Analyzing microscopy images is an extremely critical and challenging task. The process can be critical in the detection of diseases at the same time the material is complicated, causing high labour costs. This thesis delves into studying and solving several hurdles that are highly encountered in microscopic datasets, notably for cell and organoid image datasets: Accurately detecting, segmenting and classifying cell and organoid objects within the images using various deep learning methods; comparing various models on the detection of overlapping objects; improving segmentation of objects with a limited number of annotations; as well as incorporating the functionality into an end-to-end system. The main contributions of this thesis can be summarized as follows:

1- Self-supervised learning was employed and compared to the supervised approach for detecting organoid objects, under the condition that only a limited number of annotations are available (Part I, Chapter 3).

2- Various types and complexities of pretext tasks were developed, and their contribution to improving the segmentation on the main task was evaluated on the organoid images (Part I, Chapter 4).

3- Different U-Net models were developed and compared for the detection of overlapping organoid objects for a limited number of available images (Part II, Chapter 5).

Finally, two applied projects were executed, requiring the development of a publically accessible e-Science website:

4- An end-to-end system was developed to automatically detect and analyze, more specifically localize and quantify, organoid objects in microscopic images (Part III, Chapter 6).

5- Detecting and classifying cell objects was investigated using MASK R-CNN and YOLOv4 on large-size microscopic images, and the models were integrated into the end-to-end system (Part III, Chapter 7).
8.1 Answers to the Research Questions

In this section, we briefly recall the research questions that are presented in Chapter 1 while providing answers to them.

RQ1 [Self-Supervised vs. Supervised] How can self-supervised training contribute to limited annotated biological data? (Part I, Chapter 3)

It is important to find new methods to automatically detect and analyze tiny objects in microscopy images while alleviating the manual control requirements as much as possible. To tackle these issues, self-supervised learning (SSL) was employed. SSL is an approach which allows to process and extract relevant information from datasets consisting of a higher proportion of unlabeled images in comparison to the proportion of annotated images. The SSL approach is divided into two stages: (1) training models on pretext tasks and (2) training models on the main task. In the first stage, pseudo annotations were created by applying different augmentation techniques to the original images. Augmenting images normally creates a distorted version of the original images. In the pretext task, models were trained to reconstruct distorted images. As a result, half-way-trained models were obtained that are able to learn the representation of the features of the objects. Those models can be further trained on the limited amount of annotated data on the main task. It was found that models trained via SSL surprisingly outperform models trained via supervised learning, even when a limited number of annotations (114 images) was used from the total available annotated training data. Five augmentation techniques (Gaussian blur, Sobel and three-pixel drops) were used for the pretext task. Dropping 25% of the pixels to black, and training the models to reconstruct the distorted images showed to be a task that positively contributed to the main task performance in many conditions. Combining the $L_1$ loss with the SSIM loss boosted the performance on the pretext task and also lead to the best performances for the main task after secondary training on it. For training the main task itself, the $I_{0U}$ loss performed best. It appears that the F1-scores do not differ too strongly when comparing the frozen and non-frozen encoders. This means that it is not necessary to adapt the encoder part of the U-Net to new tasks, if the latter are comparable. A second implication is that a well-trained self-supervised U-Net encoder can act as a ‘foundation model’ for other similar tasks in the problem domain, in our case microscopic image segmentation.

RQ2 [Complexity of Pretext Tasks] How complex should a pretext task for the self-supervised training learning be? (Part I, Chapter 4)
In a recent survey [195], the SSL methods and their applications in medical imaging analysis were presented and discussed, which indicates that SSL currently is a hot topic in this domain. It was also explained that medical images are different from natural images, especially for the pretext task. For instance, detecting a human face should be possible under various lighting and intensities, while intensity differences provide essential information in medical imaging; for instance, the Hounsfield scale in CT scans assigns distinct values to various tissues. This opens the door to delve more into the SSL topic, more specifically the effectiveness of the pretext task on the main task. For studying the effectiveness of the pretext task on the main task, it is essential to compare various pretext tasks under the same training conditions, model architecture as well as input-output structure. Here, the organoid image dataset and U-Net architecture were used. The input of the U-Net is the corrupted image by one of the pretext tasks, whereas the target output is the original raw image. Various simple pretext tasks were defined, and their contribution to reconstructing the original image, as well as their contribution to the main task, which is accurately segmenting organoid images, was inspected. In total, eight simple pretext tasks were developed. Each task randomly distorts part of the image. This includes dropping pixels within a certain box in the image, blurring certain areas, rotating randomly selected circle areas, or shuffling randomly selected box areas. In order to answer the current research question, complex pretext tasks were designed to be combinations of two simple tasks. In total, 28 complex tasks were evaluated. It was shown that all pretext tasks, both simple and complex, were able to reconstruct the original image with an F1-score at the pixel level of at least 0.9. This is true for the different amounts of unlabeled data that were used to train the pretext task to different model variants. However, looking at the results, it is seen that the simple tasks provided a more stable, less risky training process. This risk was less apparent in complex tasks that include shuffling parts of the image, yielding trustworthy results on the main task. Moreover, most of the models trained on the main task (segmentation of organoids) obtained an F1-score higher than 0.83. At first sight, it still appears to be difficult to characterize pretext tasks as having the potential to be useful. However, it is found that it is possible to train a task space embedding which can be visualized (t-SNE). From these analyses, it appeared that ‘good’ tasks are characterized by a localized density, whereas less successful tasks often showed an elongated shape, indicative of the presence of an important but unknown latent task variable. This shape difference could be used for the selection of pretext tasks in other application domains. However, these are early results: It is clear that more research is necessary on the design of pretext tasks and their characterization, which would allow for a reliable prediction of the performance effect on the main task(s).
RQ3 [Segmenting Overlapping Objects] How to segmented tiny and overlapped organoid object? (Part II, Chapter 5)

Being able to accurately detect and segment overlapping organoid objects is necessary for the accurate computation of biological parameters from these microscope images. As this topic is not well studied in the biological literature, some challenges should be addressed. Two new U-Net models were explored: (1) A simple U-Net with one encoder and a single decoder, yet with two output channel groups - the ‘mask channel’ and a new type of output that is called the ‘overlap channel’; and (2) a double U-Net architecture with a shared encoder and two decoders, each responsible for a segmentation, i.e., the mask or overlap channel. A fundamental problem in dealing with overlap is that there will always exist a notable imbalance between the number of samples for the different pixel classes. Unlike the situation of simple classification problems (e.g., MNIST), this imbalance cannot be easily solved in the case of (semantic) segmentation. The image is as it is, and selective cropping is an attempt to balance the number of overlapping, non-overlapping pixels as well as the ‘within-object’ and background pixels are doomed to fail. This issue was addressed by comparing the combination of various loss functions. The used loss functions were Cross Entropy, Dice Loss, Focal Loss and Focal Tversky Loss. It was demonstrated that using the combination of Focal Loss and Focal Tversky Loss, which down-weight the most frequent class, significantly improved the performance of the model. The simple U-Net achieves the highest scores with complex loss functions, while the double U-Net performs best with simple loss functions. As regards the prediction of the overlapped region shapes, it was noted that the models are able to detect the occurrence of the overlaps, per se, yet, without a fully satisfying reconstruction of the full shape of these overlapping regions. However, the high intensities of the overlap indicators in the output might be used in other ways, e.g., for ignoring those regions in the measurement of biological parameters only of clean, separable objects (organoids, cells, organelles). These results indicate that more research is needed, exploring and investigating new ways to segment overlapping objects in biological images and other application domains.

RQ4 [Automatization of Organoid Analysis] How can the segmentation of organoid images be automated using an end-to-end trained system? (Part III, Chapter 6)

The goal of this research is not only to investigate how deep-learning methods can be adjusted and applied to microscopy images but also to
offer an end-to-end application that can be used by biologists with less deep-learning knowledge to develop their research. Therefore, an e-Science service that automatically detects, segments, and analyses organoid images was developed via the processing pipeline of the Mask R-CNN deep learning method. This system allows researchers to upload their brightfield images in order to quantify the number of organoids present in the image as well as to measure their area. The system allows researchers to calculate the average area of multiple organoids of a culture or only of those of interest. Further, the system was tested on a use case to evaluate the area and growth analysis of liver progenitor organoids that have been grown in either a complete medium (referred to as control) or in a medium lacking all amino acids (referred to as starvation), which are essential for growth. Additionally, the data for the use case is publicly provided. The design of the system allows for detailed user interactions with the images and their annotations. For instance, users can adjust the segmentation of a specific instance of an organoid and select particular organoids of interest to be automatically analyzed. This will be beneficial, especially, when tracking individual instances of biological objects over time.

**RQ5** [From Segmentation to Detection and Classification] How to implement automatic detection and classification of cell compartments of fluorescent images in an end-to-end system? (Part III, Chapter 7)

With an end-to-end system in place, as an ‘e-Science server’, the question is whether it can also be expanded to handle different types of data and different types of tasks, with limited requirements for intensive user labour. With this intention, a fully automated end-to-end subsystem was developed to handle yeast-cell data, with fluorescence-tagged cells containing visible organelles in the image. The resulting system is capable of solving various segmentation, detection and classification tasks on fluorescent images. This was done by constructing a pipeline that uses Mask R-CNN to automatically segment and label a large amount of yeast cell data, and YOLOv4 to automatically detect and classify individual yeast cell compartments from these images. The system was evaluated on the NOP1pr-GFP-SWAT yeast-cell data library. Since the input images are large in size, a comparison between segmenting and classifying objects on a quarter of the images versus on the complete image was studied. Experimental results show that by dividing original images into four quadrants, YOLOv4 outputs good detection and classification results with an F1-score of 98% in terms of accuracy and speed, which is optimally suited for the native resolution of the microscope and current GPU memory sizes. To be mentioned that this study utilized the fourth version of YOLO. This year (2023) the eighth version of YOLO
was released. As reported in [196], although YOLOv8 has many advantages, there are still some problems with small-size targets, which is the core issue in this dataset.

8.2 Future Research Directions

This section proposes several possible research directions that are associated with the results of this thesis. Each paragraph demonstrates a direction.

(1) Extending to different types of data

The focus of this thesis is on biological microscopy images. Compared to natural images, microscopic datasets suffer from various issues, including the limited amount of data and even the limited amount of annotated data. Since, currently, images of natural and man-made scenes (photographs) are available in vast amounts, including extensive labelling, it would be beneficial to develop AI methods and test them first on such data by artificially limiting the amount of annotation in training. In this manner, solutions to the sparse-labeling problem can be studied systematically. Also, for the overlapping objects scenario, existing benchmark image collections could be used. This is a bit of a large challenge because the occlusion is often not apparent from the existing labelling. In any case, the collected knowledge from such large-scale experiments can be transferred to the domain of sparsely-labelled microscopic image datasets, potentially also containing overlap problems.

(2) Loss functions

The results of our experiments show that the choice of the loss function is essential and is related to the actual desired response in the application domain. Many neural-network architectures, such as the U-Net, have crystallized into proven solutions and less handcrafted design is needed, at this level. However, the choice and design of optimal loss function remains an essential step in the development process. For example, for pixel-based matching, Intersection-over-Union (IoU or Jaccard) performed best in the secondary training of the main task, in the self-supervised approach. Conversely, when training the primary pretext task, the Structural Similarity Index Metric with the L1 (SSIM-L1) loss is preferred. In [197] an unsupervised loss function is proposed that specifically regularizes the network based on the variations caused by randomized data augmentation, dropout and randomized max-pooling scheme. This loss function suits the semi-supervised and self-supervised approaches especially when significantly more unlabeled data are available in comparison to the amount of labeled data. For
complicated scenarios, it is currently common to join different loss terms, as in generative adversarial networks, to fulfil the desired target output quality. Another example of ‘loss-function engineering’ is the triplet loss [198] that we have recently adapted to segmentation tasks [199]. The triplet loss encourages dissimilar pairs to be distant from any similar pairs by at least a certain margin value. Overall, more research is needed in this area.

(3) Design of effective pretext tasks

The topic of the pretext task remains extremely complex and challenging. It was found, for example, that a simple global blur followed by reconstruction is not the most effective pretext task. On the other hand, random shuffling of patches yielded stable results. It is still of interest to delve into the role of the pretext task in improving the detection and segmentation of objects for largely unlabeled datasets. The defined tasks in Chapter 4 come from the same basic problem domain, i.e., the analysis of a particular type of organoid images. A limited subset of task types in the range of local to global perturbations was explored. Therefore, no general decision can be made as yet about what type of task performs best, or how complex a pretext task should be. Future studies could fruitfully explore this issue further by defining different tasks that are derived from a different basis. For example, dividing the image into parts, then shuffling those parts, and asking the model to reorder those parts with the aim of attaining the original image. Here, the input and output of the model remain images. The latter does not always have to be the case: For some problems, artificial classifications could be used as the pretext tasks. This direction has some overlap with the research in multi-task learning [200].

(4) Overlapping objects

Although the problem of occlusion is well known in general computer vision [36], [37], the topic of overlapping objects in microscopy has not been studied extensively in medical and biological literature. Since the size, area or color of the microscopic objects can be used as an indication of the presence of disease, it is essential to accurately detect all of them and analyze their morphology. This is only possible when the input data is correctly labelled. Moreover, the approach should be compared to other datasets containing natural objects, such as ImageNet deng2009imagenet or MS-COCO [120]. New possibilities will emerge when the so-called ‘z-stack’, a number of images along the depth axis, will be included to solve the overlap problem using deep learning.

(5) Extending the website

The e-Science service [OrganelX] appeared to be interesting to end users. It
can be extended to include extra functionalities. For example, the current version of the website detects organoid objects when a certain .CZI file is uploaded, as an addition would be to enable tracking objects upon uploading multiple CZI files. Furthermore, the website can be extended to include different file formats, such as .png and .tiff, instead of being restricted to .CZI file format. Another extension would be the ability to handle different types of data, not only the organoid and fasta\(^1\), and different types of deep learning tasks.

(6) Limitations of confocal microscopy

At the start of this study, a lot was expected from the ability of deep learning to classify cell compartments (organelles). It was shown that in the case of fluorescent tagging, interesting classification results could be reached. However, the power of current deep learning is not in the area of pixel intensities, but in the domain of shape and texture classification. Due to the physical limitation of the wavelengths of the used light, the internals of the cells were a bit blurred such that fine details could not be exploited by the CNN filtering kernels. It is expected that the methods used in the current study will have interesting applications in STED (Stimulated emission depletion) and EM (Electron microscope) microscopy, where the resolution of light-fluorescence microscopy is approximately 300 nm, the resolution of EM is approximately 5 nm, and the resolution of STED is approximately 40 nm, which is higher than in light/fluorescence microscopy.

---

\(^1\) [https://www.bioinformatics.nl/tools/crab.fasta.html](https://www.bioinformatics.nl/tools/crab.fasta.html)
This thesis deals with multiple significant issues in analyzing microscopic images. It covers topics related to detecting, segmenting, classifying and automatically analyzing tiny objects found in such image datasets. Most traditional deep learning methods require the data to be massively annotated in order to process and analyze it, using traditional supervised learning techniques. In microscopy, images can be large, the target objects are tiny and annotating each image is not only time-consuming but calls for expertise in the domain. In this study, an alternative solution is proposed: Self-supervised learning. In this case, artificial pretext tasks are used as a replacement for class-labeled annotation. The main task in this case was the segmentation of organoid objects. The most striking finding is that the performance of deep-learning methods on this task can even be better on the basis of pretext tasks, than on the basis of the labeled dataset alone. There appears to be a break-even point with a proper mix of labelled instances and a large set of pretext training instances. Possible reasons for this are the ‘big data’ effect of the pretext tasks on the one hand and the problems in the quality of pixel-level labelling for the main task on the other hand. The self-supervised approach may act as a safeguard against overfitting. A topic that is often encountered in microscopy data is the difficulty of detecting the overlapping objects. This thesis proposes different approaches to investigate this topic and to allow users, e.g., to filter out non-overlapping objects or, alternatively, zoom in on the overlaps. An end-to-end system was developed which offers the opportunity for non-deep learning experts to analyze their data using deep learning techniques with limited on-line interaction and an acceptable amount of computation time.

Studying biological and medical data plays a significant role in discovering diseases, and thus is crucial for the advancements in daily life environment. Since biological data is highly complex when compared with other forms of data, it is important to develop and enrich the research in order to contribute to this domain. According to [201], the segment-anything model (SAM), introduced in April 2023, is claimed to show promise as a benchmark model and a universal solution to segment various natural images. It comes without previously-required re-training or fine-tuning specific to each new dataset. However, when used directly on medical images without re-training or fine-tuning, SAM is not as accurate as algorithms expressly created for medical-image segmentation tasks. This example shows that deep-learning research most likely still will need to focus on specialized application domains, because the currently available ‘foundation models’ are not guaranteed to provide optimal results.