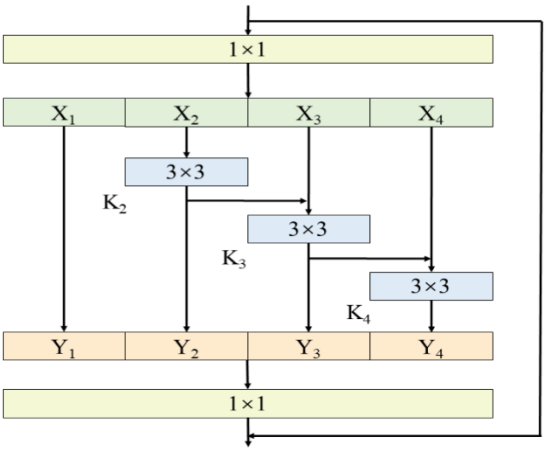

Fig. 4. Structure of Res2Net network

3.2 Adaptive Regional Attention Fusion

In YOLORG, Bineck fuses deep and shallow features but faces information redundancy issues. The subtle morphological difference between cystic and sphere organoids compounds the recognition challenge. We introduce the BRA [33] module in Bineck's first layer, where location routers assign image blocks based on rules. Upper routers capture global context, while lower routers focus on local details. BRA employs overlapping block embedding initially and blocks merging later to reduce spatial resolution while increasing channels. The basic principle of BRA is shown in Fig. 5.

The input feature map X is divided into non-overlapping regions, with each region containing $\frac{HW^2}{S^2}$ feature vectors. By reshape, $X^r \in R^{S^2\times\frac{HW}{S^2}\times C}$,we derive $Q, K, V \in R^{S^2\times\frac{HW}{S^2}\times C}$,using linear projection:

$$Q = X^r W^q, K = X^r W^k, V = X^r W^v \quad (6)$$

The letters $W^q, W^k, W^v \in R^{C\times C}$ are used to show the weights used to project Q, K, and V.We use a directed graph to figure out which regions should connect with each other. Firstly, we compute the averages of Q and K within each region, resulting in Then, We begin by calculating the regional averages of both Q and K. Subsequently, we construct an adjacency matrix to elucidate the interrelationship between Q^r and K^r across various regions. by plotting a matrix between Q^r and K^r, we obtain the region-to-region affinity graph

$$A^r = Q^r(K^r)^T \tag{7}$$

The routing index matrix $I^r \in \mathbb{N}^{s^2\times k}$ stores the indices of the top K connections row-wise.The routing index matrix stores the top K connections row by row.

$$I^r = topkIndex(A^r) \tag{8}$$

By leveraging the regional routing index matrix, we are able to implement precise token-specific attention. This attention mechanism concentrates its focus on encompassing all key-value pairs spanning across the k-routed regional intersections. The tensors for the key K and value V are gathered based on the routing indices I^r, resulting in K^g and V^g, respectively. These gathered tensors, denoted as K^g and V^g, have a shape of $\mathbb{R}^{s^2\times\frac{kHW}{s^2}\times C}$, where S represents a scaling factor, k is the number of routing regions, H and W are the height and width dimensions, and C is the channel dimension. Subsequently, attention is applied to the gathered K-V pairs.

$$O = Attention(Q, K^g, V^g) \tag{9}$$

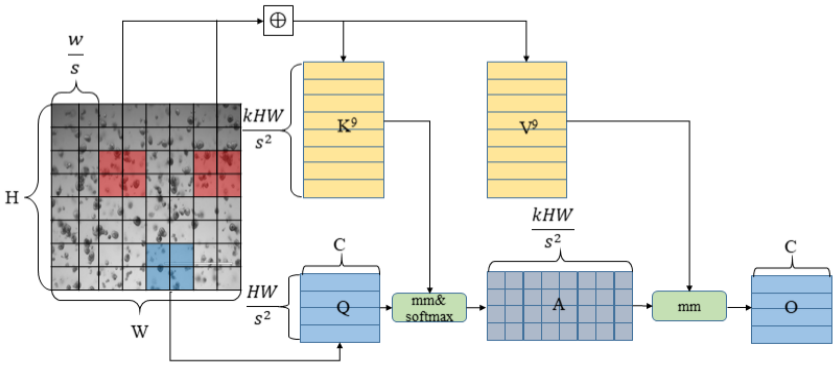

Fig. 5. H × W was divided into $\frac{HW}{s^2}$ chunks and their information was fused to compute rough regional information. For the blue region, two red regions contribute significant attention. Attention is then computed between the blue region and the red region by fusing the information from eight points into four pixels as the final information.

3.3 Decoupling classification and regression tasks

The target detection model efficiently handles classification and regression tasks via a decoupled head design. This divides the detection into two sub-tasks: classification and localization. By optimizing each sub-task separately, the model achieves improved accuracy and performance. For classification, the decoupled head extracts features, which are processed via normalization and global pooling to obtain category probabilities. The regression branch focuses on spatial location, scaling features through normalization, and deriving bounding box coordinates with relu activation. The formula is as follows:

$$P(y|x) = softmax(W_{cls} \cdot x + b_{cls}) \tag{10}$$

$$\Delta t = ReLU(W_{reg} \cdot x + b_{reg}) \tag{11}$$

$$M_1 = Conv_{3\times3}(MLConv_{1\times1}(W_{cls} * Conv_{3\times3}(F) + b_{cls})) \tag{12}$$

$$y\_cls = Softmax(GlobalAvgPool(BN(M))) \tag{13}$$

$$M_2 = Conv_{3\times3}(MLConv_{1\times1}(W_{reg} * Conv_{3\times3}(F) + b_{reg})) \tag{14}$$

$$y\_reg = ReLU(BN(M_2)) \tag{15}$$

Where, $P(y|x)$ is the probability distribution for each class given input x, and Δt is the offset of the bounding box coordinates. W_{cls} and W_{reg} represent the learnable parameters that constitute the classification and regression components respectively. The bias terms, b_{cls} and b_{reg}, act as offset values that aid in the respective processes. y_cls denotes the output of the classification branch, while y_reg signifies the output of the regression branch.

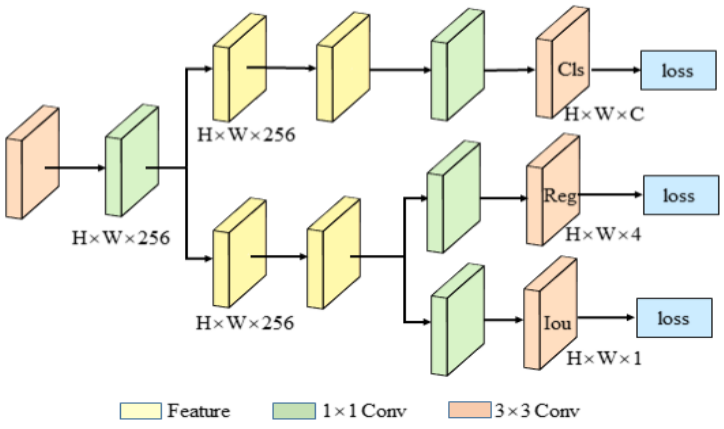

Fig. 6. Decoupling header structure

4 Experiment and Evaluation

In this part, we will compare the new method qualitatively and quantitatively with several baseline methods. Additionally, We performed ablation experiments to validate the efficacy of our proposed approach.

4.1 Implementation Details

We implemented our method in PyTorch and trained/tested it on the Tellu dataset [25], using an input size of 480x640 and a batch size of 16. The SGD algorithm with stochastic weighted averaging optimized the model. The initial learning rate was 0.01, decaying by 0.1 every 10 epochs. The training was conducted on an A100 GPU with 40GB memory for 100 epochs.

4.2 Qualitative and Quantitative Comparison

We compared our method with baselines like Tellu, YOLOv6 [34], YOLOv7 [35], YOLOv8 [36], YOLOv9 [37], and RT-DTER [38]. We quantitatively analyze the detection and classification capabilities of organoid bright field images using a series of evaluation metrics, as shown in Table 1 and Fig. 7. These metrics include mean average precision (mAP50), precision, recall, and F1 score.

Table 1. In assessing the quantitative performance of various organizations' baseline methods alongside our approach, we evaluated key metrics such as the Mean Average Precision at 50% (MAP50, with a higher value indicating better performance), Precision (where a higher value signifies improved accuracy), Recall (indicating a greater proportion of true positives captured, with a higher value preferred), and the F1-score (a harmonic mean of Precision and Recall, also favoring higher values) on intestinal organoid datasets.

Model	mAP50	Precision	Recall	F1-score
Tellu[25]	0.811	0.757	0.745	0.751
yolov6[34]	0.816	0.741	**0.781**	0.760
yolov7[35]	0.825	0.772	0.767	0.769
yolov8[36]	0.830	0.779	0.777	0.778
yolov9[37]	0.833	0.779	0.771	0.775
RT-DETR[38]	0.827	0.748	0.744	0.746
YOLORG	**0.841**	**0.789**	0.772	**0.780**

4.3 Ablation Study

To examine the impact of various components within YOLORG, we perform ablation studies on the dataset. This involves comparing the base YOLORG model with three distinct variations to evaluate their relative effectiveness.

– “w/o RE”: YOLORG without the res2net module.
– “w/o BR”: YOLORG without the BRA module
– “w/o HE”: YOLORG without the decoupling head module.
The quantitative results displayed in Table 2 convincingly illustrate that the implemented components significantly enhance the detection performance for intestinal organoids.

Table 2. The findings from our quantitative ablation studies across diverse tissue types underscore the significant contribution of our proposed components in enhancing organoid detection.

Model	mAP50	Precision	Recall	F1-score
w/o HE	0.822	0.763	0.759	0.761
w/o BR	0.817	0.755	0.749	0.752
w/o RE	0.815	0.751	**0.774**	0.762
YOLORG	**0.841**	**0.789**	0.772	**0.780**

5 Conclusion

In this paper, we introduce YOLORG, a multi-scale intestinal organoid detection framework. YOLORG leverages a robust backbone network to extract image features across various scales, which are further enhanced and optimized through the Bineck module. This approach ensures that the model can precisely capture and represent the crucial information in the input data. Moreover, by employing a decoupled prediction header, YOLORG disentangles the classification and localization tasks, enabling more flexible parameter tuning and weight adjustment to accommodate the unique characteristics of different intestinal organoids. This strategic decoupling ultimately leads to improvements in detection accuracy and efficiency. Extensive experiments validate the superiority of our proposed method over the baseline, establishing YOLORG as an automated, robust, and precise tool for intestinal organoid research.

6 Acknowledge

This work was supported in part by the Xinjiang Tianchi Talents Program (E33B9401), in part by the Natural Science Foundation of Xinjiang Uygur Autonomous Region under grant (2023D01E15) , in part by the Youth Program of Natural Science Foundation of Xinjiang Uygur Autonomous Region(Grant No.2022D01B67), and in part by the Tianshan Talent Training Program (2023TSYCLJ0021).

References

1. Zhao, Z., Chen, X., Dowbaj, A. M., Sljukic, A., Bratlie, K., Lin, L., … Yu, H. (2022). Organoids. Nature Reviews Methods Primers, 2(1).

2. Lancaster, M.A.; Knoblich, J.A. Organogenesis in a dish: Modeling development and disease using organoid technologies. Science 2014, 345, 1247125.
3. Fatehullah, A.; Tan, S.H.; Barker, N. Organoids as an In Vitro model of human development and disease. Nat. Cell Biol. 2016, 18, 246–254.
4. Rossi, G.; Manfrin, A.; Lutolf, M.P. Progress and potential in organoid research. Nat. Rev. Genet. 2018, 19, 671–687.
5. Hu P, Hu L, Wang F, et al. Computing and artificial intelligence in digital therapeutics[J]. Frontiers in Medicine, 2024, 10: 1330686.
6. Fei, K., Zhang, J., Yuan, J., & Xiao, P. (2022). Present Application and Perspectives of Organoid Imaging Technology. Bioengineering, 9(3), 121.
7. Bai, L., Wu, Y., Li, G., Zhang, W., Zhang, H., & Su, J. (n.d.). AI-enabled organoids: Construction, analysis, and application.
8. Kassis, T., Hernandez-Gordillo, V., Langer, R., & Griffith, L. G. (2019). OrgaQuant: Human Intestinal Organoid Localization and Quantification Using Deep Convolutional Neural Networks. Scientific Reports.
9. Fordham, R. P., Yui, S., Hannan, N. R. F., Soendergaard, C., Madgwick, A., Schweiger, P. J., … Jensen, K. B. (2013). Transplantation of Expanded Fetal Intestinal Progenitors Contributes to Colon Regeneration after Injury. Cell Stem Cell, 734–744.
10. Barroso, M., Monaghan, M. G., Niesner, R., & Dmitriev, R. I. . (2023). Probing organoid metabolism using fluorescence lifetime imaging microscopy (film): the next frontier of drug discovery and disease understanding. Advanced drug delivery reviews, 201.
11. Brandenberg N, Hoehnel S, Kuttler F, Homicsko K, Ceroni C, Ringel T, et al. High-throughput automated organoid culture via stem-cell aggregation in microcavity arrays. Nat Biomed Eng [Internet]. 2020 Sep [cited 2022 Apr 8];4(9):863–74. Available from:
12. Bian, X., Li, G., Wang, C., Liu, W., Lin, X., Chen, Z., … Luo, X. (2021). A deep learning model for detection and tracking in high-throughput images of organoids. Computers in Biology and Medicine, 134, 104490.
13. Jiang, S., Zhao, H., Zhang, W., Wang, J., Liu, Y., Cao, Y., … Ma, S. (2020). An Automated Organoid Platform with Inter-organoid Homogeneity and Inter-patient Heterogeneity. Cell Reports Medicine, 100161.
14. Tian, C., Yang, M., Xu, H., Zhu, M., Yue, N.-N., Zhang, Y., … Li, D. (2023). Stem cell-derived intestinal organoids: a novel modality for IBD. Cell Death Discovery, 9(1). https://doi.org/10.1038/s41420-023-01556-1
15. Beumer, J., & Clevers, H. (2021). Cell fate specification and differentiation in the adult mammalian intestine. Nature Reviews Molecular Cell Biology, 22(1), 39–53.
16. Li J, Han X, Qin Y, et al. Artificial intelligence accelerates multi-modal biomedical process: A survey[J]. Neurocomputing, 2023, 558: 126720.
17. Mergenthaler, P., Hariharan, S., Pemberton, J. M., Lourenco, C., Penn, L. Z., & Andrews, D. W. (2021). Rapid 3D phenotypic analysis of neurons and organoids using data-driven cell segmentation-free machine learning. PLOS Computational Biology, 17(2), e1008630.
18. Tong L, Li T, Zhang Q, et al. LiViT-Net: A U-Net-like, lightweight Transformer network for retinal vessel segmentation[J]. Computational and Structural Biotechnology Journal, 2024, 24: 213-224.
19. Ge, Z., Liu, S., Wang, F., Li, Z., & Sun, J. (2021). YOLOX: Exceeding YOLO Series in 2021.
20. Leng, B., Jiang, H., Wang, B., Wang, J., Luo, G., & O, A. (n.d.). Deep-Orga: An improved deep learning-based lightweight model for intestinal organoid detection.

21. Abdul, L., Xu, J., Sotra, A., Chaudary, A., Gao, J., Rajasekar, S., … Zhang, B. (2022). D-CryptO: Deep learning-based analysis of colon organoid morphology from brightfield images. Lab on a Chip, 4118–4128.
22. Deng, X., Hu, L., You, Z.-H., & Hu, P.-W. (n.d.). CAMPEOD: A Cross Attention-Based Multi-Scale Patch Embedding Organoid Detection Model.
23. X. Deng, L. Hu, Z. Jiang, L. Liu and P. -W. Hu, "A contactless automated dynamic monitoring method for organoid morphology on the time axis," *2023 IEEE International Conference on Data Mining Workshops (ICDMW)*, Shanghai, China, 2023, pp. 412-417.
24. P. Hu, X. Deng, F. Tan, and L. Hu, "Multi-Axis Attention with Convolution Parallel Block for Organoid Segmentation," in *IEEE/CAA Journal of Automatica Sinica*, vol. 11, no. 5, pp. 1295-1297, May 2024.
25. Domeǹech-Moreno, E., Brandt, A., Lemmetyinen, T., Wartiovaara, L., Mäkelä1, T., Ollila, S., & Mardis, E. (n.d.). Tellu – an object-detector algorithm for automatic classification of intestinal organoids.
26. Du, X., Cui, W., Song, J., Cheng, Y., Qi, Y., Zhang, Y., … Gu, Z. (2022). Sketch the Organoids from Birth to Death Development of an Intelligent OrgaTracker System for Multi-Dimensional Organoid Analysis and Recreation.
27. Kshirsagar, V., Bhalerao, R. H., & Chaturvedi, M. (2023). Modified YOLO Module for Efficient Object Tracking in a Video. IEEE Latin America Transactions, 21(3), 389–398.
28. Liu, C., Li, D., & Huang, P. (2021). ISE-YOLO: Improved Squeeze-and-Excitation Attention Module based YOLO for Blood Cells Detection. 2021 IEEE International Conference on Big Data (Big Data). Presented at the 2021 IEEE International Conference on Big Data (Big Data), Orlando, FL, USA.
29. Cui, C., Han, X., Wu, Y., Gao, D., Fang, Y., Xu, X., … Ju, Z. (n.d.). SDE-YOLO: A Novel Method for Blood Cell Detection.
30. Haq, I., Mazhar, T., Asif, R. N., Ghadi, Y. Y., Ullah, N., Khan, M. A., & Al-Rasheed, A. (2024). YOLO and residual network for colorectal cancer cell detection and counting. Heliyon, 10(2).
31. Liu, W., Ren, G., Yu, R., Guo, S., Zhu, J., & Zhang, L. (2022). Image-Adaptive YOLO for Object Detection in Adverse Weather Conditions. Proceedings of the AAAI Conference on Artificial Intelligence, 1792–1800.
32. Gao, S.-H., Cheng, M.-M., Zhao, K., Zhang, X.-Y., Yang, M.-H., & Torr, P. (2021). Res2Net: A New Multi-Scale Backbone Architecture. IEEE Transactions on Pattern Analysis and Machine Intelligence, 652–662.
33. Zhu, L., Wang, X., Ke, Z., Zhang, W., & Lau, R. (2023). BiFormer: Vision Transformer with Bi-Level Routing Attention.
34. Li, C., Li, L., Jiang, H., Weng, K., Geng, Y., Li, L., ... & Wei, X. (2022). YOLOv6: A single-stage object detection framework for industrial applications.
35. Wang, C.-Y., Bochkovskiy, A., & Liao, H.-Y. (n.d.). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors.2.
36. Lou, H., Duan, X., Guo, J., Liu, H., Gu, J., Bi, L., & Chen, H. (2023). DC-YOLOv8: small-size object detection algorithm based on the camera sensor. Electronics, 12(10), 2323.
37. Wang, C.-Y., Yeh, I.-H., & Liao, H.-Y. (n.d.). YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information.
38. Lv, W., Xu, S., Zhao, Y., Wang, G., Wei, J., Cui, C., ... & Liu, Y. (2023). Detrs beat yolos on real-time object detection.