Deep learning solutions for domain-specific image segmentation

Vadineanu, S.

Chapter 6

Conclusion and Outlook

The main goal of the research presented in this thesis was to develop solutions that facilitate applying deep learning algorithms under annotation constraints, with applications for cell imaging and archaeological remote sensing. Throughout this research, we targeted the challenges present for the human operator in both the annotation and the training processes within the deep learning pipeline. To do so, we relaxed the quality requirements placed on the expert annotators, we proposed annotation-efficient learning paradigms, and we introduced explainability case studies. Here, we present a summary of the main contributions of this thesis, acknowledge its limitations, and propose future research directions.

6.1 Contributions and Limitations

In Chapter 2 we modelled three types of inconsistencies that can occur when creating annotations for cell segmentation. These inconsistencies can be considered both annotator-related errors or deliberate relaxations of the annotation process to allow for creating larger quantities of annotations within a fixed time budget. We considered the effect of the omission of a certain proportion of the target cells, the inclusion of objects other than the target cells, and the effect of inconsistent cell boundaries under the form of exaggerated or reduced boundary delineations (called bias). We performed gradual reductions in the annotation quality and we tested their effect on the training of three architecturally-dissimilar segmentation networks. Our results indicated that the networks were least affected by omissions, with inclusion and bias producing more severe degradation of the performance, especially when the cells have small footprints. These findings may allow human operators to optimize quality control efforts by focusing on the most impactful error types, thereby enhancing the robustness of models even when annotation resources are constrained. This contribution directly addresses one of the annotation process bottlenecks by enabling more strategic quality trade-offs, thus supporting the deployment of robust segmentation models with fewer high-quality annotated samples. One inherent limitation of our study is the scope in which it was performed. We considered three CNN architectures, three data sets, and three types of annotation errors, which can limit the broadness of our conclusions.

Based on the findings from Chapter 2, we proposed in Chapter 3 a method that can enhance the quality of annotations suffering from various types of inconsistencies. Our main contribution is designing a learning pipeline for cell segmentation in which a small data set with high-quality annotations is leveraged to train a CNN to upgrade a low-quality annotation to a high-quality one. We achieve this by perturbing the high-quality annotations and tasking the CNN with retrieving their initial quality. We then use this upgrade CNN to enhance the quality of a larger set with low-quality annotations. We showed that by combining the initial small high-quality set with the larger set with upgraded annotations, we can train better-performing cell segmentation CNNs than on the high-quality set alone. This approach presents a practical solution for scaling annotated data in a cost-effective manner by reducing the need for extensive expert annotations. By enhancing lower-quality annotations through an automated upgrading process, our method increases the usable dataset size without a proportional increase in human effort, contributing to the applicability of deep learning in scientific imaging under constrained annotation budgets. This contribution aligns with the thesis's aim to develop annotation-efficient solutions that support the training of deep learning networks in data-limited domains. The main limitation of our approach is the necessity of engineering perturbations that replicate the errors expected in the low-quality data set. Although we showed that the match between perturbations and errors does not have to be perfect, designing them is still an additional cost in the development of the model.

The main contribution of Chapter 4 is creating a few-shot technique particularly suited for cell segmentation. This method fits into the general few-shot learning pipeline while accounting for requirements specific to cell segmentation. We leverage the high-resolution feature maps produced by MSD networks trained on the known cell classes, which we then linearly recombine to adapt to the new class of cells. We demonstrated that the few-shot learning paradigm can be effectively applied to cell images, with our method surpassing other techniques designed for natural images or medical image segmentation. Here, we targeted the training process of the deep learning pipeline and proposed a solution that reduces the reliance of the human operator on large quantities of annotations. The few-shot learning technique we developed reduced the number of labelled samples required for training new classes in the context of cell segmentation, enabling future experimentation with data sets previously unsuitable for learning setups. One important limitation of our method is its one-shot performance. When trained with a single image, the results vary based on how representative that image is for the rest of the data set. This variation, however, becomes significantly lower when using 5 or 10 shots. Another limitation is the requirement to have sufficient annotations for some of the cell types within a given data set in order to train the MSD networks. This limits the applicability of our work to scenarios in which one wants to segment new structures from an already annotated data set.

Chapter 5 contributes with an annotation process for archaeological site segmentation starting from image-level annotations. We use binary annotations for image classification, i.e., whether a site is present within an image, to train classification models from which we employ explainability techniques to extract activation maps that we further process to obtain site boundaries. In addition to producing cheap annotations for segmentation, we leverage the resulting maps to perform an analysis of the learned features by three CNNs, which can contribute to a better understanding of the CNNs operation and, consequently, to the wider adoption of these techniques in archaeological works. Moreover, we present a modification to an existing explainability technique which produces site boundaries close to the expert estimation. We observed differences in the image features that different architectures tend to highlight and we also showcased the explainability techniques' potential of highlighting biases or mis-annotated images. This contribution tackles problems in both the annotation and training processes and offers a dual benefit by providing a low-cost method for segmentation annotation in archaeology while simultaneously enhancing model interpretability. On one hand, alleviating the annotation scarcity facilitates the practical introduction of deep learning in archaeological workflows as the human operator does not need to focus much on producing annotations. On the other hand, the capacity to derive meaningful visual explanations from CNNs facilitates a greater understanding of model behaviour, which can build trust in the predictions of deep learning networks. This is important for interdisciplinary applications with non-technical fields where model transparency and interpretability are valued. The geographical area to which the study was applied constitutes the main limitation of this chapter. We did not apply our techniques to images belonging to other regions which could contain

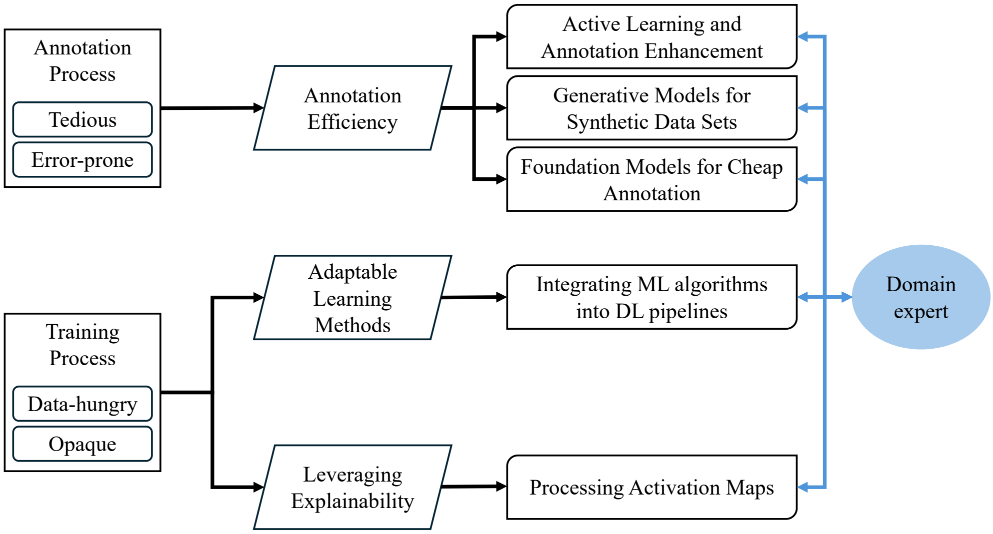

Figure 6.1: A schematic representation of future research directions emerging from this thesis, including enhanced annotation tools, adaptable learning methods and a focus on explainability techniques, highlighting the necessity of continuous collaboration with domain experts.

differently-looking sites or in which the landscape could pose more challenges in identifying a site image from a non-site one.

6.2 Outlook

Future research presents significant opportunities to streamline the advancement of deep learning within scientific fields where the demands of annotation and training processes pose challenges for human operators. In this section, we outline several potential directions for alleviating the costs associated with these processes. For a schematic representation of the envisioned directions, see Figure 6.1.

On the annotation side, one possibility emerges from combining the low-data requirements and short training time of few-shot learning with the enhancement of an upgradinig CNN into efficient annotation tools. Such tools can rely on an initial small set of manual annotations to train a few-shot model whose predictions can be further refined by an upgrading CNN. In this way, the expert can focus only on the most challenging samples, while the networks would also improve as more images are being annotated, in an active learning manner [107].

Alternatively, rather than focusing solely on increasing the throughput of the annotation process, this process may be accelerated by generating extensive synthetic segmentation datasets via generative networks, such as diffusion models, where input conditions (e.g., text queries, class labels) can be provided with significantly reduced human intervention compared to the generation of pixel-level segmentation annotations. Additionally, low-effort human input such as textual prompting, point, or box annotations can also be used to increase the number of available segmentation annotations by leveraging the predictions of foundation models, such as segment anything model (SAM)[83] or its more specialized variants, for instance, MedSAM [98].

On the training side, there remains a critical need for adaptable learning methods that can effectively exploit general image features derived from data sets with abundant annotations or even pseudo-annotations as in [112]. The key challenge lies in refining these generalized feature extractors to suit the specific requirements of the target domain. A viable solution involves embedding traditional machine learning algorithms within deep learning pipelines. This integration combines the feature extraction strength of deep learning with the efficiency and reduced data dependency of traditional machine learning methods. As demonstrated in Chapter 4, this hybrid approach has the potential to yield highly adaptable models, addressing the limitations posed by data scarcity in domain-specific applications.

When it comes to the detection of archaeological sites, the output of the explainability techniques (activation maps) can also be leveraged to derive more information about the sites than their boundaries. For example, by analysing their shape and distribution, activation maps can provide information about the morphology of archaeological sites without additional human input. This can then help in further clustering and categorization efforts.

Finally, one common theme that ties together the observations presented in this thesis is the need to strengthen collaboration between machine learning scientists and domain experts. Although scientific domains suffer from expensive data acquisition and annotation processes, these disadvantages can be mitigated by including expert knowledge directly into the development process of learning-based solutions. One way to do so is by introducing constraints based on prior knowledge. For example, in Chapter 5 we applied a Gaussian filter on site activations to generate accurate segmentation masks because we had the a priori knowledge that the area of interest contained round settlements. Thus, we were able to process the activations to better reflect this characteristic without the need of additional data or annotations. Similar approaches could also be applied in designing efficient annotation tools and accurate adaptable learning methods.