Bio-Design and Manufacturing (2026) 9:796–808
https://doi.org/10.1631/bdm.2500438

RESEARCH ARTICLE

OrganoidViT: a vision transformer-based deep learning model for segmenting viable and nonviable organoids in bright-field images

Yijun Liu[1] · Dingyuan Yu[1] · Guoxiang Fu[2] · Zhangjie Li[1] · Chenyang Zhou[1] · Jiaqi Xu[1] · Ning Zhao[3] · Lian Xuan[4] · Xiaolin Wang[1,4,5]

Received: 27 August 2025 / Accepted: 11 March 2026 / Published online: 20 June 2026

Abstract

The viability assessment of patient-derived tumor organoids is essential for preclinical drug screening, with microscopic imaging serving as a key method for evaluating drug effects. Traditional image analysis methods, such as manual evaluation and fluorescence staining, suffer from low efficiency, dye toxicity, and fluorescence degradation, making them unsuitable for high-throughput drug screening. Additionally, current artificial intelligence (AI)-based tools face challenges in precise segmentation for the accurate quantitative analysis of viable and nonviable organoids. To address these challenges, we introduce OrganoidViT, a novel deep learning model utilizing vision transformer technology for the precise segmentation of viable and nonviable colorectal cancer organoids in bright-field microscopy images, ensuring an accurate efficacy assay of different drugs. Trained on custom datasets of colorectal cancer organoids prepared using microdroplet technology, OrganoidViT facilitates high-throughput segmentation without fluorescence imaging, with experimental results indicating superior performance of OrganoidViT (accuracy of 99.7%) compared to manual annotation and traditional convolutional models. The capability of pixel-level segmentation enables accurate morphological assessments, including area, pellucidity, roundness, and grayscale entropy, which are critical for evaluating growth conditions and the effects of different therapeutic agents on colorectal cancer organoids. Thus, OrganoidViT shows promise for preclinical drug screening, enhancing both the efficiency and accuracy of organoid-based testing.

Yijun Liu and Dingyuan Yu have contributed equally to this work.

✉ Xiaolin Wang
xlwang83@sjtu.edu.cn

[1] School of Integrated Circuits (School of Information Science and Electronic Engineering), Shanghai Jiao Tong University, Shanghai 200240, China

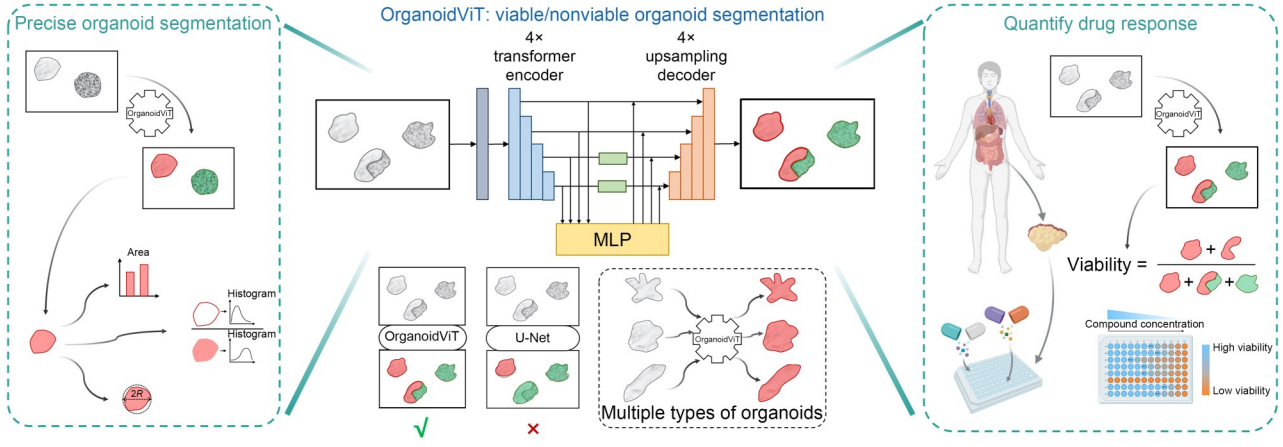
[2] Department of Research and Early Development, D1Med Technology (Hangzhou) Inc., Shanghai 200235, China

[3] Department of Orthodontics, Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine; College of Stomatology, Shanghai Jiao Tong University; National Center for Stomatology; National Clinical Research Center for Oral Diseases; Shanghai Key Laboratory of Stomatology, Shanghai 200011, China

[4] Institute of Medical Robotics, Shanghai Jiao Tong University, Shanghai 200240, China

[5] National Key Laboratory of Advanced Micro and Nano Manufacture Technology, Shanghai Jiao Tong University, Shanghai 200240, China

Graphical abstract



Keywords Organoids · Artificial intelligence (AI) · Deep learning · Vision transformer · Image segmentation

1 Introduction

Preclinical screening is a critical step in the development of therapeutic drugs for cancer. Traditional biomarker screening methods rely on two-dimensional (2D) cultured cell models or animal models. Although 2D cell models can be expanded in vitro, they considerably differ from human tissues and lose tumor heterogeneity after several passages [1]. Drug screening using patient-derived xenograft mouse models suffers from a low success rate, long testing cycles, and the impossibility of high-throughput screening [2, 3]. Patient-derived organoids, as a novel tool for cancer research, closely resemble the original tissues in terms of gene expression and drug response and can be stably passaged [4, 5]. Therefore, patient-derived organoids exhibit notable advantages in the development and screening of cancer drugs, improving success rates, shortening development cycles, and reducing costs [6–11].

When organoids are used to assist drug development and screening, statistical analyses of microscopic images of organoids under different drug treatments critically affect the accuracy and efficiency of drug screening [3, 11]. Traditional analysis methods, such as manual operations, are inefficient and prone to errors because of the large amount of data. At the same time, digital image processing methods based on fluorescent staining images suffer from poor accuracy because of dye toxicity and fluorescence decay [12, 13], and cannot continually identify nonviable organoids, as these methods limit the segmentation to endpoint assays, complicating the quantification of the long-term effects of medication [14]. Additionally, bubbles in the culture matrix can interfere with the analysis and recognition of organoid images. Therefore, an accurate and robust method that does not rely on fluorescence images must be developed to distinguish between viable and nonviable organoids, as well as quantify their drug response.

In recent years, new artificial intelligence (AI)-based methods, such as convolutional neural networks (CNNs), have been introduced for organoid recognition [12–24]. For example, Abdul et al. [23] proposed the D-CryptO tool that can automatically assess crypt formation and the pellucidity of colorectal organoids from bright-field images to determine the maturity of organoid structures. They also introduced the Deep-LUMEN tool to determine whether lung alveolar spheroids exhibited the correct morphology [24]. However, these methods depend on object detection models that generate only bounding boxes, capturing not only the target regions but also background and irrelevant areas. Consequently, these approaches exhibit limited potential for the quantitative analysis of organoids and drug response evaluation. Matthews et al. [14] proposed OrganoID, a powerful image analysis platform that can automatically identify, label, and track individual organoids pixel by pixel in bright-field microscopy images and quantify parameters such as area, roundness, and eccentricity for long-term time-lapse microscopy studies and analyses of organoid responses to chemotherapy. However, this platform cannot identify and quantify nonviable organoids, complicating the assessment of the effects of drugs on organoids.

To address these challenges, we developed a software platform named OrganoidViT, which employs a vision transformer (ViT) for organoid segmentation [25, 26]. This approach addresses the limitations of traditional pyramid-structured CNNs by preserving segmentation accuracy

while effectively distinguishing between viable and nonviable organoids [27]. ViT is a deep learning model built on the transformer architecture. Its core components consist of the "linear projection of flattened patches" module for serializing input image data and the "transformer encoder" module for capturing global information and extracting features [26, 28]. Unlike pyramidal tree structures, the transformer architecture minimizes information loss during feature extraction [28]. OrganoidViT includes an input module, a transformer encoder, and a prediction branch. It also integrates multiple loss functions, such as dice, focal, and compact losses, to improve the accuracy of organoid segmentation.

2 Results

2.1 Transformer model for the segmentation of colorectal cancer organoids

We developed a semantic segmentation model based on ViT, named OrganoidViT, which can segment organoids in bright-field microscopy images pixel by pixel and classify them as either viable or nonviable (Fig. 1). The model takes the bright-field images of organoids as input and outputs multichannel probability maps, where the pixel value in each channel represents the probability of the pixel belonging to viable or nonviable organoids. The model employs an encoder–decoder architecture, with the encoder derived from ViT and utilizing 2D positional attention to retain the 2D positional information between image tokens [29], thereby providing richer spatial details for segmenting small-sized organoids. The encoder is structured into four stages, with the transformer encoder in each stage consisting of two transformer layers and an "overlap patch merging" module for feature extraction and downsampling of the feature maps [21, 29]. As a result, the four stages output multiscale feature maps with resolutions of 1/4, 1/8, 1/16, and 1/32 of the original image, respectively. The decoder is derived from the U-Net architecture, utilizing a stepwise upsampling approach for the fine-grained segmentation of the feature maps. Notably, we utilized the dilated residual connection [30] module to establish connections between the decoder and the 3rd and 4th stages of the encoder, replacing the traditional skip connection in U-Net. This design preserves a richer global semantic context, enhancing the ability of the model to more accurately distinguish ambiguous organoids within the image. In contrast, the 1st and 2nd stages use direct skip connections to transmit features to the decoder, effectively reducing computational load while ensuring accuracy.

2.2 Dataset creation and model training

To facilitate the segmentation of viable and nonviable organoids using the OrganoidViT model, we prepared colorectal cancer organoids using droplet microfluidics and transferred them into 96-well cell culture plates for subsequent culture and treatment (Figs. 2a–2c). A Nikon Ti2-E microscope (Nikon, Japan) equipped with a 3× objective lens was employed to capture images of the colorectal cancer organoids

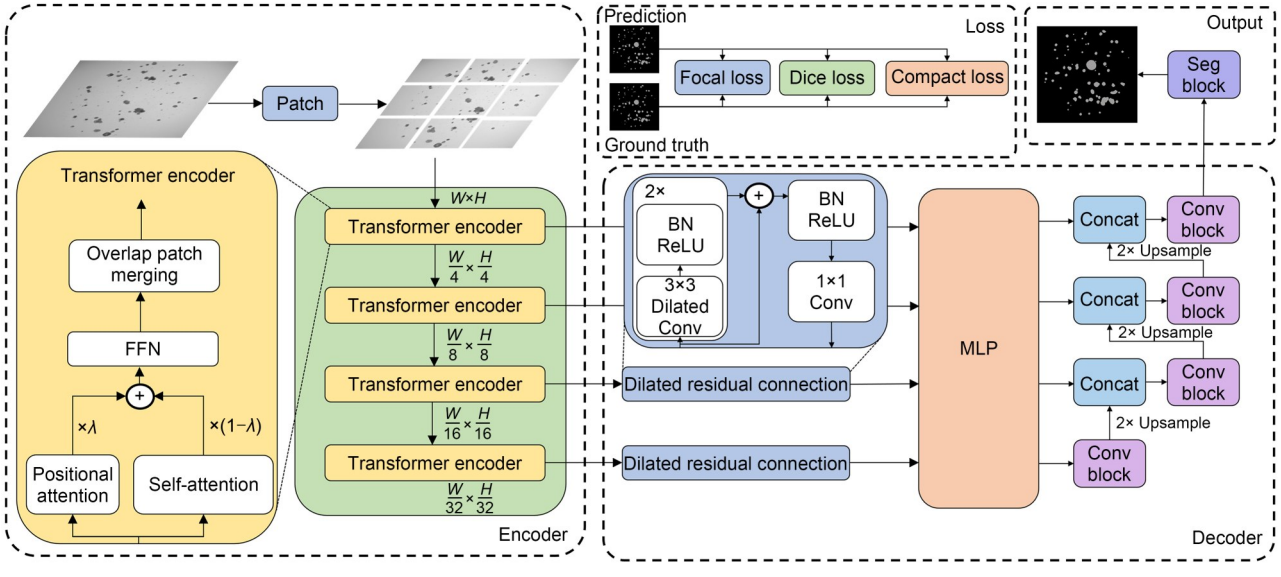


Fig. 1 Architecture of the OrganoidViT model. The model is based on the encoder–decoder structure. The input must be bright-field microscopy images of organoids, from which multiscale features are extracted in multistage transformer layers, yielding feature maps at scales of 1/4, 1/8, 1/16, and 1/32 of the input image size. These features are then connected to an MLP via a dilated residual connection module, yielding a mask image of organoids as the output. FFN: feedforward network; Conv: convolution; Concat: concatenate; BN: batch normalization; ReLU: rectified linear unit; MLP: multilayer perceptron

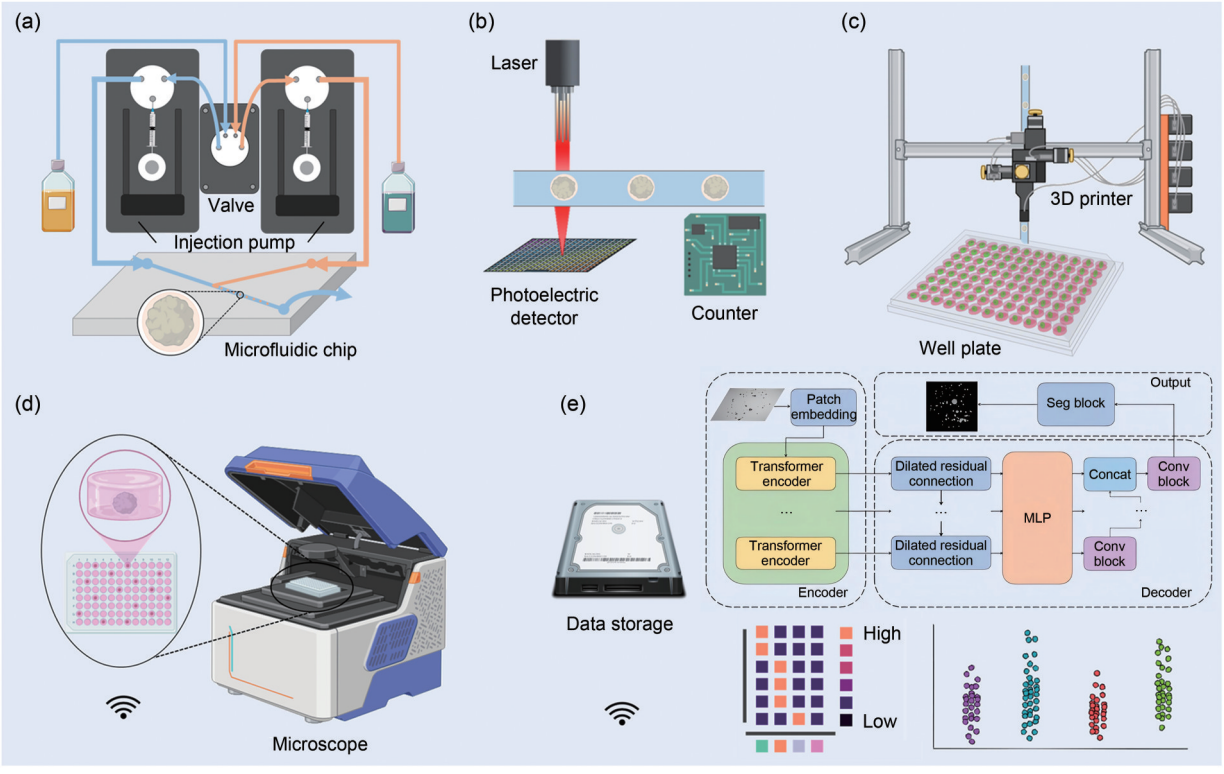


Fig. 2 Overview of the development of the OrganoidViT model. We utilized droplet microfluidics to standardize the cultivation and drug treatment procedures for colorectal cancer organoids, enabling automated analyses via our OrganoidViT model. (a–c) Standardized organoid cultivation: droplets encapsulating organoids and extracellular matrices were generated at high throughput using a syringe pump and microfluidic chips, and then dispensed into 96-well cell culture plates using a three-dimensional (3D) droplet printer for subsequent cultivation, treatment, and imaging. (d, e) Acquisition of organoid image data and analysis using OrganoidViT

(Fig. 2d), which were subsequently used for model training and inference (Fig. 2e). We acquired organoid images using the *Z*-stack method implemented in the OpenCV (version 4.5.1.48) framework, following a timeline (Fig. 3a). This method involves capturing multiple images at different focal planes and then synthesizing them to create a clear full-field view of the organoids on different days (Figs. 3b–3d). A total of 812 microscopy images of colorectal cancer organoids were collected, and each organoid was annotated using the Labelme software by an expert team based on morphology, pellucidity, and dispersion, as well as fluorescent images, classifying each organoid as viable or nonviable (Fig. 3e). Specifically, regions with high pellucidity or distinct textures were classified as viable organoids, whereas regions with low pellucidity or dispersed structure were labeled as nonviable organoids. Additionally, bubbles in the images were separately labeled to prevent misidentification by the model. The data were divided into training, validation, and test sets in a 60%, 20%, and 20% split and used to train the OrganoidViT model through transfer learning from the ImageNet dataset. The OrganoidViT model was implemented in PyTorch and trained for 40,000 steps. Training configuration parameters (Table S1 in the supplementary information) were adjusted and optimized, with training curves recorded in TensorBoard (version 2.10.1) (Fig. S1 in the supplementary information) to assess convergence and prevent overfitting.

The model performance was quantified using multiple metrics on the test set to comprehensively evaluate its predictive capability: intersection over union (IoU, for semantic region-level assessment), Dice coefficient (for semantic region-level assessment), and Recall, Precision, and F1-score (for organoid-level evaluation). All test images met the predefined benchmark criteria, with the model achieving a mean IoU of 0.745 (a mean IoU above 0.5 indicates good predictive capability [14]), a Dice coefficient of 0.854 for semantic segmentation, a Recall of 0.935, a Precision of 0.987, and an F1-score of 0.960. These results confirm the reliability of the training outcomes and validate the superior performance of OrganoidViT. To further assess the performance of OrganoidViT, we implemented a U-Net [27] model in PyTorch on the same dataset and compared the prediction results with those of OrganoidViT (Fig. 4). The viable organoids in the images taken in the initial stage of treatment (Days 0 and 6) can be effectively predicted by both OrganoidViT and U-Net, though the pixel counts predicted by OrganoidViT are closer to the ground truth, achieving an accuracy of 99.7%. In the later stages of treatment (Days 6, 9, and 12), U-Net struggled to reliably distinguish between viable and nonviable

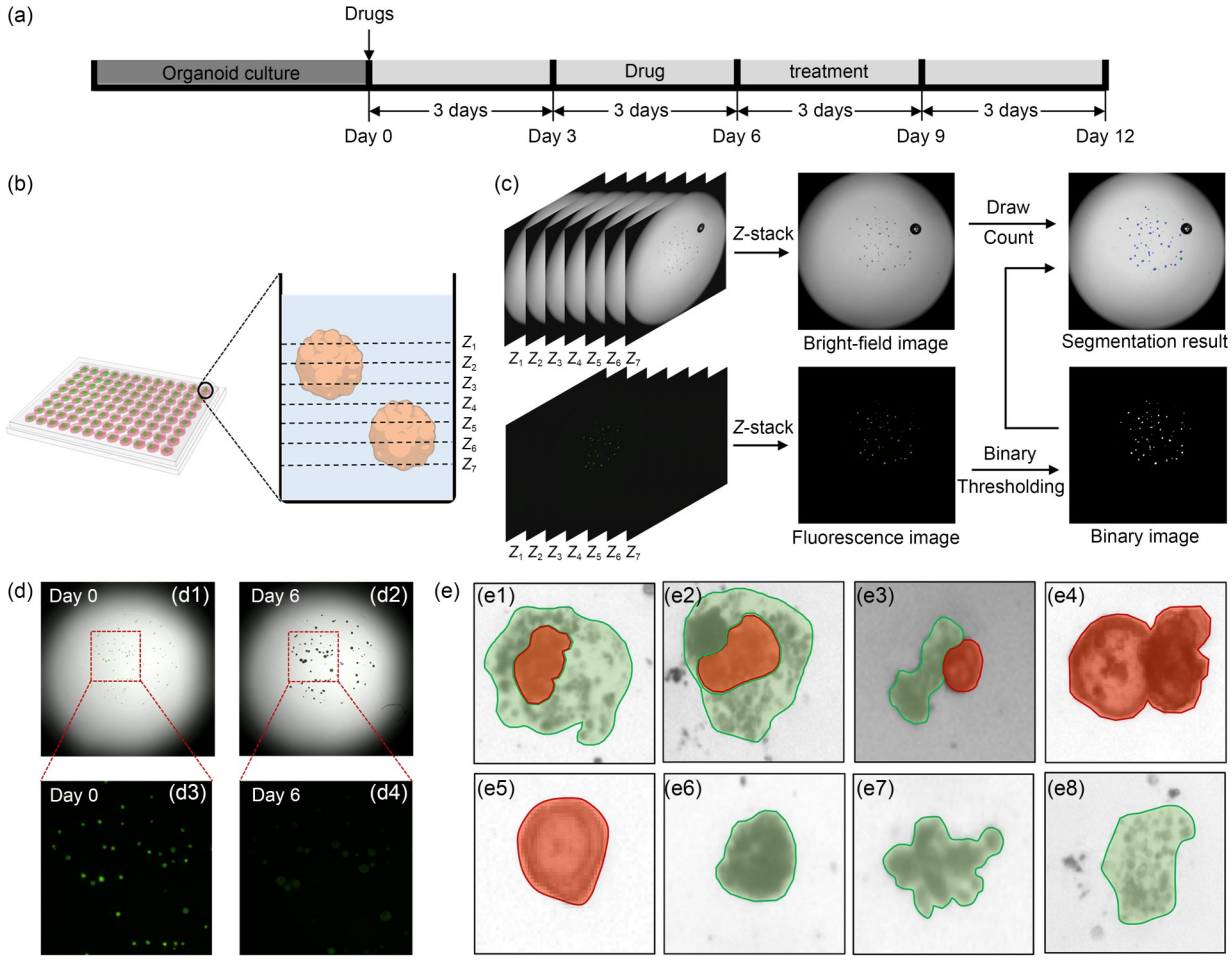


Fig. 3 Imaging and dataset annotation of organoids. (a) Schematic timeline of organoid image acquisition. (b) *Z*-stack method for organoid image acquisition, in which the final image is a composite of seven images taken at different focal planes. (c) A workflow diagram of organoid segmentation using digital image processing based on fluorescent staining. The segmentation results were used in the annotation of the organoid datasets. Bright-field and fluorescence images of organoids were collected through the *Z*-stack method, and segmentation was performed by applying a set threshold for binarization, filtering, and edge detection of the fluorescent images. (d) A schematic showing fluorescence decay over the organoid culture period. Panels (d1, d3) show the bright-field and corresponding fluorescence images within the red boxed area on Day 0. Panels (d2, d4) show the same for Day 6, highlighting a substantial reduction in fluorescence intensity. (e) An example of dataset annotations, where red areas indicate viable organoids and green areas indicate nonviable organoids. Panels (e1–e3) show the annotation approach for organoids containing both viable and nonviable regions, panels (e4, e5) show viable regions, and panels (e6–e8) show nonviable regions

organoids, leading to substantial pixel count discrepancies. Conversely, OrganoidViT exhibited occasional misclassifications, but its relative error in predicting viable and nonviable organoids on Day 12 was 2.14% and 1.73%, respectively, which are acceptable.

To further validate the accuracy of OrganoidViT, we conducted statistical analyses of the relative errors of predictions from U-Net, U-Net++, DeepLabv3, and OrganoidViT against the ground truth on the test set. In predicting viable and nonviable organoids, OrganoidViT shows a substantially smaller relative error than the other three models when benchmarked against ground truth data (Figs. 5a and 5b). However, in bubble prediction, OrganoidViT exhibits a higher relative error than the other three models (Fig. 5c) because of its ability to detect edge bubbles (Fig. S2 in the supplementary information), which were not included in the training dataset. These results indicate that OrganoidViT outperforms traditional CNN-based networks in organoid segmentation tasks and further confirm its ability to learn more complex and deeper features from images.

We also trained the OrganoidViT model on additional publicly available organoid datasets [14, 22], achieving the segmentation of human small intestinal organoids, human pancreatic ductal adenocarcinoma organoids, and mouse small intestinal organoids. As these public datasets lack annotations for nonviable organoids, we focused exclusively on viable organoids (Fig. 6). OrganoidViT achieved mean IoU scores of 0.9182, 0.9320, and 0.9618 on the three datasets. Despite the lack of annotations for nonviable organoids, OrganoidViT exhibits strong performance compared to the

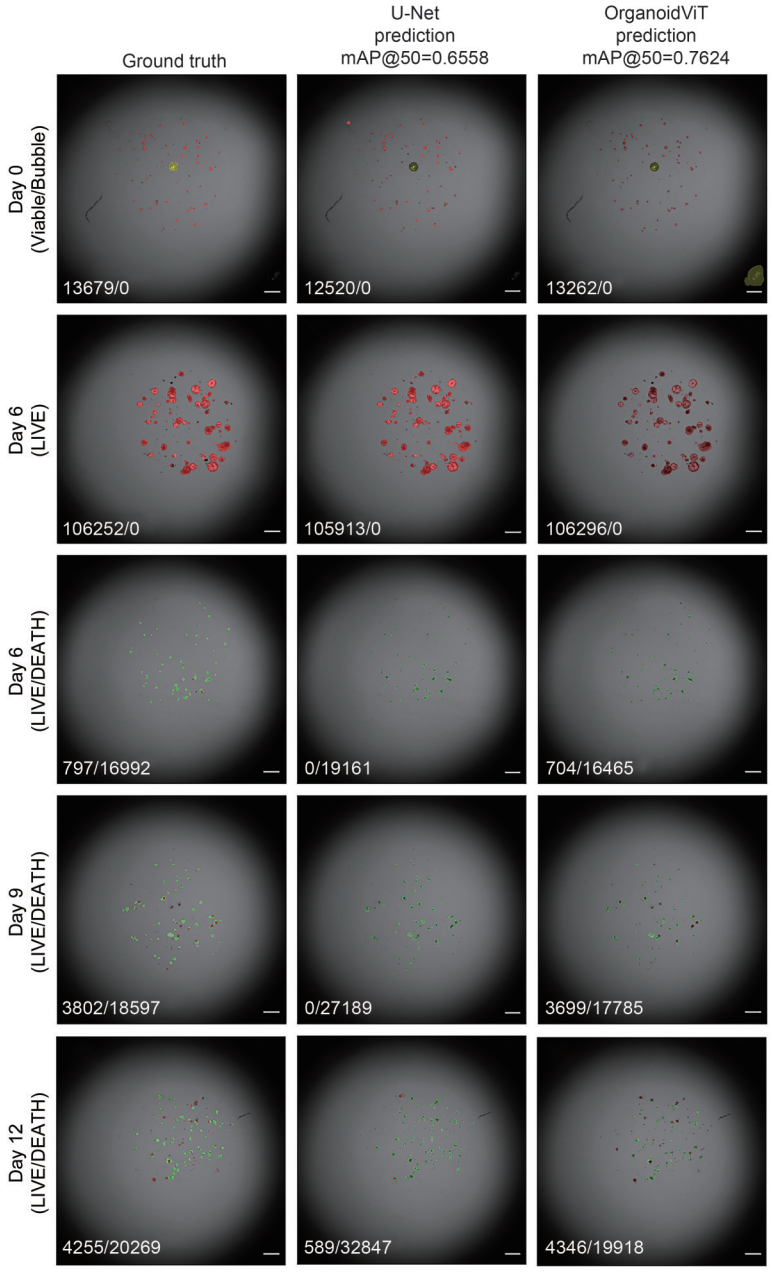


Fig. 4 Comparison between the ground truth and prediction results from different models on various days. Two numbers in the lower left corner of each image indicate the areas of viable/nonviable organoids (pixels). In all images, red-marked regions represent viable organoids, green-marked regions represent nonviable organoids, and yellow-marked regions represent bubbles. Scale bars: 500 μm. mAP@50: mean average precision at intersection over union (IoU) threshold of 0.50

predictions of viable organoids on the publicly available datasets (Table 1), highlighting its potential for the analysis of other types of organoids.

2.3 OrganoidViT uncovers changes in 2D morphological parameters during organoid expansion

The 2D morphological parameters of organoids—such as area, roundness, and pellucidity (Fig. 7a)—progressively change as the organoids mature [12]. These parameters provide a more precise assessment of organoid growth and quality. The pixel-level recognition capability of OrganoidViT enables the precise identification and quantification of organoid morphology, offering better accuracy than D-CryptO in morphological assessments [23]. We analyzed growth patterns in 37 groups of colorectal cancer organoids cultured under drug-free conditions. Images were obtained from 20 groups on Days 0 and 6 and from 17 groups on Days 0, 3, 6, 9, and 12. These images were processed using OrganoidViT

to generate masks of viable organoids, followed by the application of our custom extractor of morphological parameters to obtain the organoid count for each set and assess individual organoid area (Fig. 7b), pellucidity (Fig. 7c), roundness (Fig. 7d), and grayscale entropy (Fig. 7e).

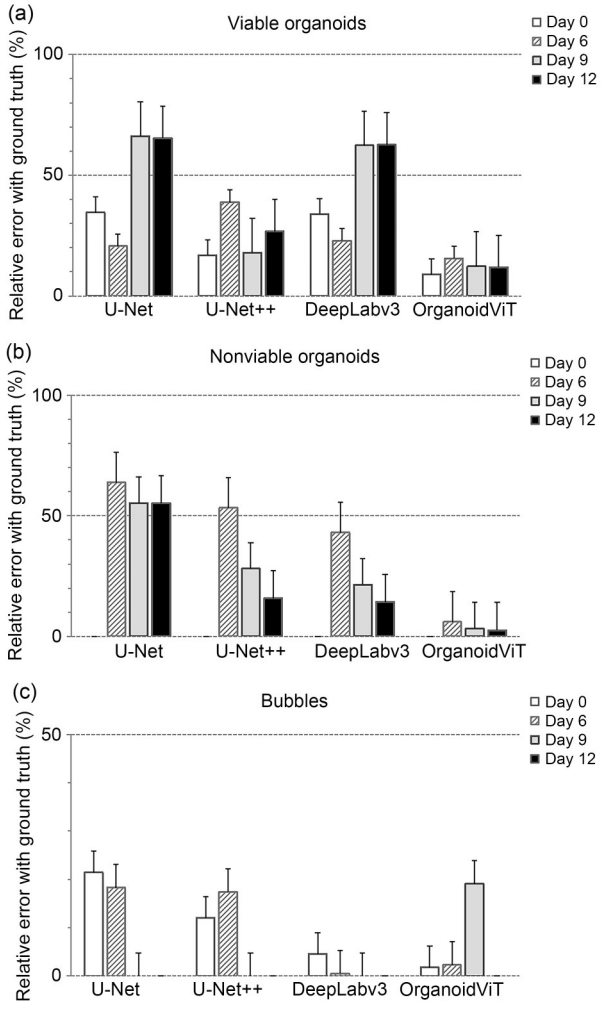


Fig. 5 Comparison of the prediction results obtained from U-Net, U-Net++, DeepLabv3, and OrganoidViT against the ground truth on the test set. (a) Relative errors between the predictions of all four models for viable organoids on different days and the ground truth, with OrganoidViT showing smaller relative errors. (b) Relative errors between the predictions of all four models for nonviable organoids on different days and the ground truth, with OrganoidViT showing smaller relative errors. It should be noted that no data for nonviable organoids are available for Day 0, resulting in the absence of relative error values for this day. (c) Relative errors between the predictions of all four models for bubbles compared to the ground truth. As some edge bubbles were not annotated in the ground truth, and OrganoidViT was able to partially predict these edge bubbles, its relative error was greater in this case. Data are expressed as mean±standard deviation (n=37 organoid groups)

The area was calculated as the total number of pixels predicted by OrganoidViT, representing the projection of the 3D organoid onto the X–Y plane and reflecting its overall volume. Pellucidity was determined by the contrast between the histograms of the edge and central regions within the OrganoidViT-predicted area, indicating the size of the organoid along the Z-axis. Roundness was measured as the ratio of the area to the square of the perimeter, where the perimeter is defined as the number of edge pixels in the predicted region. This metric reflects the degree of budding and maturation of an organoid. Grayscale entropy, calculated using Shannon's entropy formula, quantifies the randomness of grayscale values in the OrganoidViT-predicted region, assessing the complexity of the internal structure of the organoid. Consequently, as the culture time increases and the organoid matures, the area expands, the pellucidity decreases, the roundness diminishes, and the grayscale entropy rises. Statistical analyses from Day 0 to Day 12 reveal that the average area of 37 organoid groups expanded by 638.60% (from 193.30 to 1427.71), the average pellucidity decreased by 21.05% (from 0.0095 to 0.0075), the average roundness decreased by 13.94% (from 0.775 to 0.667), and the average grayscale entropy increased by 11.54% (from 5.46 to 6.09). Although organoid counts vary considerably across the experimental group, the trend in 2D morphological parameters within each experimental group remains consistent (Fig. S3 in the supplementary information). Additionally, as OrganoidViT cannot distinguish spatially overlapping organoids, the calculated roundness may slightly deviate from the actual values. However, statistical analyses of a substantial number of standard samples indicate that OrganoidViT is suitable for analyzing colorectal cancer organoid cultures.

2.4 OrganoidViT assists in quantifying the drug response of colorectal cancer organoids

The pixel-level recognition capability of OrganoidViT enables precise differentiation between viable and nonviable organoids, establishing it as a powerful tool for the quantification of drug effects on tumor organoids and suggesting its promising potential in drug screening applications [10]. Herein, eight distinct inhibitor drugs were administered to colorectal cancer organoids cultured in 96-well cell culture

Table 1 Comparison of intersection over union (IoU) results obtained through different methods

Type of organoids	Data source/Mean IoU of the corresponding method	Mean IoU of OrganoidViT
Human intestinal	OrgaSegment/0.76	0.92
Human pancreatic ductal adenocarcinoma	OrganoID/0.74	0.93
Mouse small intestinal	OrganoID/0.70	0.96

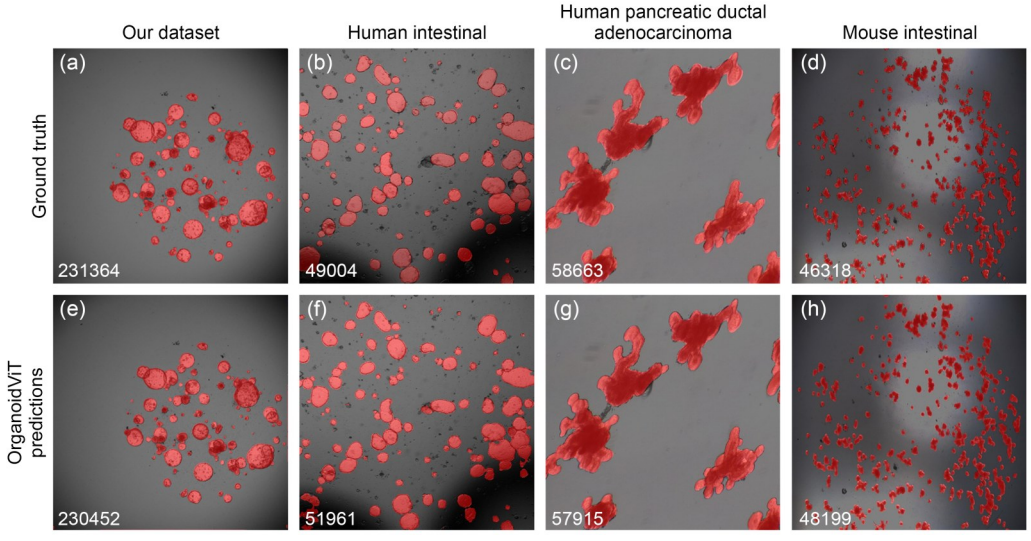


Fig. 6 Performance of OrganoidViT on additional public datasets. (a, e) Ground truth and predictions on our custom colorectal cancer organoid dataset. (b, f) Ground truth and predictions on human small intestinal organoids from the OrgaSegment dataset [22]. (c, g) Ground truth and predictions on human pancreatic ductal adenocarcinoma organoids from the OrganoID dataset [14]. (d, h) Ground truth and predictions on mouse small intestinal organoids from the OrganoID dataset [14]. The number in the lower left corner of each image indicates the area of viable organoids (pixels)

plates. The drugs were applied at various concentrations according to a specified timeline (Fig. 8a), with untreated controls (0 μmol/L concentration) included in each experimental setup. Specifically, five inhibitors, including Adavosertib, Encorafenib, Lapatinib, Regorafenib, and Sotorasib, were administered for 6 d, with images captured on Days 0 and 6. The remaining three inhibitors—5-Fluorouracil, CPT-11, and Oxaliplatin—were administered for 12 d, with imaging performed on Days 0, 3, 6, 9, and 12.

Using OrganoidViT, we analyzed microscopic images of colorectal cancer organoids treated with these inhibitors at different concentrations, quantifying changes in the distribution of organoid area (Figs. 8b–8d; Fig. S4 in the supplementary information) and organoid viability (Figs. 8e–8g; Fig. S5 in the supplementary information). The viability of organoids was quantified as the ratio of the viable organoid area to the total organoid area. Our previous work showed that this indicator may reliably reflect the response of locally advanced rectal cancer to chemoradiation treatment. The organoid sizing method was validated via direct comparison with the CellTiter-Glo 3D cell viability assay [31]. Statistical analyses indicated that, relative to the control group, colorectal cancer organoids treated with different concentrations of Sotorasib exhibited no significant differences in area or viability, suggesting limited clinical applicability of Sotorasib for this case of colorectal cancer. Additionally, prior studies have shown that Sotorasib is ineffective against specific mutations of colorectal cancer [32, 33]. Conversely, the organoid growth rate decreases with increasing concentrations of Adavosertib, Encorafenib, Lapatinib, Regorafenib, 5-Fluorouracil, CPT-11, and Oxaliplatin (Fig. S4 in the supplementary information), suggesting their potential effectiveness in growth inhibition. Furthermore, the organoid viability declines with the extension of treatment duration and increase in drug concentrations (Figs. 8e–8g; Fig. S5 in the supplementary information), indicating a possible cytotoxic effect on colorectal cancer organoids [32–34].

Additionally, a comprehensive analysis of changes in organoid viability across various concentrations and treatment durations facilitates the determination of the optimal therapeutic concentration and exposure time for each inhibitor in colorectal cancer organoids. These findings indicate that the pixel-level segmentation capabilities of OrganoidViT enable the precise quantification of organoid area and viability, effectively capturing the therapeutic effects of drugs on colorectal cancer organoids, thereby providing valuable insights for preclinical drug screening.

3 Materials and methods

3.1 Culturing of colorectal cancer organoids

Human colorectal cancer organoids were obtained from D1Med Technology (Hangzhou) Inc. (China) and cultured as described by Yao et al. [31]. Briefly, organoids were cultured in an advanced Dulbecco's modified Eagle's medium (DMEM)/F12 medium (12634-010, Gibco, USA), containing 500 ng/mL R-spondin-1 (11083-HNAS, Sino Biological, China), 100 ng/mL Noggin (50688-M02H, Sino Biological), 50 ng/mL epidermal growth factor (EGF; 50482-MNCH, Sino Biological), 1× 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES; 15630080, Gibco), 1× GlutaMAX (35050061, Gibco), 100 μg/mL Normocin

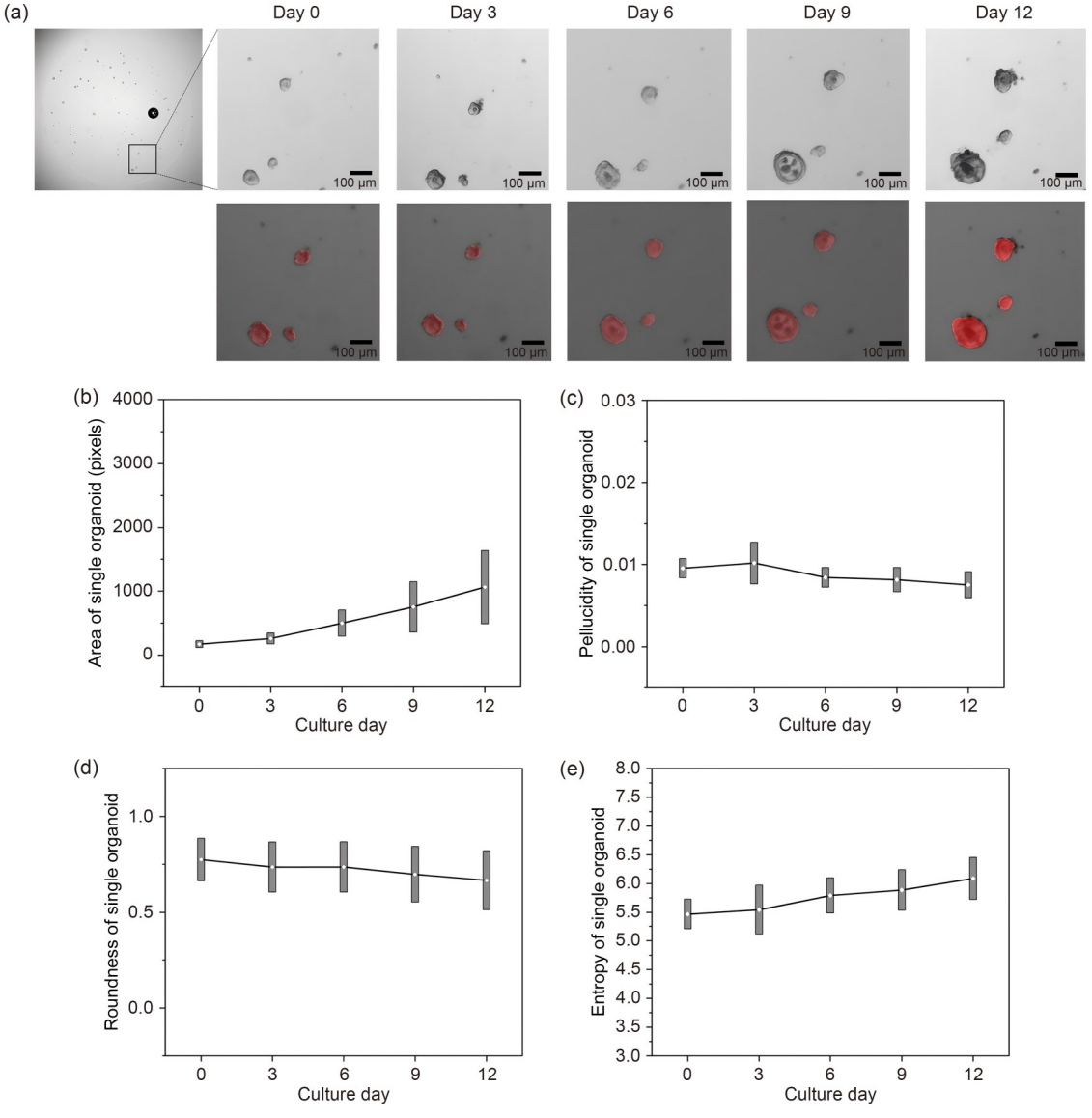


Fig. 7 Analyses of 2D morphological parameters of organoid growth using OrganoidViT. (a) Schematic of organoid growth dynamics, with original images shown at the top and OrganoidViT predictions shown at the bottom. (b) Area distribution and mean area trend of individual organoids across 37 organoid groups during cultivation. (c) Pellucidity distribution and mean pellucidity trend for individual organoids across 37 organoid groups during cultivation. (d) Roundness distribution and mean roundness trend for individual organoids across 37 organoid groups during cultivation. (e) Grayscale entropy distribution and mean grayscale entropy trend for individual organoids across 37 organoid groups during cultivation. The white diamond symbol on the line indicates the mean value. Data are expressed as mean±standard deviation (n=37 organoid groups)

(ant-nr-1, InvivoGen, France), 1× gentamicin/amphoteritin B (R01510, Gibco), 1× B27 (17504-044, Invitrogen, USA), 1× N2 (17502-048, Invitrogen), 1 mmol/L *N*-acetylcysteine (A9165, Sigma-Aldrich, Germany), 10 mmol/L nicotinamide (N0636, Sigma-Aldrich), 0.5 μmol/L A-83-01 (2939, Tocris, UK), 10 μmol/L Y-27632 (Y0503, Sigma-Aldrich), 10 nmol/L gastrin (G9145, Sigma-Aldrich), and 1 μmol/L prostaglandin E2 (P6532, Sigma-Aldrich). The organoid culture medium was refreshed every 3 d. Often, organoids were passaged every 1–2 weeks [31, 35].

3.2 OrganoidViT model implementation and training

PyTorch (version 1.8.1) and Python (version 3.7.6) were used to implement and train the OrganoidViT model. The input image dimensions for the model were set to 512×512 pixels, with a batch size of 4. A pretrained version of the ViT model, specifically trained on the ImageNet dataset, was loaded. The use of the pretrained model benefited OrganoidViT by providing a vast learned base of visual features, thereby reducing the need for labeled data specific

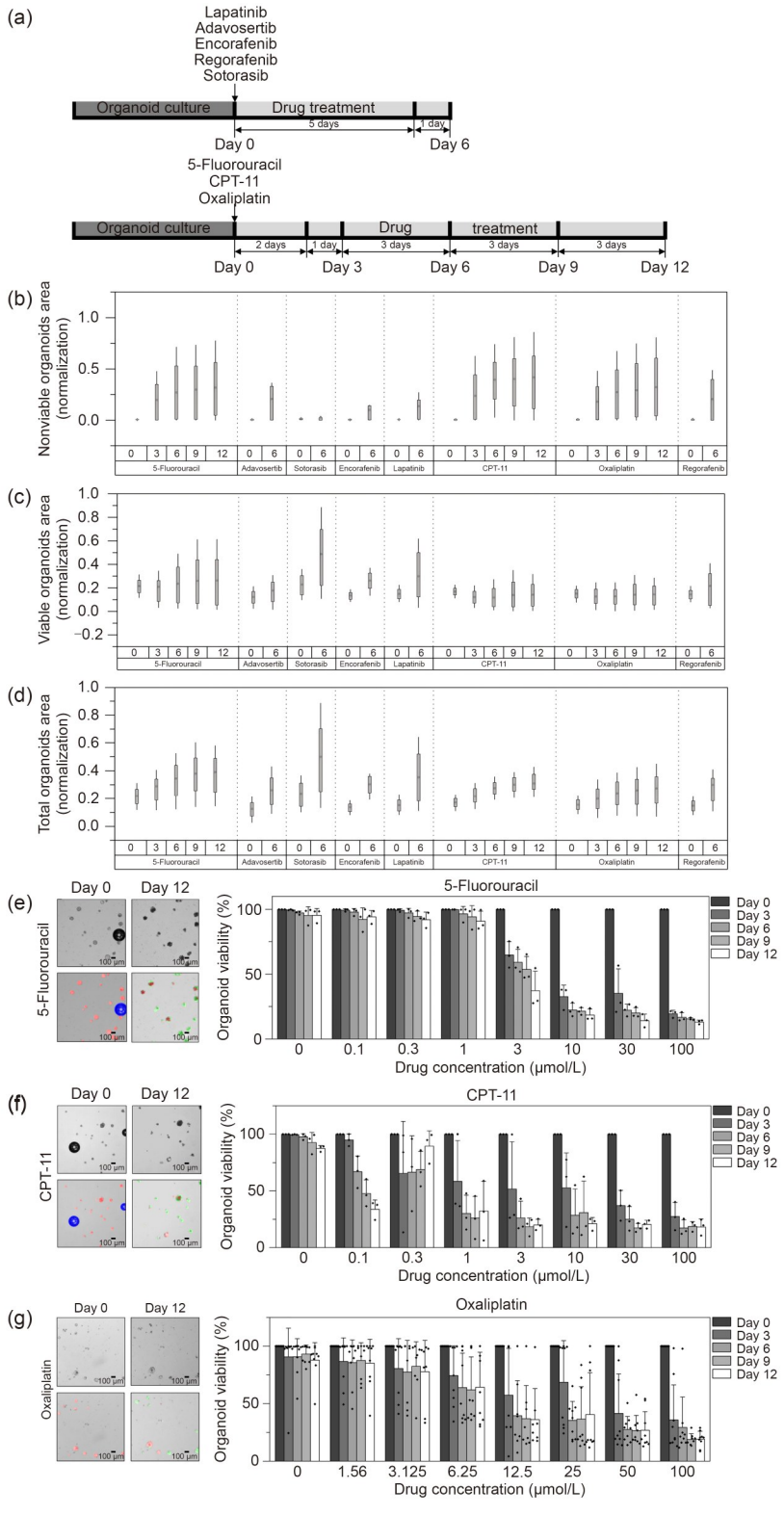


Fig. 8 Quantitative analyses of the effects of various inhibitors on colorectal cancer organoids using OrganoidViT, with some results presented in Fig. S5 in the supplementary information. (a) Timeline of the sequence of inhibitor administration and image acquisition for colorectal cancer organoids. (b–d) Trends in the nonviable, viable, and total average area of colorectal cancer organoids treated with different inhibitors over the entire culture period. The gray center symbol on the plot indicates the mean value. (e–g) On the left side, from top to bottom and left to right: an enlarged bright-field image on Day 0, the OrganoidViT detection result on Day 0, an enlarged bright-field image on the final day of culture, and the OrganoidViT detection results on the final day. On the right side: quantified average organoid viability, using OrganoidViT analyses, across the culture period at various drug concentrations. Data are expressed as mean±standard deviation (n=421 organoids)

to organoids and speeding up convergence during fine-tuning. The architecture of the model was carefully configured to handle variations in organoid morphology, which are crucial for tasks such as viability prediction and area measurement. Adam optimizer was used for model training, with an initial learning rate of 5×10^{-6} for 40,000 steps. The training set was augmented, which included random resizing within a range of scales, horizontal flipping, and random rotation. These augmentations aimed to increase the robustness of the model to variations in orientation and scale, which are typical in organoid imaging. All training was conducted on a server equipped with an NVIDIA 2080Ti graphics processing unit (GPU).

3.3 Extractor of morphological parameters

The morphological parameter extractor was developed using Python and OpenCV to extract specific characteristics (area, pellucidity, roundness, and grayscale entropy) from individual organoids identified in OrganoidViT segmentation results. The morphological parameter extractor takes the organoid mask output by OrganoidViT as input. It first applies a contour detection algorithm to generate a label map of all connected regions, where each pixel value corresponds to the label of a specific connected region, and identical pixel values indicate membership in the same region. The number of organoids and the area of each organoid are calculated by iterating through each connected region. During iteration, individual masks for each organoid are created to facilitate the computation of grayscale histograms. These histograms represent the distribution of grayscale pixel values within each organoid. Therefore, the area is measured by counting the number of pixels within each organoid, pellucidity is calculated as the standard deviation of the grayscale histogram within the segmented region, roundness is determined using Eq. (1) with A_{organoid} representing the organoid area and P_{organoid} representing the organoid perimeter, and grayscale entropy is computed using Eq. (2), where p_i is the probability of the occurrence of a pixel with grayscale value i in an image, and L is the number of gray levels.

$$\text{Roundness} = \frac{4\pi \times A_{\text{organoid}}}{P^2_{\text{organoid}}}, \tag{1}$$

$$\text{Entropy} = -\sum_{i=0}^{L-1} p_i \cdot \log_2 p_i. \tag{2}$$

3.4 Drug administration

Colorectal cancer organoids in good condition were harvested and seeded in 96-well cell culture plates (3548, Costar, USA). The density of organoids was adjusted to 5 μL^{-1} Matrigel (356231, Corning, USA) before seeding. Finally, every well contained about (50±20) organoids in 10 μL Matrigel with 300 μL culture medium.

Organoids derived from different patients exhibit varying growth rates and morphologies, making it impossible to achieve uniformity in characteristics such as cell lines. In our previous work, we established an assay for the imaging and analysis of 3D organoids [3]. We optimized the organoid size, density, and *Z*-stack layers for drug assay based on previous results, which showed that organoids had good detection performance at 60–100 μm. Herein, we employed a diameter of 50 μm to define the optimal time point for in vivo treatments. Specifically, when the organoids grew to about 50 μm, we performed drug treatment and replated the organoids, with drug testing started on the second day—by this time, most organoids had grown to the 60–100 μm range, exhibiting favorable detection performance.

During drug treatment, the organoid culture medium was removed and replaced with 300 μL drug-containing culture medium with 5-Fluorouracil (S1209, Selleck, USA), CPT-11 (S2217, Selleck), Oxaliplatin (S1224, Selleck), Adavosertib (S1525, Selleck), Encorafenib (S7108, Selleck), Lapatinib (S2111, Selleck), Regorafenib (S1178, Selleck), or Sotorasib (S8830, Selleck). Drug-containing culture medium was refreshed again after 3 d. For 5-Fluorouracil, CPT-11, and Oxaliplatin, the culture medium was refreshed on Days 6 and 9. Organoids were photographed during drug treatment (Ti2-E, Nikon).

The control group refers to the nondrug group, which was treated with the same volume of dimethyl sulfoxide as that used for dissolving the drug. This group is essential for ensuring the accuracy of the experiment. During the experiment, it is necessary to observe that the organoids maintain a normal growth state. If a large number of organoids in the blank control group die, the drug treatment experiment conducted on this batch of organoids should be judged unqualified.

4 Discussion

With the advancement of AI technologies, ViT has been applied to visual recognition tasks [21]. Owing to self-attention mechanisms, ViT can capture relationships within images, enhancing the capacity of the model to interpret visual information and providing superior representational power compared to traditional CNNs [24]. We employed ViT technology for the recognition of organoids in bright-field microscopy images, introducing the OrganoidViT tool, which can automatically identify viable and nonviable colorectal cancer organoids at the pixel level and analyze their morphological parameters. The effectiveness of this tool was further validated by analyzing the images of long-term-cultured colorectal cancer organoids exposed to various inhibitor drugs.

OrganoidViT offers several advantages over previous organoid analysis tools [14, 16–20, 22–24]. First, by learning deeper image features, OrganoidViT can distinguish between

viable and nonviable organoid regions without the need for fluorescent staining, providing quantitative data for evaluating drug effects on tumor organoids. Second, OrganoidViT achieves superior segmentation accuracy, offering notable advantages in the extraction of morphological parameters from organoids, facilitating a more accurate automated organoid analysis. Lastly, as it does not require fluorescent staining, OrganoidViT can be used for continuous monitoring throughout culturing. When combined with the OrganoID organoid tracking algorithm, it can be employed for monitoring the entire developmental and drug response processes of individual organoids.

Despite its many advantages, OrganoidViT has some limitations. First, because of limited training data, we have only evaluated its performance on colorectal cancer data. This limitation could be addressed by expanding the dataset to a greater number of organoid types, although culturing, drug testing, image collection, and annotation remain challenging tasks. Second, it misclassified some nonviable organoids because postmortem organoids exhibit diverse morphologies, and early-stage clustered nonviable colorectal cancer organoids may visually resemble highly mature viable organoids. To enhance the feature extraction capabilities of the model, further fine-tuning of the transformer encoder with more diverse data is required. Finally, OrganoidViT struggles with identifying blurry or overlapping organoids, necessitating model refinement, increased data input, and additional postprocessing.

Notably, OrganoidViT still has great potential for optimization. Its detection ability and applicability can be improved by optimizing datasets and algorithms. Although OrganoidViT is currently capable of analyzing viable and nonviable colorectal cancer organoids at the 2D level by detecting their areas, the semantic segmentation of 3D organoids remains to be addressed, particularly in solving issues such as overlapping, occlusion, and fusion. Additionally, the current output of OrganoidViT is limited to the area parameters of organoids, rather than core indicators such as cell density that more directly reflect the viability of organoids. Therefore, future research can employ high-quality 3D organoid datasets with cell density annotations for model training to further enhance the performance of the model.

5 Conclusions

Herein, we introduced OrganoidViT, a ViT-based automated tool for the analysis of organoid images. It can automatically identify viable and nonviable colorectal cancer organoids at the pixel level and analyze their morphological parameters, enabling the assessment and quantification of organoid growth and drug responses. To validate OrganoidViT, we first used it to analyze the growth of colorectal cancer organoids. By identifying organoids at the pixel level, OrganoidViT extracted parameters such as area, roundness, pellucidity, and grayscale entropy, providing a detailed representation of the morphological changes occurring during organoid growth. Subsequently, we employed OrganoidViT to assess the effects of eight inhibitor drugs on colorectal cancer organoids at various concentrations. By performing the pixel-level identification of viable and nonviable regions, we calculated organoid viability, thus assessing the therapeutic effects of the tested inhibitor drugs across different concentrations. Thus, OrganoidViT offers valuable insights for organoid cultivation and drug screening.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1631/bdm.2500438.

Acknowledgements This work was supported by grants from the National Key R&D Program of China (No. 2024YFB4707800), the Interdisciplinary Program of Shanghai Jiao Tong University (No. YG2023LC04), the National Natural Science Foundation of China (Nos. 32471473, 62231025, and 82171011), the Research Program of Shanghai Science and Technology Committee (Nos. 24141900900 and 25JC3201100), and Chongqing Natural Science Foundation (No. CSTB2022NSCQ-MSX0767). The authors are also grateful to the Center for Advanced Electronic Materials and Devices (AEMD) of Shanghai Jiao Tong University.

Author contributions YJL and DYY designed and trained the OrganoidViT model, performed the experiments and data analyses, and prepared the manuscript. GXF was responsible for colorectal cancer organoid cultivation, and helped with dataset creation. ZJL, CYZ, and JQX conducted the experiments related to droplet microfluidics. NZ and LX collected the data and edited the manuscript. XLW conceived and administrated the research, and contributed to manuscript editing. All authors reviewed the final manuscript.

Declarations

Conflict of interest GXF is an employee and XLW is a technical consultant of D1Med Technology (Hangzhou) Inc., a company involved in the application of organoid technology. The remaining authors declare no competing interests with relevance to this study.

Ethical approval The human colorectal cancer organoids used in this study were purchased from D1Med Technology (Hangzhou) Inc. The supplier confirmed that these organoids were derived from human tissue that was obtained with informed consent from the donor, and all related ethical regulations and approvals were complied with (ethics approval number: 2103232-7).

Data availability The datasets can be acquired by requesting from the corresponding author.

Code availability The code can be downloaded from https://github.com/Notalove/Orgvit.

Use of generative AI tools No generative AI tools were used in the preparation of this manuscript.

References

1. Suggitt M, Bibby MC (2005) 50 years of preclinical anticancer drug screening: empirical to target-driven approaches. *Clin Cancer Res* 11(3):971–981.

https://doi.org/10.1158/1078-0432.971.11.3
2. Golebiewska A, Fields RC (2023) Advancing preclinical cancer models to assess clinically relevant outcomes. *BMC Cancer* 23(1): 230. https://doi.org/10.1186/s12885-023-10715-7
3. Li XM, Fu GX, Zhang L et al (2022) Assay establishment and validation of a high-throughput organoid-based drug screening platform. *Stem Cell Res Ther* 13(1):219. https://doi.org/10.1186/s13287-022-02902-3
4. Li M, Izpisua Belmonte JC (2019) Organoids: preclinical models of human disease. *N Engl J Med* 380(6):569–579. https://doi.org/10.1056/nejmra1806175
5. Yang HY, Sun LJ, Liu MX et al (2018) Patient-derived organoids: a promising model for personalized cancer treatment. *Gastroenterol Rep* 6(4):243–245. https://doi.org/10.1093/gastro/goy040
6. Fatehullah A, Tan SH, Barker N (2016) Organoids as an *in vitro* model of human development and disease. *Nat Cell Biol* 18(3): 246–254. https://doi.org/10.1038/ncb3312
7. Yan HHN, Chan AS, Lai FP et al (2023) Organoid cultures for cancer modeling. *Cell Stem Cell* 30(7):917–937. https://doi.org/10.1016/j.stem.2023.05.012
8. Abdolahi S, Ghazvinian Z, Muhammadnejad S et al (2022) Patient-derived xenograft (PDX) models, applications and challenges in cancer research. *J Transl Med* 20(1):206. https://doi.org/10.1186/s12967-022-03405-8
9. Ashok A, Choudhury D, Fang Y et al (2020) Towards manufacturing of human organoids. *Biotechnol Adv* 39:107460. https://doi.org/10.1016/j.biotechadv.2019.107460
10. Corrò C, Novellasdemunt L, Li VSW (2020) A brief history of organoids. *Am J Physiol Cell Physiol* 319(1):C151–C165. https://doi.org/10.1152/ajpcell.00120.2020
11. Drost J, Clevers H (2018) Organoids in cancer research. *Nat Rev Cancer* 18(7):407–418. https://doi.org/10.1038/s41568-018-0007-6
12. Du X, Chen ZZ, Li QW et al (2023) Organoids revealed: morphological analysis of the profound next generation in-vitro model with artificial intelligence. *Bio-Des Manuf* 6(3):319–339. https://doi.org/10.1007/s42242-022-00226-y
13. Du GT, Cao X, Liang JM et al (2020) Medical image segmentation based on U-Net: a review. *J Imaging Sci Technol* 64(2):020508. https://doi.org/10.2352/j.imagingsci.technol.2020.64.2.020508
14. Matthews JM, Schuster B, Kashaf SS et al (2022) OrganoID: a versatile deep learning platform for tracking and analysis of single-organoid dynamics. *PLoS Comput Biol* 18(11):e1010584. https://doi.org/10.1371/journal.pcbi.1010584
15. Bai L, Wu Y, Li GF et al (2024) AI-enabled organoids: construction, analysis, and application. *Bioact Mater* 31:525–548. https://doi.org/10.1016/j.bioactmat.2023.09.005
16. Park T, Kim TK, Han YD et al (2023) Development of a deep learning based image processing tool for enhanced organoid analysis. *Sci Rep* 13(1):19841. https://doi.org/10.1038/s41598-023-46485-2
17. Powell RT, Moussalli MJ, Guo L et al (2022) deepOrganoid: a brightfield cell viability model for screening matrix-embedded organoids. *SLAS Discov* 27(3):175–184. https://doi.org/10.1016/j.slasd.2022.03.004
18. Gritti N, Le Lim J, Anlaş K et al (2021) MOrgAna: accessible quantitative analysis of organoids with machine learning. *Development* 148(18):dev199611. https://doi.org/10.1242/dev.199611
19. Bian XS, Li G, Wang C et al (2021) A deep learning model for detection and tracking in high-throughput images of organoid. *Comput Biol Med* 134:104490. https://doi.org/10.1016/j.compbiomed.2021.104490
20. Wang XW, Wu CY, Zhang SD et al (2022) A novel deep learning segmentation model for organoid-based drug screening. *Front Pharmacol* 13:1080273. https://doi.org/10.3389/fphar.2022.1080273
21. Hörst F, Rempe M, Heine L et al (2024) CellViT: Vision Transformers for precise cell segmentation and classification. *Med Image Anal* 94:103143. https://doi.org/10.1016/j.media.2024.103143
22. Lefferts JW, Kroes S, Smith MB et al (2024) OrgaSegment: deep-learning based organoid segmentation to quantify CFTR dependent fluid secretion. *Commun Biol* 7(1):319. https://doi.org/10.1038/s42003-024-05966-4
23. Abdul L, Xu J, Sotra A et al (2022) D-CryptO: deep learning-based analysis of colon organoid morphology from brightfield images. *Lab Chip* 22(21):4118–4128. https://doi.org/10.1039/d2lc00596d
24. Abdul L, Rajasekar S, Lin DSY et al (2020) Deep-LUMEN assay – human lung epithelial spheroid classification from brightfield images using deep learning. *Lab Chip* 20(24):4623–4631. https://doi.org/10.1039/d0lc01010c
25. Vaswani A, Shazeer N, Parmar N et al (2017) Attention is all you need. In: 31st Conference on Neural Information Processing Systems. https://doi.org/10.48550/arXiv.1706.03762
26. Dosovitskiy A, Beyer L, Kolesnikov A et al (2021) An image is worth 16×16 words: transformers for image recognition at scale. In: International Conference on Learning Representations. https://doi.org/10.48550/arXiv.2010.11929
27. Ronneberger O, Fischer P, Brox T (2015) U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells W et al (Eds.), Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, p.234–241. https://doi.org/10.1007/978-3-319-24574-4_28
28. Khan S, Naseer M, Hayat M et al (2022) Transformers in vision: a survey. *ACM Comput Surv* 54(10s):200. https://doi.org/10.1145/3505244
29. Xie EZ, Wang WH, Yu ZD et al (2021) SegFormer: simple and efficient design for semantic segmentation with transformers. In: 35th Conference on Neural Information Processing Systems, p.12077–12090
30. He KM, Zhang XY, Ren SQ et al (2016) Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, p.770–778. https://doi.org/10.1109/CVPR.2016.90
31. Yao Y, Xu XY, Yang LF et al (2020) Patient-derived organoids predict chemoradiation responses of locally advanced rectal cancer. *Cell Stem Cell* 26(1):17–26. https://doi.org/10.1016/j.stem.2019.10.010
32. Matsubara H, Miyoshi H, Kakizaki F et al (2023) Efficacious combination drug treatment for colorectal cancer that overcomes resistance to KRAS G12C inhibitors. *Mol Cancer Ther* 22(4): 529–538. https://doi.org/10.1158/1535-7163.MCT-22-0411
33. Durinikova E, Reilly NM, Buzo K et al (2022) Targeting the DNA damage response pathways and replication stress in colorectal cancer. *Clin Cancer Res* 28(17):3874–3889. https://doi.org/10.1158/1078-0432.CCR-22-0875
34. Ramzy GM, Norkin M, Koessler T et al (2023) Platform combining statistical modeling and patient-derived organoids to facilitate personalized treatment of colorectal carcinoma. *J Exp Clin Cancer Res* 42:79. https://doi.org/10.1186/s13046-023-02650-z
35. Lee SY, Koo IS, Hwang HJ et al (2023) In vitro three-dimensional (3D) cell culture tools for spheroid and organoid models. *SLAS Discov* 28(4):119–137. https://doi.org/10.1016/j.slasd.2023.03.006