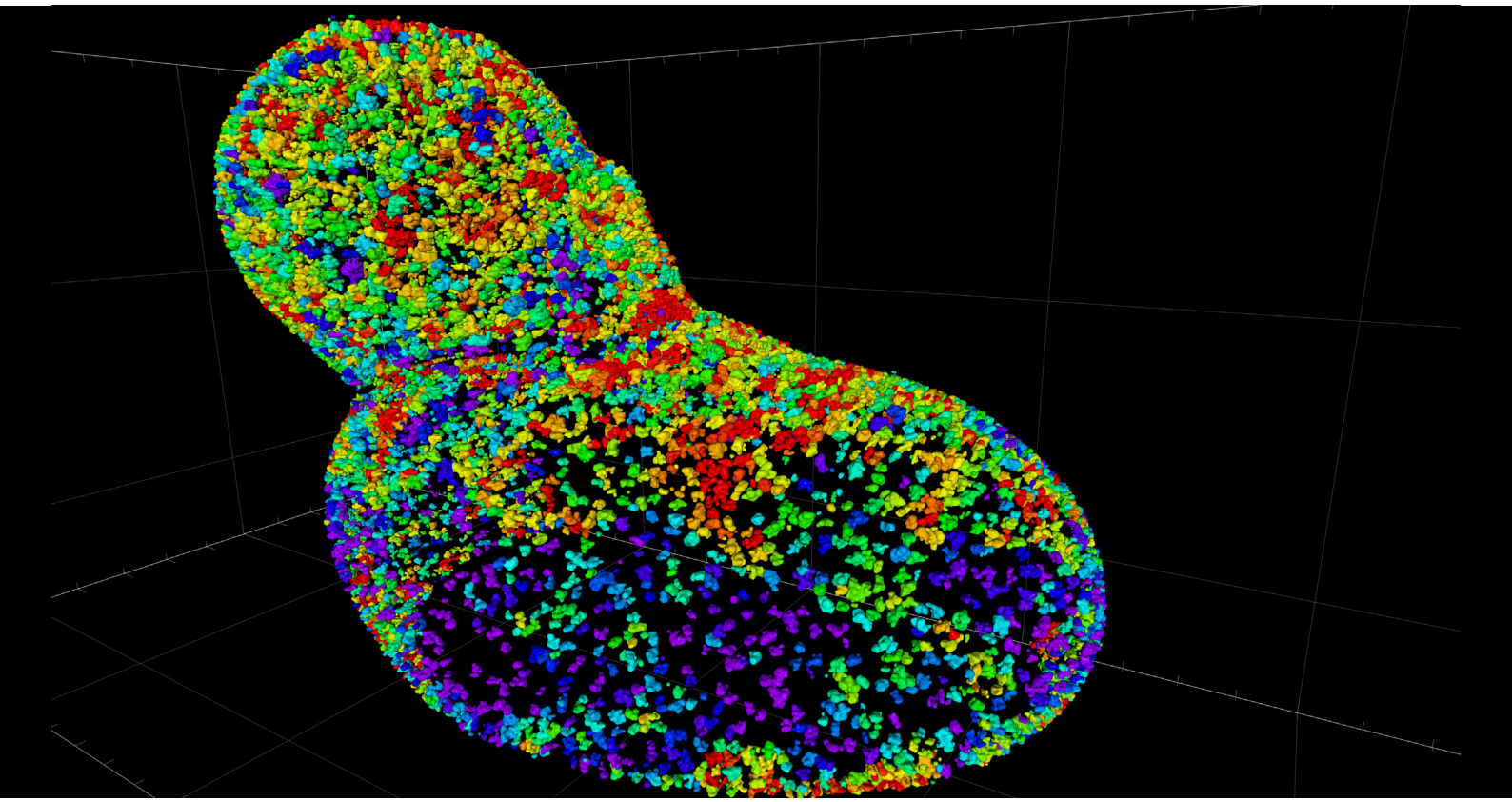
Application Note

How Artificial Intelligence is Reshaping Microscopy Image Analysis



ZEISS

Seeing beyond

Author: Dr. Sreenivas Bhattiprolu
Carl Zeiss Microscopy, LLC, USA

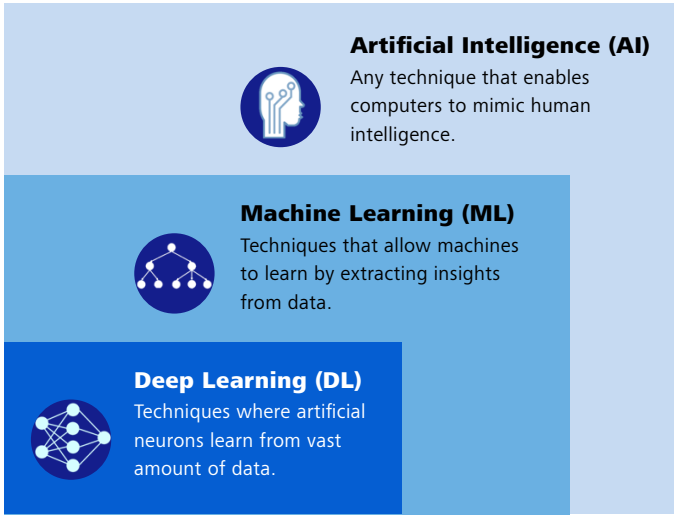
Date: February 2022

The COVID pandemic accelerated the digital transformation of organizations and even individuals. It increased our dependence on technology, primarily powered by AI. From background removal in zoom meetings to drug target identification in Biotech, AI services are rapidly becoming indispensable. Several Biotech companies have embraced AI as a much-needed solution to various applications including in the field of microscopy image analysis.

Modern automated microscopes offer the advantage of collecting large amount of data and hence statistically significant results. Unfortunately, due to its complexity, image analysis remains a bottleneck for many researchers, in their image to insights journey. Manual analysis is laborious and consumes researcher's valuable time, the time that could be put to good use on tasks that require human discretion. AI is the answer as it allows easy customization of solutions for specific bioimage analysis challenges.

AI refers to any technique that enables computers to mimic human intelligence. Machine learning is a subset of AI that includes techniques for machines to learn extracting insights from data. Deep learning is a subset of machine learning in which artificial neurons learn from vast amount of training data.

Traditional machine learning (non-deep learning) relies on engineered 'features' to minimize the complexity of the data presented to the algorithm. A feature is an independent variable that represents an attribute of the object of interest. For example, in image analysis, digital filters can be used to extract features that highlight various aspects of the objects of interest. Sobel filter (Figure 2) is an edge detecting feature extractor whereas the entropy filter (Figure 3) is a good example of feature extractor that highlights regions of different texture.



Features are provided as inputs to the machine learning system. Feature engineering process involves engineering appropriate feature extractors based on the application. This is usually done by the domain experts. For example, experts with a good knowledge of microscopy applications can design features for bioimage analysis. Traditional algorithms (e.g., Random Forest, SVM) typically work with a few tens of features making it possible for them to be used on regular image processing workstations. Image processing software with machine learning tools (e.g., ZEN Intellesis from ZEISS) pre-engineer the features for the

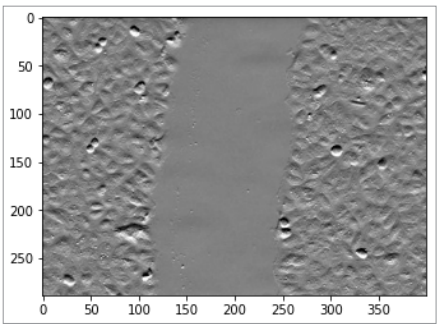


Figure 1 *Image showing scratch on a cell monolayer. Smooth region showing the scratch and textured region showing cells.*

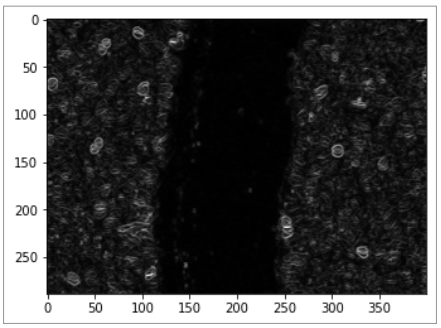


Figure 2 *Sobel filtered image highlighting the edges based on the intensity variation. This image can be used as one of the feature inputs to the ML algorithm.*

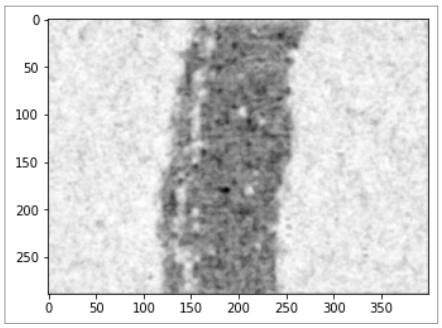


Figure 3 *Entropy filtered image highlighting the textured vs. non-textured regions. This image can be used as another feature input to the ML algorithm.*

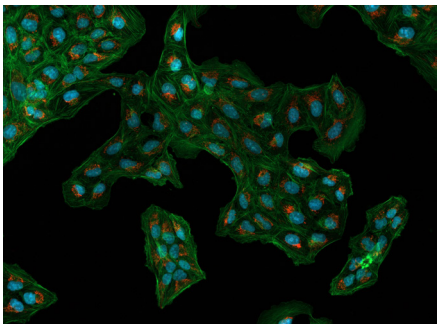


Figure 4 *Osteosarcoma cells showing DAPI-stained nuclei in blue.*

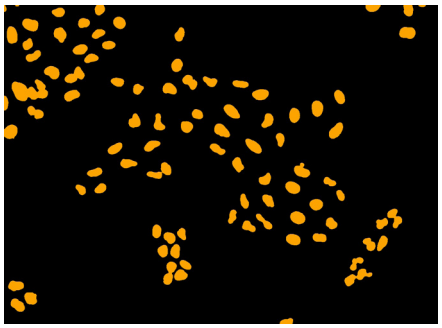


Figure 5 *Segmented nuclei using traditional machine learning (33 features + Random Forest classifier)*

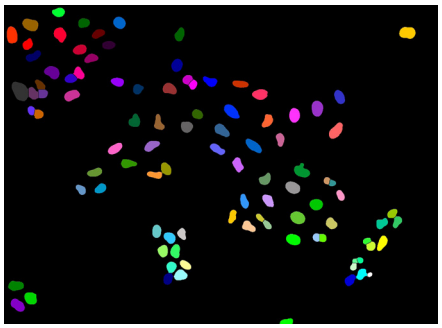


Figure 6 *Segmentation followed by watershed separation of individual nuclei.*

convenience of its users. This makes the technology accessible to any researcher and allows automated training of machine learning algorithms in an iterative way via visual feedback after each iteration. A bulk of biological microscope images can be segmented using traditional machine learning. For example, the researcher can annotate a few nuclei in a confocal image, train an ML model to segment the nuclei (Figure 5), and apply image processing tools such as watershed to separate the touching nuclei (Figure 6).

The traditional machine learning algorithms start to fail as the complexity of objects in images increases. A handful of engineered features are not enough to represent the complexity of glomeruli in renal pathology images (Figure 7) or high-density cellular structures in confocal images (Figure 8). That's where deep learning holds more promise.

Deep learning involves tuning of millions of parameters in its hidden layer structure to learn features incrementally, from simple lines to shapes to objects. It outperforms traditional machine learning but requires GPU acceleration for faster computation and a large amount of data to constrain the large number of parameters. Fortunately, techniques such as transfer learning enable 'fast' training of models using relatively small data sets reducing the training time from days to a few hours in most cases. Transfer learning refers to methods where pre-trained models trained on one task get repurposed on a second related task. Data scientists and ML engineers are well versed with such tricks and possess the required skills in preparing the data such that deep neural networks can be optimized efficiently. Biotechnologists typically do not have the skills nor the time to focus on such tasks. Their time is too valuable to perform manual tasks every time a data set needs segmentation. This is where APEER can offer help.

APEER is a cloud platform designed to automate image analysis workflows using the state-of-the-art deep learning algorithms. It's no-code user interface makes deep learning accessible and easy to use for Biotech researchers.

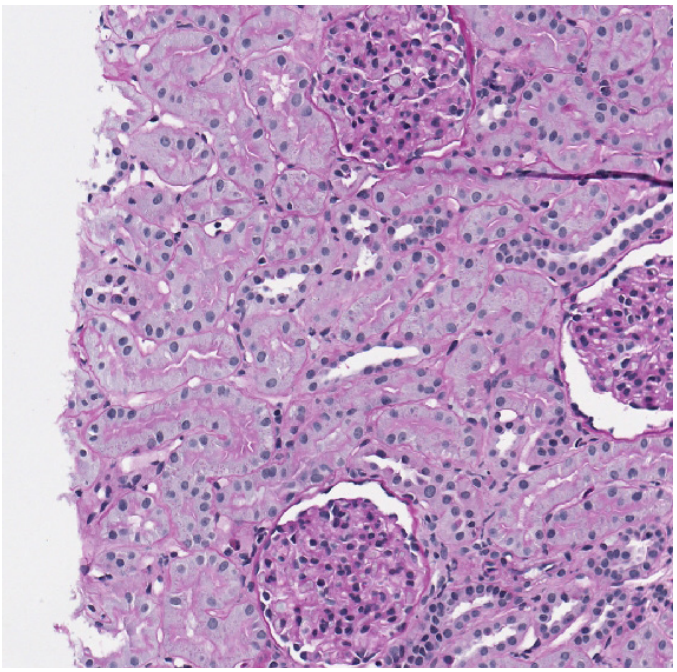


Figure 7 *Tissue sample stained using Periodic acid–Schiff (PAS) showing glomeruli.*

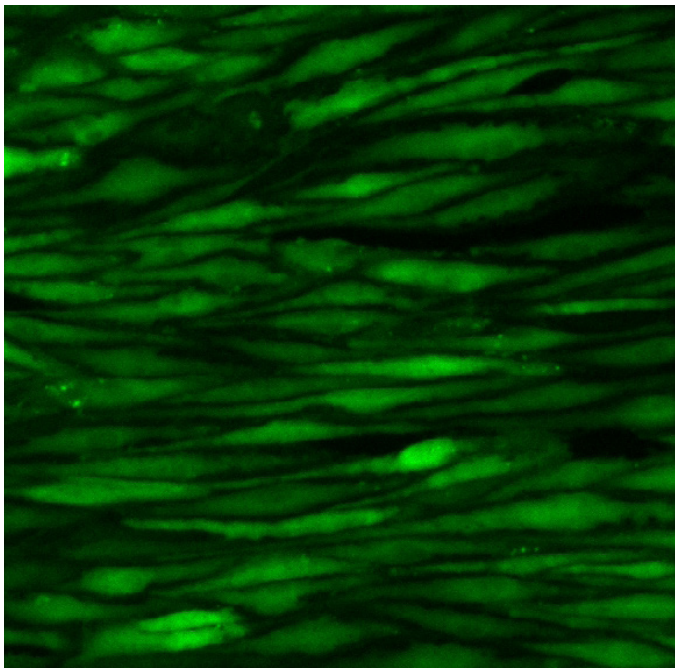
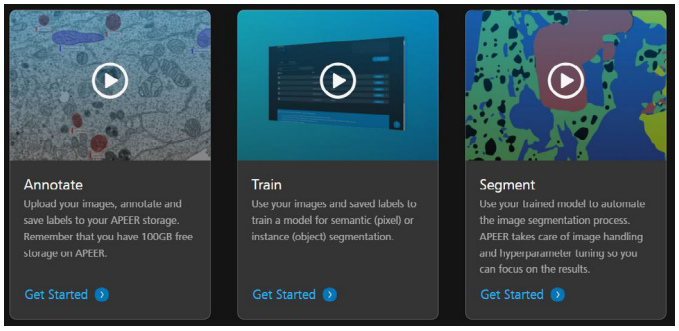


Figure 8 *Monolayer of human aortic endothelial cells transfected with GCaMP5G*

9 ways APEER makes it easy for a biotech researcher to segment images using deep learning

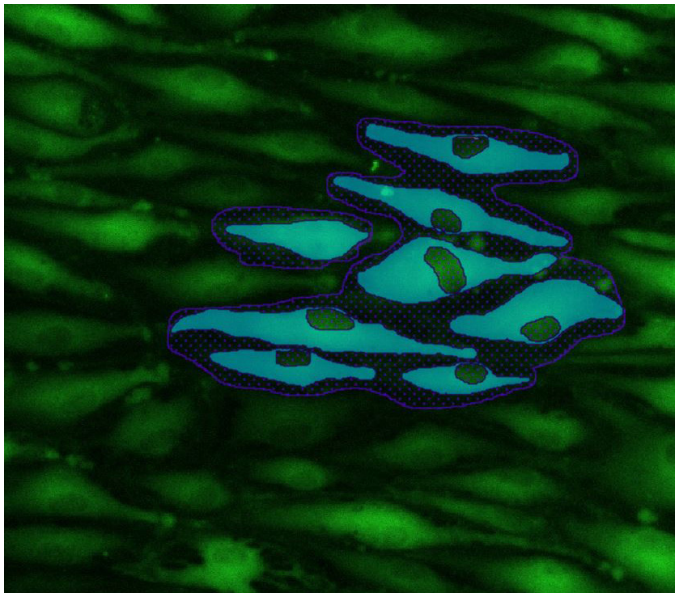
1. Walk through interface to annotate, train, and segment images

Apeer provides a walk-through interface for the biotech researcher to easily load training images, annotate them using a multidimensional web viewer, train the deep learning model, and segment future images using the trained model.



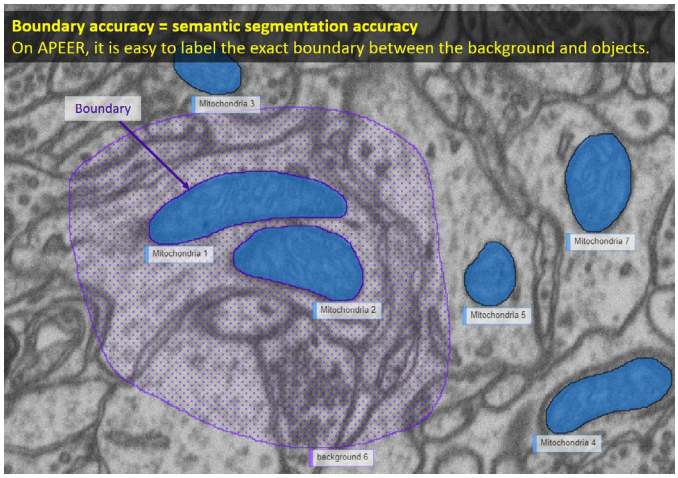
2. Partial annotations

Deep learning based semantic segmentation usually requires dense labeling where every pixel in every training image needs to be annotated. Dense labeling is laborious, it forces the researcher to annotate over-represented regions in every image. APEER introduced partial labeling for segmentation as part of its deep learning workflow. Partial labeling is efficient as it allows the researcher to focus on under-represented regions in several images. This is particularly relevant for biotech applications, where images are large.



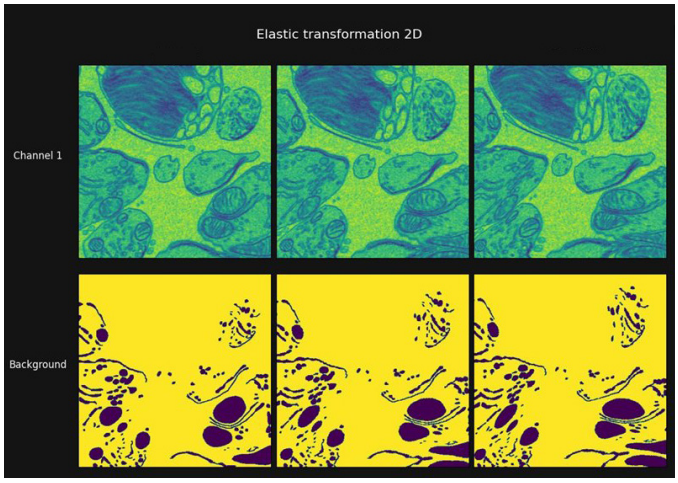
3. Automatic boundary labeling

In many applications, it is easier to segment the central pixels of objects compared to their boundaries. The boundaries between objects and background are usually associated with high uncertainty (low probability). Therefore, it is very important for the researcher to appropriately define the boundary pixels during the training process. APEER makes it easy for the researcher to define these boundaries precisely.



4. Data augmentation

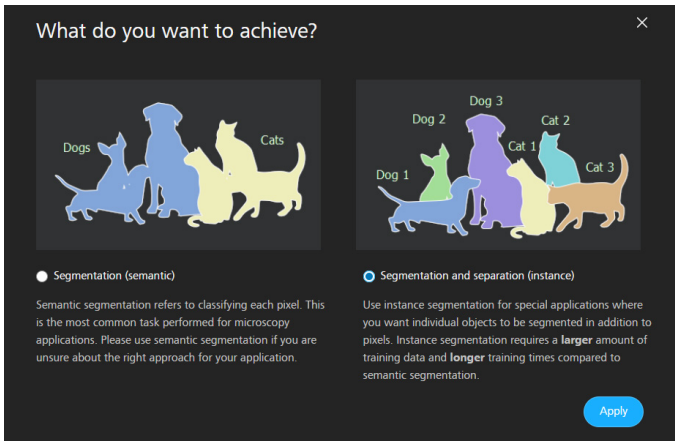
Data augmentation is a technique to generalize the trained model. This technique supplies the training data to the algorithm by performing various transformations including rotation, zooming, and stretching images. This allows generalization of the model and improves accuracy on future data whose representation the model may have seen via one of the transformations. APEER performs many tasks in the background and image augmentation is one such task.



5. Choice of semantic or instance segmentation.

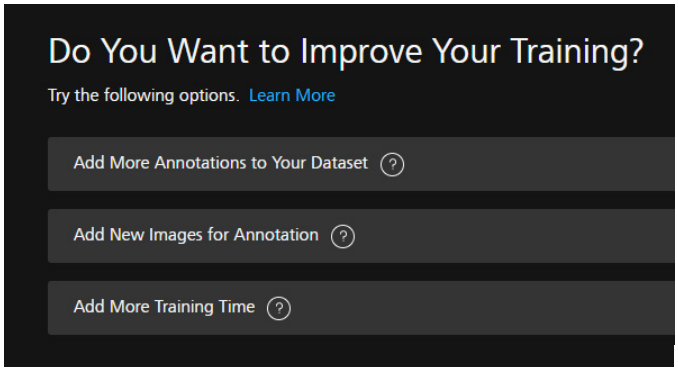
The researcher may want to choose the type of segmentation based on the application. For example, tissue classification requires semantic segmentation where every pixel gets classified into a specific tissue class. Whereas nuclei segmentation often involves object level classification (instance segmentation)

as morphological parameters need to be extracted for each segmented nucleus. APEER provides both options to allow the researcher to make the decision based on the application.



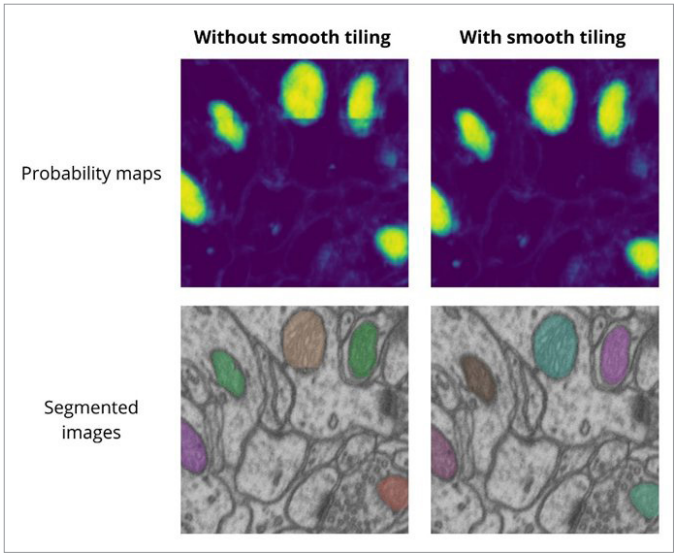
6. Iterative model development

APEER allows you to focus your annotation effort on regions where the algorithm learns the most from your manual effort. This is done by our iterative model development cycle: In a first step, the user provides a base set of partial annotations, then trains a first model. APEER shows the resulting segmentation across all provided images. From now on the user can focus their annotation effort on regions where the model is not yet performing as desired and then train again. This iterative workflow avoids manual work in regions where the algorithm is already performing well and saves valuable time.



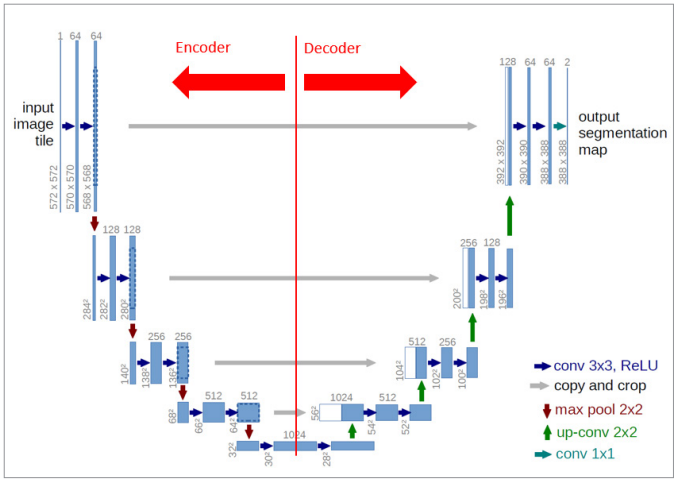
7. Smooth tiling

A deep neural network is only able to process a limited image size at once due to hardware restrictions. However, many bio-tech applications require the processing of much larger images. Therefore, it is customary to crop large images to process them one by one. The segmentation results of these cropped images are then assembled back to the original image. This process can create segmentation artifacts around the edges of cropped image tiles. To address this issue, APEER uses a smooth tiling approach, allowing cropped images to overlap and smoothly merging segmentation results in overlapping regions.



8. Algorithm designed for Bioimage analysis

The U-Net convolutional network architecture was developed for fast and precise segmentation of biomedical images. APEER uses the same architecture but modified using custom encoder and decoder steps that allow the model to be easily scaled with fewer trainable parameters providing higher accuracy.



9. Pretraining

Deep learning involves the training of millions of parameters. These parameters are usually initialized using random numbers and makes the training process time consuming. This is because, the model needs to learn about detecting lines, shapes, and objects from scratch. Pretraining the model on millions of images provides a better starting point for it to be adapted to segment other images. APEER uses pretrained weights to reduce the training time from days to hours, about 3 hours in most cases. Such quick segmentation feedback can be helpful to iteratively train the model by adding new annotations or images or extra training time.

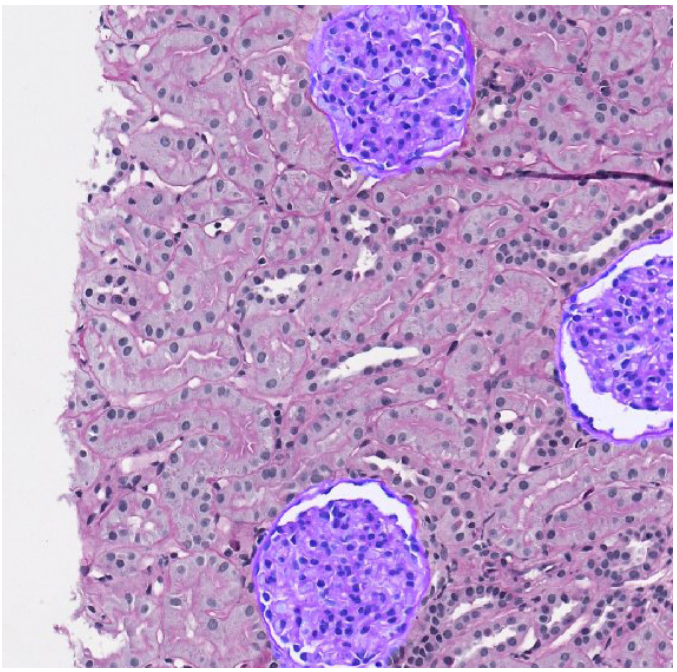


Figure 9 *Segmented glomeruli (blue overlay) in the tissue sample stained using Periodic acid–Schiff (PAS).*

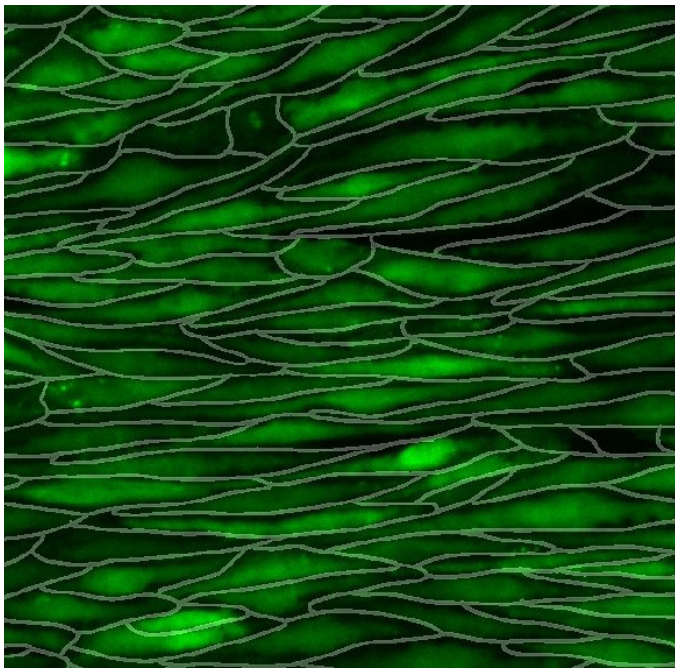


Figure 10 *Segmented endothelial cells from human aorta.*

In summary, AI enables the automation of end-to-end workflows by simplifying complex segmentation process which makes image analysis no longer the bottleneck in the image to insight journey. The volume of data analyzed per hour is far beyond what human researchers can achieve in days or weeks once a trained model can reproduce the experts image annotation. Such automation improves reproducibility and can help minimizing human bias in experiment results by aggregating the knowledge of multiple experts. APEER provides an easy-to-use interface to automate the image segmentation process using deep learning.

With AI, the Biotech industry is set to witness an increase in productivity and reproducibility with their microscopy image analysis workflows.

Cover image

Nuclear pore density analysis of HeLa cell volume EM imaging, provided by Yannick Schwab, EMBL. For each nuclear pore object, the average distance to the nearest 8 other nuclear pore objects was measured. Color-coding these objects according to this average distance to the nearest 8 objects was used to represent the density of pores across the nuclear membrane. Dataset provided by Anna Steyer and Yannick Schwab, EMBL

Carl Zeiss Microscopy GmbH
07745 Jena, Germany
microscopy@zeiss.com
www.zeiss.com/apeer