Cytotherapy 25 (2023) 1361−1369

Contents lists available at ScienceDirect

CYTOTHERAPY

journal homepage: www.isct-cytotherapy.org

FULL-LENGTH ARTICLE
Manufacturing

Real-time semantic segmentation and anomaly detection of functional images for cell therapy manufacturing

Rui Qi Chen[1], Benjamin Joffe[2], Paloma Casteleiro Costa[3], Caroline Filan[3], Bryan Wang[3], Stephen Balakirsky[2], Francisco Robles[3], Krishnendu Roy[3], Jing Li[1,*]

[1] *H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia, USA*
[2] *Georgia Tech Research Institute, Georgia Institute of Technology, Atlanta, Georgia, USA*
[3] *Wallace H. Coulter Department of Biomedical Engineering at Georgia Tech and Emory University, Atlanta, Georgia, USA*

ARTICLE INFO

Article History:
Received 22 April 2023
Accepted 24 August 2023

Key Words:
anomaly detection
cell therapy manufacturing
machine learning
semantic segmentation

ABSTRACT

Background aims: Cell therapy is a promising treatment method that uses living cells to address a variety of diseases and conditions, including cardiovascular diseases, neurologic disorders and certain cancers. As interest in cell therapy grows, there is a need to shift to a more efficient, scalable and automated manufacturing process that can produce high-quality products at a lower cost.
Methods: One way to achieve this is using non-invasive imaging and real-time image analysis techniques to monitor and control the manufacturing process. This work presents a machine learning-based image analysis pipeline that includes semantic segmentation and anomaly detection capabilities.
Results/Conclusions: This method can be easily implemented even when given a limited dataset of annotated images, is able to segment cells and debris and can identify anomalies such as contamination or hardware failure.

Introduction

Cell therapy uses living cells to treat complex diseases. It has some significant advantages over small molecule or biologics-based therapies, including a targeted and multimodal treatment approach, leveraging natural repair and surveillance mechanisms, long-lasting effects and the potential to be personalized to each individual patient [1,2]. Recent advancements have identified cell therapy to be capable of treating a wide range of conditions, including cardiovascular diseases, neurologic disorders and certain cancers [3−5]. These advancements have led to a few U.S. Food and Drug Administration−approved therapies, including chimeric antigen receptor T-cell therapies for the treatment of acute lymphoblastic leukemia and various lymphomas [6,7]. However, patient access to these new treatments is limited due to high costs and highly complex manufacturing and supply chain logistics. The complex manufacturing process coupled with high variability in patient or donor cell materials, along with lack of in/at-line quality control, significantly affects the yield, quality, risk and cost of the therapeutic biologic products [8]. There is a significant unmet need for efficient and automated manufacturing processes with embedded quality-by-design and process control to enable consistent, reproducible and quality-driven cell manufacturing to ultimately improve product consistency and quality, lower failure risk, control costs and increase access. One proposed system involves the use of bioreactors that can autonomously monitor, decide on the process status and product quality and control the manufacturing process to efficiently and reliably produce high-quality products in every batch [9].

Quantitative oblique back-illumination microscopy (qOBM) is a label-free, scalable and fast technique well-suited for functional imaging of cells inside bioreactors [10−13]. This technology can be applied to image cells in-process, in real time, during manufacturing to monitor their quality and potentially optimize/improve the therapeutic product. However, to automate manufacturing, real-time segmentation and anomaly detection are needed [9,14−17]. Machine learning, specifically deep learning, is a promising approach for these tasks. Machine-learning algorithms are more adaptable, generalizable and robust than traditional computer vision methods [18,19]. Using machine-learning techniques can improve cell therapy manufacturing by segmenting images for further image analysis and detecting anomalies such as contamination and faulty imaging equipment.

Semantic segmentation classifies each pixel based on the class it belongs to. In qOBM phase images, the primary goal is to

---

[*] Correspondence: Jing Li, Groseclose Building, 765 Ferst Dr NW, Atlanta, GA 30332.
E-mail address: jli3175@gatech.edu (J. Li).

https://doi.org/10.1016/j.jcyt.2023.08.011
1465-3249/

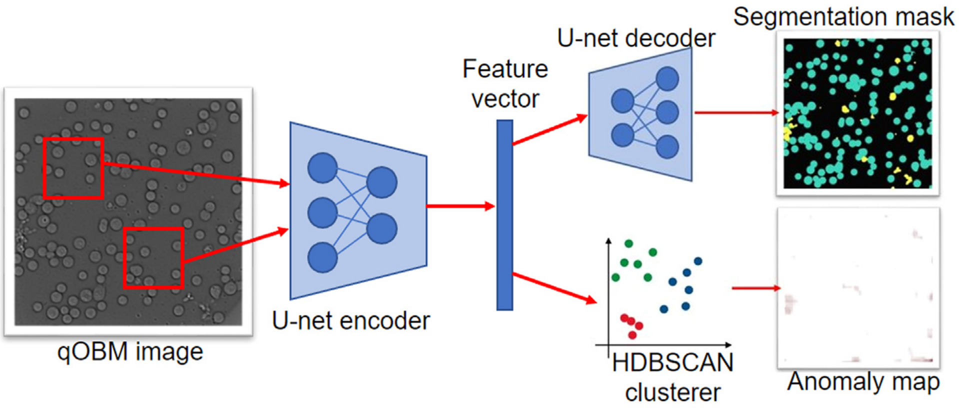

Figure 1. Proposed machine learning−based pipeline. Every patch from an image map to a feature vector via the U-net encoder. The feature vector is used to produce a segmentation mask and an anomaly heatmap in parallel via the U-net decoder and HDBSCAN clusterer, respectively. (Color version of figure is available online.)

distinguish cells and debris from the background, as these objects are indicative of the bioreactor health. Traditional cell-segmentation pipelines use techniques such as filtering, thresholding, edge detection and region growing [15,20]. Since these pipelines are constructed on a case-by-case basis for specific cells and imaging techniques, they may not be adaptable to different types of images [21]. More recently, deep-learning approaches for segmentation have been developed to be more accurate and robust than traditional techniques [22]. Deep learning−based semantic segmentation models usually follow the encoder−decoder architecture [23]. In such a model, images first pass through a series of encoding layers that result in a low-resolution feature representation, which is then mapped to full-input resolution for pixel-wise classification by a series of decoding layers. Proposed in 2015, the U-net became a popular choice for medical image segmentation due to its outstanding performance [24]. Recent works have demonstrated that U-net−based architectures can achieve high accuracy in numerous segmentation tasks even with a small amount of annotated training data [25−27]. These attributes make the U-net architecture appealing for qOBM image segmentation, as image annotation is time-consuming and training data are scarce.

The goal of anomaly detection is to identify observations of unusual events that do not conform to the expected behavior. Since anomalous samples are typically inaccessible, rare or different from future anomalies, the problem of anomaly detection often requires unsupervised solutions that learn from only normal data. Similar to semantic segmentation, deep-learning approaches for anomaly detection have demonstrated success for addressing the complexity of image data [28]. Major approaches use a deep neural network to learn a feature map of the data. This can be done by autoencoders [29], generative adversarial networks [30], self-supervised methods [31,32] or transferring pre-trained representations from other supervised tasks [33]. This work adopts the last-mentioned idea by leveraging the supervised segmentation task to generate lower-dimensional feature representations of images, which are used to train the anomaly detection model. Feature representations of normal data should be close to each other, whereas anomalous features should be distant in feature space. Therefore, clustering models are sometimes the go-to choices for real-time anomaly detection, as they are naturally unsupervised and are more computationally lightweight than other machine-learning algorithms [34]. Hierarchical clustering and density-based spatial clustering of applications with noise (HDBSCAN) stands out, as it automatically calculates the number of clusters, handles clusters with different density and shapes and can output anomaly scores with the global−local outlier scores from hierarchies (GLOSH) algorithm [35]. Past works have also shown the suitability of HDBSCAN in various anomaly detection applications [36−38].

This work proposes an efficient machine learning−based image analysis pipeline for real-time monitoring of *in vitro* T cells via qOBM images, requiring as little as two annotated images to train. The pipeline combines semantic segmentation and anomaly-detection capabilities in a single system that can be used in closed-loop cell therapy manufacturing processes. Figure 1 illustrates the pipeline, which consists of a sliding window that takes square patches from a full-sized image, which are mapped to feature vectors by a modified U-net encoder [24]. The corresponding modified U-net decoder then maps the feature vectors to generate a segmentation mask, whereas the HDBSCAN clusterer generates an anomaly heatmap [35]. This setup allows naturally parallel computation of both tasks. The anomaly scores in the heatmap can also be used as uncertainty scores of segmentations; not only locating anomalous objects in the image, but also revealing regions where segmentation may be inaccurate. This real-time pipeline can aid in the monitoring of T-cell quality in the manufacturing process, further enabling feedback control.

Methods

Data acquisition

The data in this study were collected from multiple T-cell expansion runs (institutional review board protocol no. H17348). During each expansion, human CD3+ T cells are seeded in PBS-Mini Vertical-Wheel Bioreactor from PBS Biotech with TexMACS media from Miltenyi Biotec (San Diego, CA, USA). The cells are expanded for 14 days with the goal of achieving a high yield. Every day, starting on day 7, the cells in the bioreactor are imaged with qOBM. The raw images are 2048 by 2048 pixels, with a single channel for quantitative phase. Based on the observed intensity distribution of quantitative phase, contrast normalization is applied to the images by clipping at two fixed thresholds and scaling to a range of 0 to 1. This step reduces noise from extreme values, enhances the visibility of features crucial for segmentation and helps the machine learning pipeline by stabilizing gradient flow during model training.

Along with imaging, the cell concentration and viability (i.e., proportion of cells alive) are measured with a Via1-Cassette and NucleoCounter NC-200 from ChemoMetec to monitor the bioreactor health status. To do this, a small cell sample is loaded into a cassette, where they are stained by fluorophores acridine orange and 4′,6-diamidino-2-phenylindole. Acridine orange stains both live and dead cells, providing a total cell count, whereas 4′,6-diamidino-2-phenylindole stains the dead cells, from which viability is calculated.

The data were collected from three different expansion runs, denoted by *A*, *B* and *C*. Run *A* had 10 images, run *B* had 17 and run *C* had 19. To produce labels for semantic segmentation, all images from run *A* and 6 images from run *B* were manually annotated with three classes: background, cell and debris. These annotations were considered the ground truth for model training and testing. Run *A* was used for training and validation, then the subsequent run *B* was used for generating performance metrics. In addition, images from the third expansion run, *C*, were analyzed for anomaly detection. A set of six normal unlabeled images from run *C* were used to train the anomaly detection clusterer, whereas the remaining 13 images were used for testing. Since anomalies can be global to the whole image or local to a specific part, the test images from run *C* were cropped into multiple smaller patches and manually labeled as normal or anomalous. The test dataset contained three types of anomalies: abnormal imaging leading to reduced image contrast (e.g., poor focusing conditions), delayed imaging resulting in dead and expired cells and contamination with yeast during cell expansion.

Quantitative oblique back illumination microscopy

qOBM enables label-free three-dimensional quantitative phase imaging using epi-illumination. This technology is well suited to image thin transparent samples (e.g., cells *in vitro*) like traditional phase imaging technologies but can also analyze complex scattering samples (e.g., thick tissues [39] and adherent/suspended cells inside bioreactors [10]), which cannot be analyzed with traditional phase technologies. The qOBM system uses a traditional bright-field microscope configuration and can also be miniaturized into a flexible fiber-based system that can also be applied to image inside bioreactors [39,40] among other applications. Illumination for qOBM comprises four low-cost LEDs, coupled to 1-mm multimode fibers, which illuminate the samples sequentially in epi-mode. The fibers are configured 90° from each other around the microscope objective. LED light deployed via the fibers enters the sample, where it is multiply scattered. The scattered light eventually generates a virtual light source within the sample, emulating a transmission-based microscope. In thin samples, a scattering medium, such as a polydimethylsiloxane device with titanium–dioxide nanoparticles with well characterized scattering properties, can be applied (as was the case in this work). Quantitative phase image reconstruction is achieved via deconvolution of the raw intensity images with the transfer function of the system. More details on the system hardware and reconstruction algorithm can be found in previous works [12,11,40]. The processed qOBM images provide quantitative phase information that is proportional to the optical path length of the sample (the product of the refractive index and thickness of the sample), with a sensitivity of ~6 nm. Using a 60×/0.7na objective and illumination at 720 nm, the lateral resolution is approximately 0.63 μm.

Modified U-net

The U-net is a popular deep-learning architecture used for semantic segmentation tasks, where the goal is to classify each pixel within an image. Since the U-net architecture is well established and is shown to have good segmentation performance on biomedical images, a modified U-net was used for semantic segmentation [24]. The qOBM images only have a single-input channel and are relatively simple, so the U-net was modified to have 16 times fewer channels in all convolutional layers. This drastically reduced the number of training parameters and resulted in a model better suited for training with a small dataset.

The model was implemented in PyTorch and an NVIDIA Tesla V100 GPU was used for training and inference [41]. Various loss functions were experimented with, including Cross-Entropy (CE) loss, Weighted Cross-Entropy (WCE) loss, and Dice loss [42,43]. Preliminary experiments using images from run *A* determined an appropriate optimizer, learning rate scheduler and model hyperparameters. An AdamW optimizer with an initial learning rate of 0.001 and a cosine annealing scheduler with warm restarts, doubling periods and a 150-epoch initial period were found to work best for training [44,45]. With an additional weight decay of 0.01, the model was trained for 2250 epochs. Rather than using full-sized images, the modified U-net used patches of smaller size, denoted by size S_U, for training and inference. All patches used in this work are square, where their sizes are denoted by their widths. The segmentation performance of the modified U-net was measured across different S_U, loss functions, and number of training images. From run *A*, seven images were used for training, and three were used for validation.

To effectively make use of the small, labeled dataset, the images were scaled, augmented with randomized transformations, and cropped to S_U. First, the pixel values were scaled to values ranging from 0 to 1, for faster training convergence. Next, the image was randomly horizontally flipped. Then, an affine transformation was applied with a random rotation, scaling and shear. The last step was a random square crop with size S_U within the transformed image. During each epoch in training, each image in the training set was augmented 100 times independently, resulting in 100 distinct training samples per image.

At inference time, the test images are also cropped into S_U-sized patches. The crops are done with a sliding window, with one third overlap and repeated on a vertical flip of the image. The resulting predictions are then stitched together, with overlapping predictions averaged out.

HDBSCAN-GLOSH outlier detection

In this anomaly detection task, the model has no access to anomalous images at training time. Therefore, the model must learn the common characteristics of normal samples and be able to discern anomalous samples at inference time. For this unsupervised machine-learning task, the encoder portion of the modified U-net was leveraged to produce low-dimensional latent representations of image patches. Considering the modified U-net was trained with S_U-sized patches, the set of normal samples used for training were cropped to this size with a sliding window, striding across with steps of 32 pixels, and generating numerous training patches for the anomaly detection clusterer. After being encoded by the modified U-net, the output is mean-pooled across the height and width to produce a 64-length feature vector. To learn the characteristics of normal samples, feature vectors from normal images were fed into a HDBSCAN clusterer [46].

HDBSCAN is an extension of the classic DBSCAN [47] that can not only deal with clusters of different shapes but also of different sizes. A high-level description of the HDBSCAN implementation consists of five steps [48]. First, the space is transformed by spreading apart data points with low densities, to make the subsequent clustering more robust to noise. Second, the new space is viewed as a graph, where the data points are vertices, and the edges are weighted by the transformed distance. The minimum spanning tree of this graph is constructed. Third, the edges of the tree are sorted in increasing order. Iterating through the edges, a cluster hierarchy is formed by creating a new merged cluster for each edge. Fourth, the cluster hierarchy is condensed down into a smaller tree, where each node represents a group of data points. This is done by continuously splitting the cluster hierarchy starting from the root, retaining clusters that are at least of the specified minimum size. Lastly, flat clusters are extracted when the stability of the cluster is greater than the sum of its children. The stability is measured by how long the points in the cluster persisted after cluster creation and before being split in the previous step. Any points not extracted into a cluster is labeled as noise.

For anomaly detection, the GLOSH algorithm assigns a score ranging from 0 to 1 [35]. Higher scores are assigned to points that have a substantially lower density than the densities of points associated within the current and child clusters. This model is well suited for the anomaly detection task, as it works well with noisy data and performs efficiently with 64-dimensional features.

The main clusterer parameter that affects anomaly detection is the number of samples in a neighborhood for a point to be considered a core point. Increasing this parameter increases the number of points considered as noise. Since the clusterer is trained on normal data free of anomalies, this parameter was set to 1, minimizing noise labels.

At inference time, depending on the size of the evaluated image or patch, denoted by S_A, the model can be used to directly perform classification or generate an anomaly heatmap. When S_A is equal to S_U, the model assigns one outlier score for the patch. For classification, a cutoff threshold is used. Having a score greater than the threshold classifies the evaluated patch as anomalous. When S_A is greater than S_U, such as evaluating a full-sized image, a stride 32 sliding window is first used to crop S_U-sized patches. The clusterer then assigns outlier scores to each S_U-sized patch, which can be combined to create an anomaly heatmap for the larger S_A-sized patch or image. To do this, the scores of the S_U-sized patches were averaged at overlapping areas. For classification, the anomaly score was assigned to be the maximum score in the heatmap. Encoders from modified U-nets with different numbers of training data were compared for anomaly detection performance. In addition, different S_A were tested to find the most appropriate inference window size for anomaly detection.

Results

Experimental setup

The segmentation performance of models trained with different loss functions and S_U were compared. The set of loss functions included CE loss, WCE loss and Dice loss. The weights for WCE were 0.5, 1, 1.5 for the classes of Background, Cell and Debris respectively to account for class imbalance in the training data. The compared S_U ranged from 128 to 512. Multiple evaluation metrics were used to assess the segmentation performance, including accuracy, class-wise Dice coefficients and correlation coefficients with bioreactor attributes of cell concentration and viability.

The accuracy and Dice coefficients are calculated for the six annotated test images from run *B* [49]. Dice coefficients for each class and the total accuracy are reported. The Dice coefficient is the harmonic mean of the precision and recall, which can be calculated from the number of true positives (TP), false positives (FP), and false negatives (FN) as in Equation (1). For instance, for the Cell class, TP is the number of correctly classified Cell pixels, FP is the number of pixels from Background or Debris that are misclassified as Cell and FN is the number of Cell pixels misclassified as Background or Debris.

$$\text{Dice Coefficient} = 2\frac{\text{precision}\cdot\text{recall}}{\text{precision}+\text{recall}} = \frac{2\text{TP}}{2\text{TP}+\text{FP}+\text{FN}} \tag{1}$$

To further validate the segmentation results, the segmentation outputs of all 17 images in run *B* are used to calculate proxy indicators for cell concentration and viability. Since Cell and Debris labels represent healthy and dead cells, respectively, the proportion of different labels in each image should be highly indicative of cell concentration and viability. The indicator for cell concentration is measured as the number of Cell pixels divided by the total number of pixels, and the indicator for viability is the number of Cell pixels divided by the number of non-Background pixels, as shown in Equation (2) and Equation (3). The Pearson correlation coefficient is calculated to measure the linear correlation between the indicators and measurements taken directly from the bioreactor [50].

$$\text{Cell Concentration Indicator} = \frac{|\text{Cell}|}{|\text{Background}|+|\text{Cell}|+|\text{Debris}|} \tag{2}$$

$$\text{Viability Indicator} = \frac{|\text{Cell}|}{|\text{Cell}|+|\text{Debris}|} \tag{3}$$

After the optimal loss function and S_U were determined, the effect of training set size on model performance was explored. Starting at seven images, the training set was reduced image by image, discarding the perceived least informative image until 1 image remained. For instance, images that only contained a few objects were primarily discarded.

For anomaly detection, images from run *C* were divided into two subsets for training and testing. The training set contained only normal images, whereas the test set contained both normal and anomalous images. The clusterer is then trained with the training set, adhering to the procedure outlined in the section "HDBSCAN-GLOSH outlier detection." For testing, the test images were cropped into non-overlapping square patches. Patch sizes ranging from 128 to 512 pixels were tested to find an optimal S_A for anomaly detection.

Each patch from the test set was manually labeled with ground truths, either being normal or anomalous. The area under the receiver operating characteristic curve (AUC) score was used to measure anomaly detection performance. Classification used the 98th percentile of the outlier scores of the normal training patches. Precision scores, recall scores and contingency tables for different anomalies were reported. In addition, full-sized anomaly heatmaps were qualitatively assessed.

Cell segmentation

Using secen training images, modified U-nets were trained with varying loss functions and S_U. The performance of these models is summarized in Table 1. As can be seen, the accuracy and correlation coefficients are relatively good across all models, with all values above 0.9. The class-specific Dice coefficients imply that the models are best at segmenting background and worst at segmenting debris.

Table 1
Segmentation performance comparison of modified U-nets trained with different loss functions and S_U.

			Dice coefficient			Correlation coefficient	
Loss Function	S_U	Accuracy	Background	Cell	Debris	Cell Concentration	Viability
CE	128	0.9486	0.9697	0.9160	0.7116	0.9585	**0.9937**
	256	0.9409	0.9654	0.8902	0.7259	0.9507	0.9931
	512	0.9334	0.9613	0.8707	0.7008	0.9613	0.9878
WCE	128	0.9407	0.9647	0.8873	0.8131	0.9607	0.9881
	256	0.9263	0.9560	0.8599	0.7833	0.9578	0.9921
	512	0.9075	0.9427	0.8253	0.7715	0.9532	0.9805
Dice	128	0.9529	0.9733	0.9098	**0.8343**	0.9621	0.9899
	256	**0.9566**	**0.9758**	**0.9180**	0.8247	**0.9635**	0.9922
	512	0.9434	0.9663	0.8890	0.8193	0.9578	0.9901

The training dataset contained seven images. In each column, bold font denotes the best value.

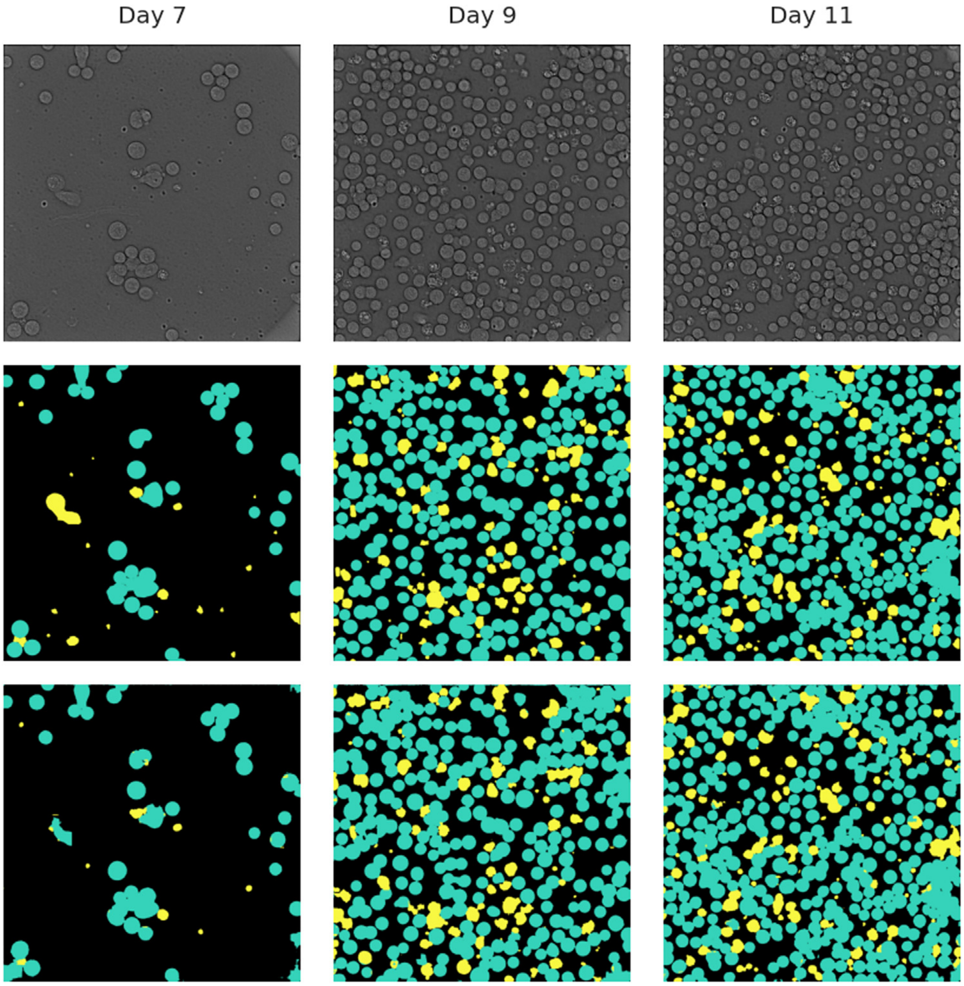

Figure 2. Example segmentation results of segmentation with Dice-128 at different stages of cell expansion. The top row shows qOBM images, the second row shows the manually annotated labels, and the bottom row shows the predicted segmentation map. In the segmentation maps, green represents Cells, yellow represents Debris, and black represents Background. (Color version of figure is available online.)

Another observation is that the model performance generally decreases as S_U increases. Moreover, for each S_U, Dice loss outperforms WCE loss and CE loss across accuracy and Dice coefficient metrics. Weighting the CE loss drastically improved the debris segmentation capabilities. The best performing models across Dice coefficients and accuracy scores were trained with Dice loss and S_U of 128 (Dice-128) or 256 (Dice-256). These two models had similar performance. Figure 2 qualitatively illustrates the segmentation performance of Dice-128 on various normal images. The segmentation is generally accurate, only visibly underperforming on objects not in the focal plane and in the corners, which are plagued by imaging artifacts.

Correlation coefficients were similar across all models. The correlation coefficients were generally above 0.95 for cell concentration and around 0.99 for viability. Overall, Dice loss was the best performing loss function, and performance increased with smaller S_U. Therefore, the Dice-128 model was chosen to explore the effects of decreasing the training dataset size on segmentation performance. The results are summarized in Table 2.

As expected, in Table 2, it can be observed that as the number of training images decreases, the model performance generally decreases across all evaluation metrics. The model slightly deteriorates when removing training samples, until 2 images are used for training. When going from 2 images to 1, the model performance sees a drastic decline.

Anomaly detection

As the Dice-128 and Dice-256 models were deemed to be the best at semantic segmentation, their encoders would then be used to generate features for anomaly detection. Since U-nets trained with fewer training images also had good segmentation performances, the 2-training-image variants of these models were also examined. The results are summarized in Table 3.

Since the test patches are cropped from a dataset in a non-overlapping manner, the number of inference patches decreases as S_A increases, as evidenced by the contingency table. Despite this, the evaluated dataset remains relatively balanced between normal and anomalous patches, and the number of samples is sufficient for analysis even for a S_A of 512. Inspecting the contingency table, there does not seem to be a type of anomaly that is easier to classify than others.

The tested models all produced AUC scores around the range of 0.7–0.8. When we compared AUC scores, modified U-nets trained with seven images were marginally better than the corresponding ones trained with only two images. Furthermore, having S_A slightly larger than S_U seems to be the optimal choice in all cases. The model with the highest AUC is Dice-128-7 with approximately 0.83 when S_A is 256.

When examining the tradeoff between precision and recall, as S_A increases, the recall is improved at the cost of decreased precision. This is because using the classification method outlined in “HDBSCAN-

Table 2
Segmentation performance comparison of modified U-nets trained with different number of training images.

		Dice coefficient			Correlation coefficient	
Number of training images	Accuracy	Background	Cell	Debris	Cell Concentration	Viability
7	**0.9529**	**0.9733**	**0.9098**	**0.8343**	**0.9621**	0.9899
6	0.9339	0.9605	0.8701	0.8227	0.9540	0.9889
5	0.9425	0.9665	0.8876	0.8207	0.9605	**0.9900**
4	0.9257	0.9543	0.8533	0.8302	0.9440	0.9878
3	0.9324	0.9608	0.8640	0.8139	0.9490	0.9848
2	0.9209	0.9511	0.8448	0.8264	0.9399	0.9867
1	0.8198	0.8740	0.6872	0.8105	0.6701	0.9832

The model used Dice loss and S_U of 128. In each column, bold font denotes the best value.

GLOSH outlier detection,” the cutoff threshold does not change, but the score used for classification is taken as the maximum of the scores in the inference window. This resulted in a recall score of 1 in many cases where S_A was 512, correctly classifying all anomalous samples. In these cases, a low precision score of approximately 0.6 was seen, as most normal samples are also misclassified as anomalous.

Anomaly heatmaps of full-sized images are depicted in Figure 3. For visualization purposes, the anomaly scores are clipped to the interval from the 95th percentile of the outlier scores of the training samples to 1.0, corresponding to fully transparent and fully opaque respectively. As can be seen from the expired cells and yeast contamination examples, the anomaly localization performance is good, generally assigning higher scores to areas with anomalies and lower scores to areas filled with normal cells and background. Another observation is that the anomaly score is higher around the corners and the edges of images, where there exist anomalies in the form of artifacts caused by imaging hardware. Additionally, in the normal image, objects not in the focal plane get assigned higher anomaly scores, as they become objects with reduced contrast. Recalling Figure 2, the segmentation performance on these objects is correspondingly subpar, suggesting that the anomaly score can be used as a proxy for segmentation uncertainty.

Discussion

The use of artificial intelligence and machine-learning techniques has the potential to greatly improve cell therapy manufacturing by allowing for increased automation and control in the production processes. These novel techniques can be used to analyze sensor and image data, monitor cell culture conditions and provide insights for process optimization and quality control. For effective real-time monitoring of a cell therapy manufacturing process through images, cell segmentation is an important step to generate regions of interest [14]. Since the therapeutic product consists of these cells, identifying them within the image is essential for further analysis, be it study of its dynamics or generating features for phenotype classification. The results signify that a modified U-net is capable of segmenting both Cell and Debris well. The slightly lower performance for Debris is expected, as Debris is the least abundant class.

When we compared different S_U and loss functions, smaller S_U are preferred, and Dice loss yields the best results. Using a smaller S_U poses a more difficult segmentation task for the model to learn, as less context information is provided for each object and more objects wind up on patch boundaries. This provides additional regularization during training to reduce overfitting [51]. As a result, such a model will be more generalizable on new data at inference time. For the compared loss functions, CE does not account for the class imbalance in the training data, whereas WCE explicitly uses weights provided to account for the imbalance. On the other hand, Dice loss implicitly takes class imbalance into consideration [52]. Since the primary performance metric of interest is the Dice coefficient, using a loss function that directly optimizes for this yielded the best results.

When reducing the number of training images, the modified U-net could still achieve a surprisingly high segmentation performance. This suggests that the segmentation of qOBM images is a relatively simple task, and the data-augmentation techniques were enough to train a generalizable model. However, a single image does not capture all image scenarios and is insufficient for training, whereas two images can include both sparsely and densely populated images. Consequently, for deployment in an automated, closed-loop system,

Table 3
Anomaly detection performance comparison of different modified U-net encoders.

					Normal		Contrast		Expired		Yeast	
Model	S_A	AUC	Precision	Recall	TN	FP	TP	FN	TP	FN	TP	FN
Dice-128-7	128	0.7113	0.6928	0.4172	1547	278	426	618	95	96	106	162
	256	0.8258	0.6359	0.9532	172	233	253	11	91	2	63	7
	512	0.7884	0.6000	1.0000	18	76	64	0	32	0	18	0
Dice-128-2	128	0.7056	0.7310	0.2369	1694	131	272	772	51	140	33	235
	256	0.7967	0.7500	0.7307	301	104	207	57	64	29	41	29
	512	0.7002	0.6644	0.8684	44	50	61	3	27	5	11	7
Dice-256-7	256	0.7502	0.7092	0.7026	282	123	159	105	75	18	66	4
	512	0.7980	0.5672	1.0000	7	87	64	0	32	0	18	0
Dice-256-2	256	0.6745	0.6756	0.5902	284	121	137	127	53	40	62	8
	512	0.7596	0.5700	1.0000	8	86	64	0	32	0	18	0

The last number in the model name denotes the number of images used to supervise segmentation training. The contingency table shows the instances of classification for the classes of normal patches (Normal) and anomalies of reduced contrast (Contrast), expired cells (Expired), and yeast contamination (Yeast). The table represents the number of normal patches classified as normal (TN), normal classified as anomalous (FP), anomalous classified as anomalous (TP), and anomalous classified as normal (FN).
FN, false negative; FP, false positive; TN, true negative; TP, true positive.

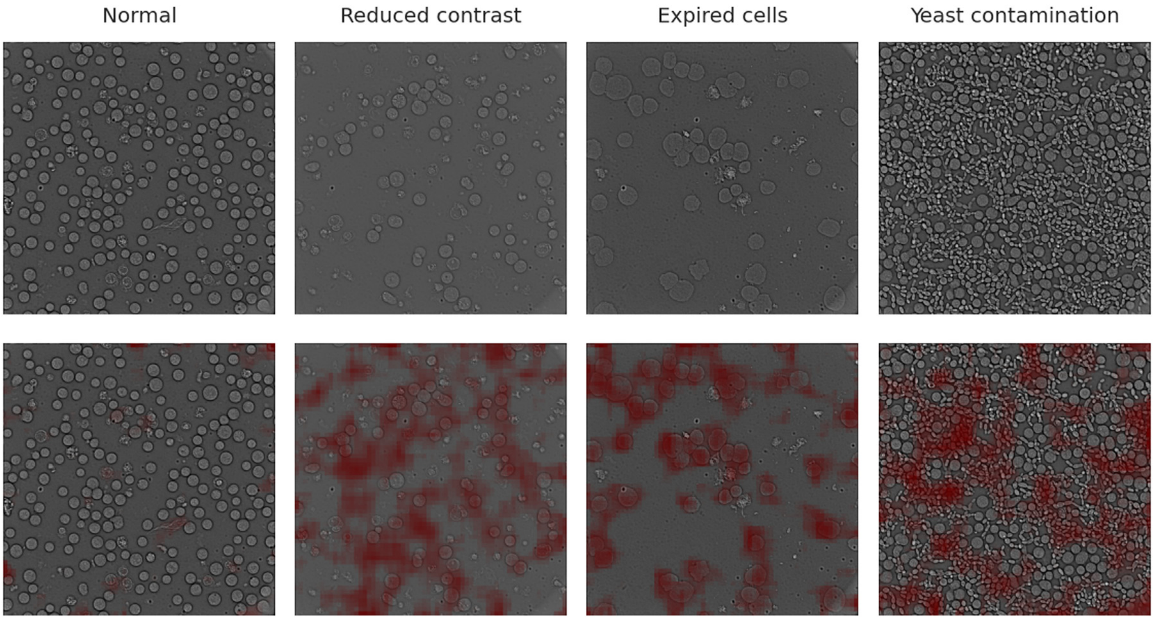

Figure 3. Anomaly detection of Dice-128-7 model on normal and anomalous images. The first row illustrates raw qOBM images, and the second has an anomaly heatmap overlaid. Greater opacity of maroon indicates greater anomaly scores. (Color version of figure is available online.)

training with more images leads to better results, but having only a few training images is also sufficient, as long as they are varied and sufficiently encompass future images.

The correlation coefficient for viability is consistently greater than cell concentration. This is because the viability measurements of the test images have a bimodal distribution, either being very high or very low. In contrast, cell concentration measurements were more evenly distributed across a large range. Due to the limited variability in viability measurements, it is easier for a model to achieve a higher correlation coefficient. Nevertheless, the astounding correlation coefficients consistently above 0.9 for both metrics further reinforce that the proposed machine learning-based pipeline segments sufficiently well to inform further decisions.

In anomaly detection, modified U-nets trained with more images likely performed better because they have seen more data during training and are more generalizable at producing encodings for unseen data. The achieved 0.83 AUC score by Dice-128-7 for a S_A of 256 is high, considering no post-processing other than a simple mean-pooling was employed for the modified U-net encoded feature vector. Despite less supervision during the encoder training, Dice-128-2 still achieved an AUC score of 0.8 for a S_A of 256. This suggests that anomaly detection is still possible even when the availability of annotated data is severely limited. It is also important to note that the semantic segmentation module struggles with out-of-focus cells, which have similar features to the anomalous “reduced contrast” images, potentially leading to inaccurate segmentation. However, the anomaly detection module can flag these out-of-focus cells as anomalous, rendering the incorrect labeling less critical.

Although the image analysis pipeline presented in this work has demonstrated promising results, it also has several limitations that should be acknowledged. First, since the proposed pipeline uses a full-sized image for inference, the different AUC scores attained by different S_A are less meaningful. The smaller S_A were used to validate the proposed method, given that only a handful of anomalous images were available for testing. Therefore, when classifying full-sized images, the anomaly score should simply use the maximum. A different image scoring method should be used to reduce the influence of out-of-plane objects in normal images. Otherwise, it is recommended to only use regions with low anomaly scores for decision-making, as they are most likely to have accurate segmentations.

Second, the anomaly detection module may miss small anomalies due to the feature vector generation method, which averages across each patch. If only a small section of the original patch contains the anomaly, then the pooling would lessen the contribution from the anomaly. Therefore, setting S_U smaller than 128 may help catch these smaller anomalies, but this would make segmentation harder. There are also several other methods that can potentially improve anomaly detection performance at the cost of added complexity. Building on the patch-wise approach, multiple U-nets with different S_U could be used to generate multiple anomaly heatmaps, which are then aggregated as in an ensemble [53,54]. Moreover, additional processing could be used after the encoder step to generate feature vectors more suited for anomaly detection [54,55].

Despite the aforementioned limitations and possible improvements, the proposed pipeline is easy to deploy and performs well with low amounts of data. The segmentation is reliable on normal images, and if segmentation fails, the anomaly detection module will likely catch it. Since the same encoder is used, a failed segmentation results from an out-of-domain feature vector being passed into the decoder, which would also be flagged by the anomaly detection module. Therefore, the anomaly heatmap could be used to inform the uncertainty of the segmentation at different locations in the image. When used together, the segmentation mask and anomaly heatmap can prioritize the analysis of cells in regions with low anomaly scores while excluding those with high anomaly scores.

Conclusions

This work presents a machine learning–based image analysis pipeline that can provide real-time segmentation masks and anomaly heatmaps to improve closed-loop cell therapy biomanufacturing. The results suggest that a U-net based model can effectively perform semantic segmentation on qOBM phase images with a limited number of annotated images. In addition, the encoder features generated by the U-net can be used to detect anomalies including those that the network has never seen, which is extremely powerful.

In deployment, non-invasive imaging paired with the real-time image analysis pipeline can dynamically monitor the growth and behavior of cultured cells. The accurate segmentation results also offer a wealth of possibilities for further analysis. For instance, the segmented regions of interest enable feature extraction for cell phenotype classification or quality prediction. These insights can be used to drive real-time decisions for regulating crucial parameters like nutrient supply, growth factors, and environmental conditions to consistently yield high-quality therapeutic products. Furthermore, the anomaly detection capabilities swiftly identify deviations such as contamination or hardware issues, prompting immediate corrective actions.

Looking ahead, an important future direction is the refinement of the anomaly detection capabilities. Detecting and distinguishing different anomalies can help maintain experimental consistency, prevent unreliable data generation, and avoid resource wastage. By incorporating aspects of lifelong learning, a model can retain knowledge about past seen anomalies, enabling it to distinguish previously encountered anomaly types while continuously adapting to new, previously unseen anomaly types. Overall, the presented pipeline represents a promising stride towards improved cell therapy manufacturing processes.

Funding

The authors acknowledge the support from Georgia Tech Research Institute, via their Independent Research and Development budget; the Billie and Bernie Marcus Foundation; the Georgia Research Alliance; National Institutes of Health National Institute of General Medical Sciences (R35GM147437); Burroughs Wellcome Fund (CASI BWF 1014540) and Georgia Tech.

Declaration of Competing Interest

The authors have no commercial, proprietary or financial interest in the products or companies described in this article.

Author Contributions

Conception and design of the study: RQC, BJ, JL. Acquisition of data: PCC, CF, BW, FR, KR. Analysis and interpretation of data: RQC. Drafting or revising the manuscript: RQC, BJ, SB, FR, KR, JL. All authors have approved the final article.

Acknowledgments

The authors thank Tony Pan (Emory University) for their insightful discussions.

References

[1] Ben Jehuda Ronen, Shemer Yuval, Binah Ofer. Genome editing in induced pluripotent stem cells using crispr/cas9. Stem Cell Reviews and Reports 2018;14(3):323–36.

[2] Takahashi Toshio. Organoids for drug discovery and personalized medicine. Annual Review of Pharmacology and Toxicology 2019;59:447–62.

[3] Muthu Sathish, Bapat Asawari, Jain Rashmi, Jeyaraman Naveen, Jeyaraman Madhan. Exosomal therapy—a new frontier in regenerative medicine. Stem Cell Investigation 2021;8.

[4] Alessandrini M, Preynat-Seauve O, De Bruin K, Pepper Michael Sean. Stem cell therapy for neurological disorders. South African Medical Journal 2019;109(8 Suppl 1):S71–8.

[5] Dai Hanren, Wang Yao, Lu Xuechun, Han Weidong. Chimeric antigen receptors modified T-cells for cancer therapy. Journal of the National Cancer Institute 2016;108(7):djv439.

[6] Mullard Asher. FDA approves first CAR T therapy. Nature Reviews Drug Discovery 2017;16(10):669–70.

[7] Rosa Saez-Ibañez Ana, Upadhaya Samik, Partridge Tanya, Shah Monica, Correa Diego, Campbell Jay. Landscape of cancer cell therapies: trends and real-world data. Nature Reviews Drug Discovery 2022;21(9):631–2.

[8] Mikhael Joseph, Fowler Jessica, Nina Shah. Chimeric antigen receptor t-cell therapies: barriers and solutions to access. JCO Oncology Practice OP—22. 2022.

[9] Wang Bryan, Bowles-Welch Annie C, Yeago Carolyn, Roy Krishnendu. Process analytical technologies in cell therapy manufacturing: state-of-the-art and future directions. Journal of Advanced Manufacturing and Processing 2022;4(1):e10106.

[10] Costa Paloma Casteleiro, Wang Bryan, Filan Caroline, Bowles-Welch Annie, Yeago Carolyn, Roy Krishnendu, Robles Francisco E. Functional imaging with dynamic quantitative oblique back-illumination microscopy. Journal of Biomedical Optics 2022;27(6):066502.

[11] Ledwig Patrick, Robles Francisco E. Quantitative 3D refractive index tomography of opaque samples in epi-mode. Optica Jan 2021;8(1):6–14.

[12] Ledwig Patrick, Robles Francisco E. Epi-mode tomographic quantitative phase imaging in thick scattering samples. Biomed. Opt. Express Jul 2019;10(7):3605–21.

[13] Ledwig Patrick, Sghayyer Moses, Kurtzberg Joanne, Robles Francisco E. Dual-wavelength oblique back-illumination microscopy for the non-invasive imaging and quantification of blood in collection and storage bags. Biomed. Opt. Express Jun 2018;9(6):2743–54.

[14] Grys Ben T, Lo Dara S, Sahin Nil, Kraus Oren Z, Morris Quaid, Boone Charles, Andrews Brenda J. Machine learning and computer vision approaches for phenotypic profiling. Journal of Cell Biology 2017;216(1):65–71.

[15] Kraus Oren Z, Frey Brendan J. Computer vision for high content screening. Critical Reviews in Biochemistry and Molecular Biology 2016;51(2):102–9.

[16] Zhurikhina Anastasia, Qi Timothy, Hahn Klaus M, Elston Timothy C, Tsygankov Denis. Edgeprops: a computational platform for correlative analysis of cell dynamics and near-edge protein activity. Rho GTPases: Methods and Protocols 2018: 47–56.

[17] Pilcher William, Yang Xingyu, Zhurikhina Anastasia, Chernaya Olga, Xu Yinghan, Qiu Peng, Tsygankov Denis. Shape-to-graph mapping method for efficient characterization and classification of complex geometries in biological images. PLoS Computational Biology 2020;16(9):e1007758.

[18] Sabokrou Mohammad, Fathy Mahmood, Hoseini Mojtaba, Klette Reinhard. Real-time anomaly detection and localization in crowded scenes. In: In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops; 2015. p. 56–62.

[19] Long Jonathan, Shelhamer Evan, Darrell Trevor. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2015. p. 3431–40.

[20] Meijering Erik. Cell segmentation: 50 years down the road [life sciences]. IEEE Signal Processing Magazine 2012;29(5):140–5.

[21] Bengtsson Ewert, Wahlby C, Lindblad Joakim. Robust cell image segmentation methods. Pattern Recognition Image Analysis 2004;14(2):157–67.

[22] Du Getao, Cao Xu, Liang Jimin, Chen Xueli, Zhan Yonghua. Medical image segmentation based on u-net: a review. Journal of Imaging Science and Technology 2020;64:1–12.

[23] Taghanaki Saeid Asgari, Abhishek Kumar, Cohen Joseph Paul, Cohen-Adad Julien, Hamarneh Ghassan. Deep semantic segmentation of natural and medical images: a review. Artificial Intelligence Review 2021;54(1):137–78.

[24] Ronneberger Olaf, Fischer Philipp, Brox Thomas. U-net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells W, Frangi A, eds. International conference on medical image computing and computer-assisted intervention, Cham: Springer; 2015:234–41.

[25] Bardis Michelle, Houshyar Roozbeh, Chantaduly Chanon, Ushinsky Alexander, Glavis-Bloom Justin, Shaver Madeleine, Chow Daniel, Uchio Edward, Chang Peter. Deep learning with limited data: organ segmentation performance by u-net. Electronics 2020;9(8):1199.

[26] Le An Pei Zhang, Adeli Ehsan, Wang Yan, Ma Guangkai, Shi Feng, Lalush David S, Lin Weili, Shen Dinggang. Multi-level canonical correlation analysis for standard-dose pet image estimation. IEEE Transactions on Image Processing 2016;25(7):3303–15.

[27] Zhu Jinhan, Zhang Jun, Qiu Bo, Liu Yimei, Liu Xiaowei, Chen Lixin. Comparison of the automatic segmentation of multiple organs at risk in CT images of lung cancer between deep convolutional neural network-based and atlas-based techniques. Acta Oncologica 2019;58(2):257–64.

[28] Ruff Lukas, Kauffmann Jacob R, Vandermeulen Robert A, Montavon Grégoire, Samek Wojciech, Kloft Marius, Dietterich Thomas G, Müller Klaus-Robert. A unifying review of deep and shallow anomaly detection. Proceedings of the IEEE 2021;109(5):756–95.

[29] Nick Pawlowski, Matthew CH Lee, Martin Rajchl, Steven McDonagh, Enzo Ferrante, Konstantinos Kamnitsas, Sam Cooke, Susan Stevenson, Aneesh Khetani, Tom Newman, et al. Unsupervised lesion detection in brain CT using Bayesian convolutional autoencoders. 2018.

[30] Schlegl Thomas, Seeböck Philipp, Waldstein Sebastian M, Langs Georg, Schmidt-Erfurth Ursula. f-anogan: fast unsupervised anomaly detection with generative adversarial networks. Medical Image Analysis 2019;54:30–44.

[31] Li Chun-Liang, Sohn Kihyuk, Yoon Jinsung, Pfister Tomas. Cutpaste: self-supervised learning for anomaly detection and localization. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2021. p. 9664–74.

[32] Sheynin Shelly, Benaim Sagie, Wolf Lior. A hierarchical transformation-discriminating generative model for few shot anomaly detection. In: Proceedings of the IEEE/CVF international conference on computer vision; 2021. p. 8495–504.

[33] Reiss Tal, Cohen Niv, Bergman Liron, Hoshen Yedid. Panda: adapting pretrained features for anomaly detection and segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2021. p. 2806–14.

[34] Muller Steve, Lancrenon Jean, Harpes Carlo, Traon Yves Le, Gombault Sylvain, Bonnin Jean-Marie. A training-resistant anomaly detection system. Computers & Security 2018;76:1–11.

[35] Campello Ricardo JGB, Moulavi Davoud, Zimek Arthur, Sander Jörg. Hierarchical density estimates for data clustering, visualization, and outlier detection. ACM Transactions on Knowledge Discovery from Data (TKDD) 2015;10(1):1–51.

[36] Abdullah Johari, Chanderan Navein. Hierarchical density-based clustering of malware behaviour. Journal of Telecommunication, Electronic and Computer Engineering (JTEC) 2017;9(2-10):159–64.

[37] Xia Yan, Pan Shuguo, Meng Xiaolin, Gao Wang, Ye Fei, Zhao Qing, Zhao Xingwang. Anomaly detection for urban vehicle gnss observation with a hybrid machine learning system. Remote Sensing 2020;12(6):971.

[38] Qin Kun, Wang Qixin, Lu Binbin, Sun Huabo, Shu Ping. Flight anomaly detection via a deep hybrid model. Aerospace 2022;9(6):329.

[39] Costa Paloma Casteleiro, Guang Zhe, Ledwig Patrick, Zhang Zhaobin, Neill Stewart, Olson Jeffrey J, Robles Francisco E. Towards in-vivo label-free detection of brain tumor margins with epi-illumination tomographic quantitative phase imaging. Biomed. Opt. Express Mar 2021;12(3):1621–34.

[40] Guang Zhe, Ledwig Patrick, Costa Paloma Casteleiro, Filan Caroline, Robles Francisco E. Optimization of a flexible fiber-optic probe for epi-mode quantitative phase imaging. Opt. Express May 2022;30(11):17713–29.

[41] Paszke Adam, Gross Sam, Massa Francisco, Lerer Adam, Bradbury James, Chanan Gregory, Killeen Trevor, Lin Zeming, Gimelshein Natalia, Antiga Luca, Desmaison Alban, Kopf Andreas, Yang Edward, DeVito Zachary, Raison Martin, Tejani Alykhan, Chilamkurthy Sasank, Steiner Benoit, Fang Lu, Bai Junjie, Chintala Soumith. PyTorch: an imperative style, high-performance deep learning library editors. In: Wallach H, Larochelle H, Beygelzimer A, d'Alché Buc F, Fox E, Garnett R, eds. Advances in neural information processing systems 32, Red Hook, NY: Curran Associates, Inc; 2019:8024–35.

[42] Eugene Khvedchenya. Pytorch toolbelt. https://github.com/BloodAxe/pytorch-toolbelt; 2019. [accessed 09.27.22].

[43] Milletari Fausto, Navab Nassir, Ahmadi Seyed-Ahmad. V-net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 fourth international conference on 3D vision (3DV), Piscataway, NJ: IEEE; 2016:565–71.

[44] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101, 2017.

[45] Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts. arXiv preprint arXiv:1608.03983, 2016.

[46] Campello Ricardo JGB, Moulavi Davoud, Sander Jörg. Density-based clustering based on hierarchical density estimates. In: Pacific-Asia conference on knowledge discovery and data mining, New York: Springer; 2013:160–72.

[47] Ester Martin, Kriegel Hans-Peter, Sander Jörg, Xu Xiaowei, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of 2nd international conference on knowledge discovery and data mining (KDD-96), 96; 1996. p. 226–31.

[48] McInnes Leland, Healy John, Astels Steve. hdbscan: hierarchical density based clustering. J. Open Source Softw. 2017;2(11):205.

[49] Dice Lee R. Measures of the amount of ecologic association between species. Ecology 1945;26(3):297–302.

[50] Pearson Karl. VII. Note on regression and inheritance in the case of two parents. Proc. Roy. Soc. Lond. 1895;58(347-352):240–2.

[51] Strudel Robin, Garcia Ricardo, Laptev Ivan, Schmid Cordelia. Segmenter: Transformer for semantic segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV); October 2021. p. 7262–72.

[52] Jadon Shruti. A survey of loss functions for semantic segmentation. In: 2020 IEEE conference on computational intelligence in bioinformatics and computational biology (CIBCB); 2020. p. 1–7.

[53] Yi Jihun, Yoon Sungroh. Patch svdd: Patch-level svdd for anomaly detection and segmentation. In: In: Proceedings of the Asian conference on computer vision; 2020.

[54] Tsai Chin-Chia, Wu Tsung-Hsuan, Lai Shang-Hong. Multi-scale patch-based representation learning for image anomaly detection and segmentation. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision (WACV); January 2022. p. 3992–4000.

[55] Ruff Lukas, Vandermeulen Robert, Goernitz Nico, Deecke Lucas, Siddiqui Shoaib Ahmed, Binder Alexander, Müller Emmanuel, Kloft Marius. Deep one-class classification. In: Dy Jennifer, Krause Andreas, eds. Proceedings of the 35th international conference on machine learning, volume 80 of Proceedings of Machine Learning Research; 2018. p. 4393–402.