Contribution ID: **62** Type: **not specified**

ConText Transformer: Text-guided Instance Segmentation in Scientific Imaging

Scientific imaging gives rise to a multitude of different segmentation tasks, in many cases addressed with manually annotated datasets. We collected a large number of such heterogeneous datasets, consisting of over 10 million instance annotations, and demonstrate that in a multi-task setting, segmentation models at this scale cannot be trained effectively by only using image-based supervised learning. A major reason is that images of the same domain may be used to address different research questions, with varying annotation procedures and styles. For example, images of biological tissues may be evaluated for nuclei or cell bodies despite using the same staining. To overcome these challenges, we propose using simple text-based task descriptions to provide models the necessary context for solving a given objective. We introduce the ConText Transformer, which implements a dual-stream architecture, processing and fusing both image and text data. Based on the provided textual descriptions, the model learns to adapt its internal feature representations to effectively switch between segmenting different classes and annotation styles observed in the datasets. These descriptions can range from simple class names (e.g. "white blood cells")—prompting the model to only segment the referenced class—to more nuanced formulations such as toggling the use of overlapping segmentations in model predictions or segmenting a cell's nuclei during cell segmentation if the respective cell boundary is not visible, as it is common for example in the TissueNet dataset. Since interpreting these descriptions is part of the model training, it is also possible to define dedicated terms abbreviating very complex descriptions. ConText Transformer is designed for compatibility. It can be used with existing segmentation frameworks, including Contour Proposal Network (CPN) or Mask R-CNN. Our experiments on over 10 million instance annotations show that ConText Transformer models achieve competitive segmentation performance and outperform specialized models in several benchmarks; confirming that a single, unified model can effectively handle a wide spectrum of segmentation tasks; and eventually allowing to replace specialist models in scientific image segmentation.

Primary author: Mr UPSCHULTE, Eric (Forschungszentrum Jülich)

Co-authors: Prof. AMUNTS, Katrin (Forschungszentrum Jülich); Prof. DICKSCHEID, Timo (Forschungszentrum Jülich)

Session Classification: AI for Image Analysis Part 2