ETH zürich

Scientific IT Services, ETH

From Prototyping to Production: How SIS is Supporting AI in Research!

Franziska Oschmann
Data Scientist, SIS, ETH Zurich
30th of March 2022, AMLD

Scientific IT Services (SIS): Team and Mission

- **Section of ETH ITS** and a **pool of experts** (~45 members)

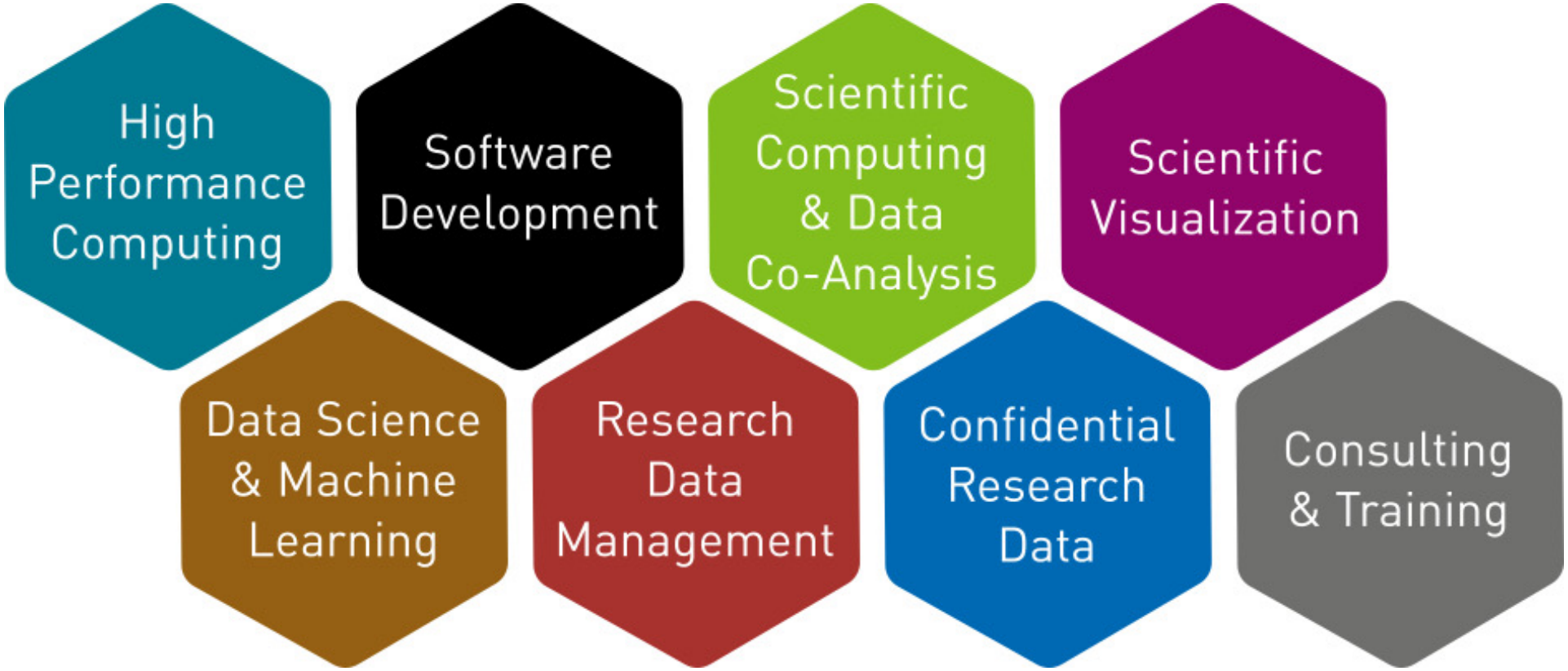
- **Main tasks:**
  - Enable research
  - Improve efficiency

- **Scientific computing experts & data scientists** from various scientific backgrounds
- **Scientific software developers** with computer science background and industry experience
- **System administrators & DevOps** for the HPC infrastructure and research IT platforms

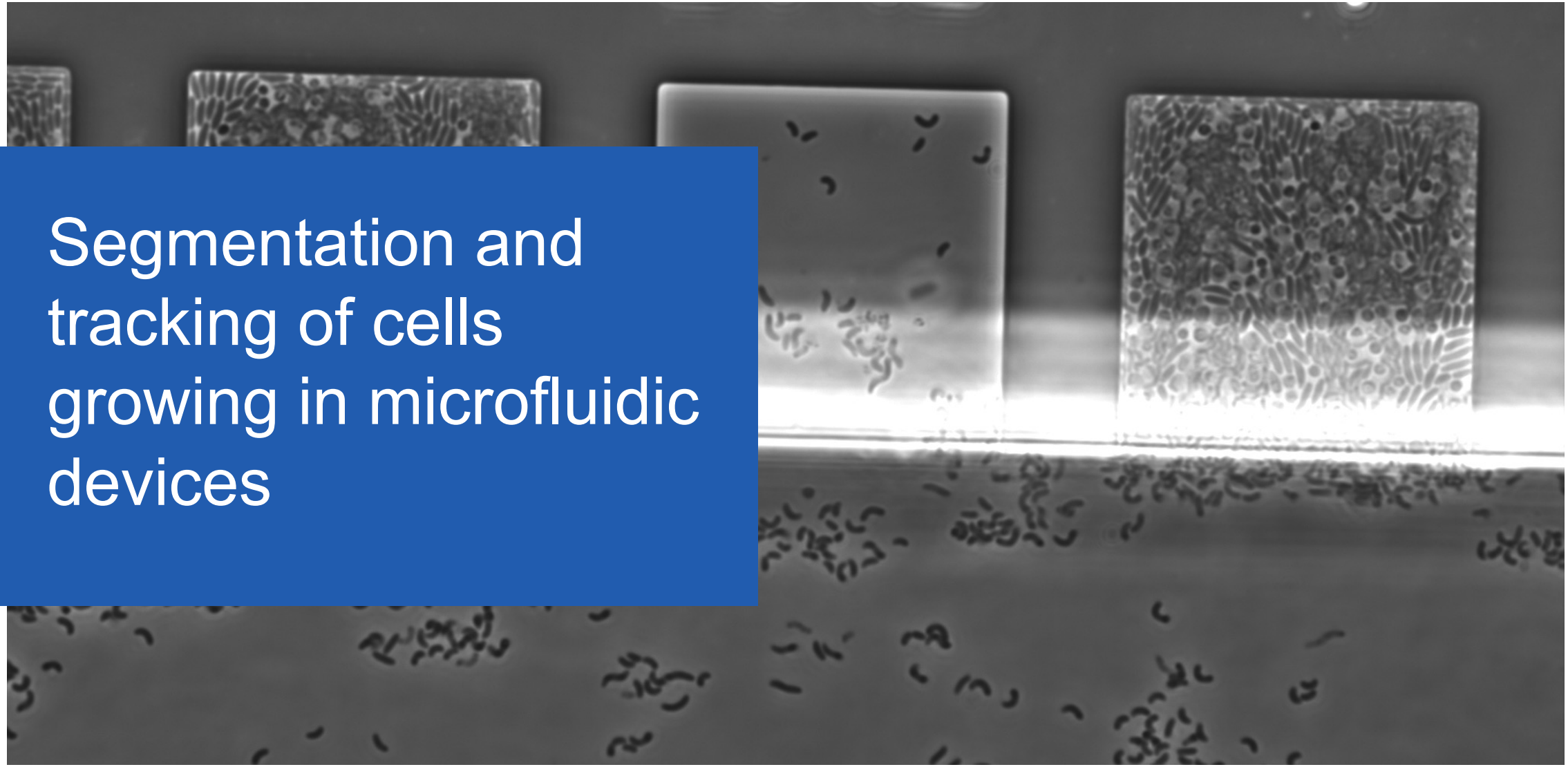
Working model

- Contracts with ETH researchers to the extent of:
  - a few days full-time support
  - one or several years part-time (20-50%) support

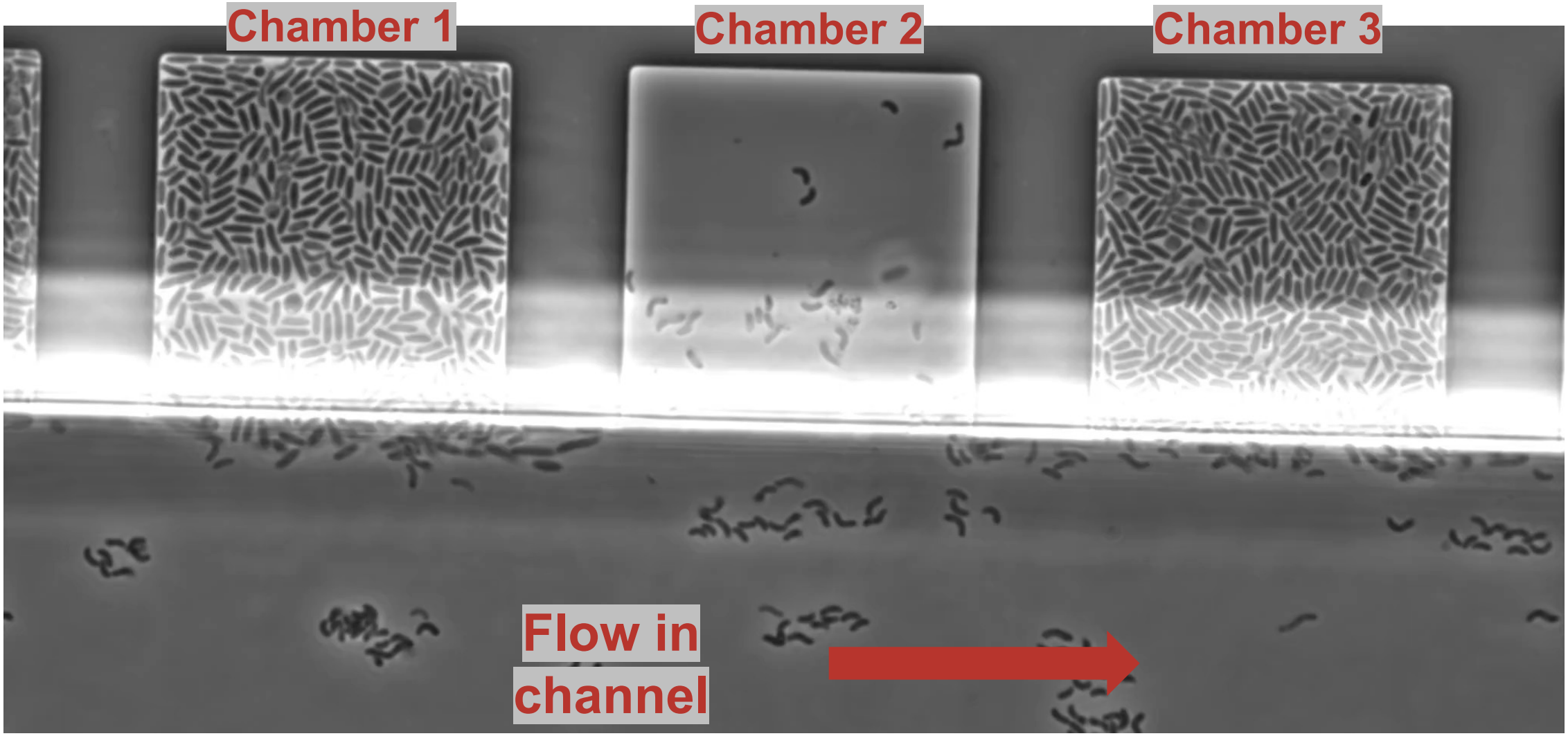
SIS Services



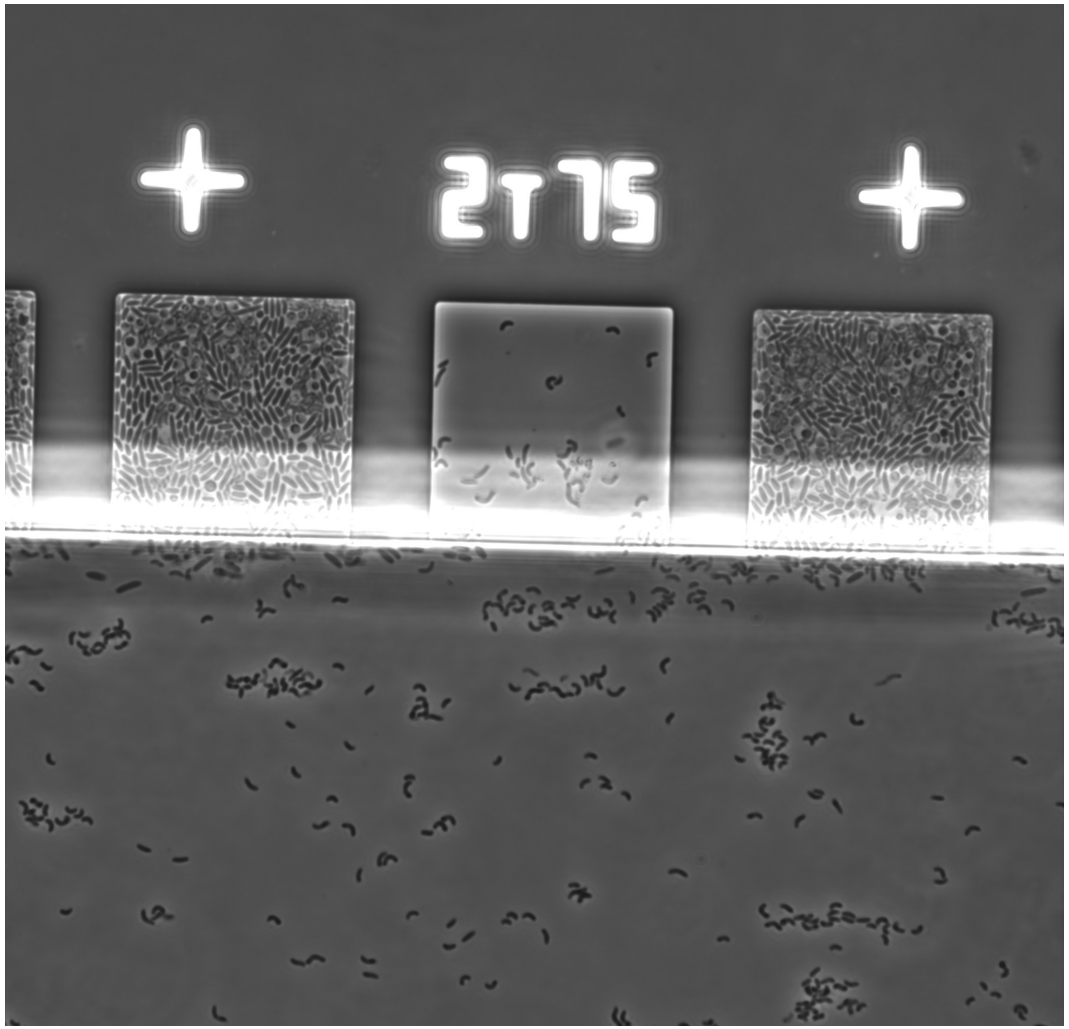
Segmentation and tracking of cells growing in microfluidic devices



Cell growth in microfluidic devices



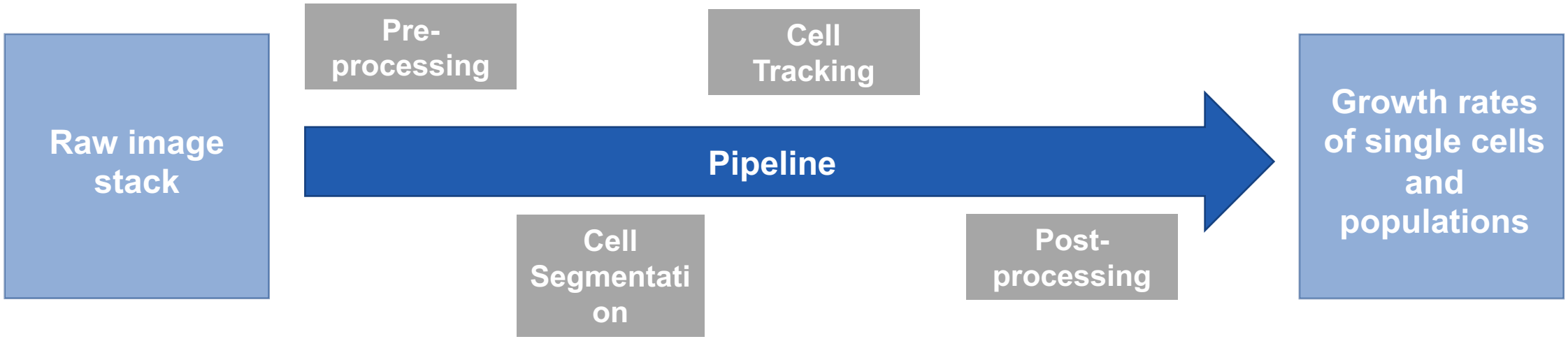
Quantification of population growth and dynamics



Required results:

- Number of cells
- Position of cells
- Relationship between cells
- Movement of cells
- Interaction between cell populations

Requirements for image analysis



Methods used before:

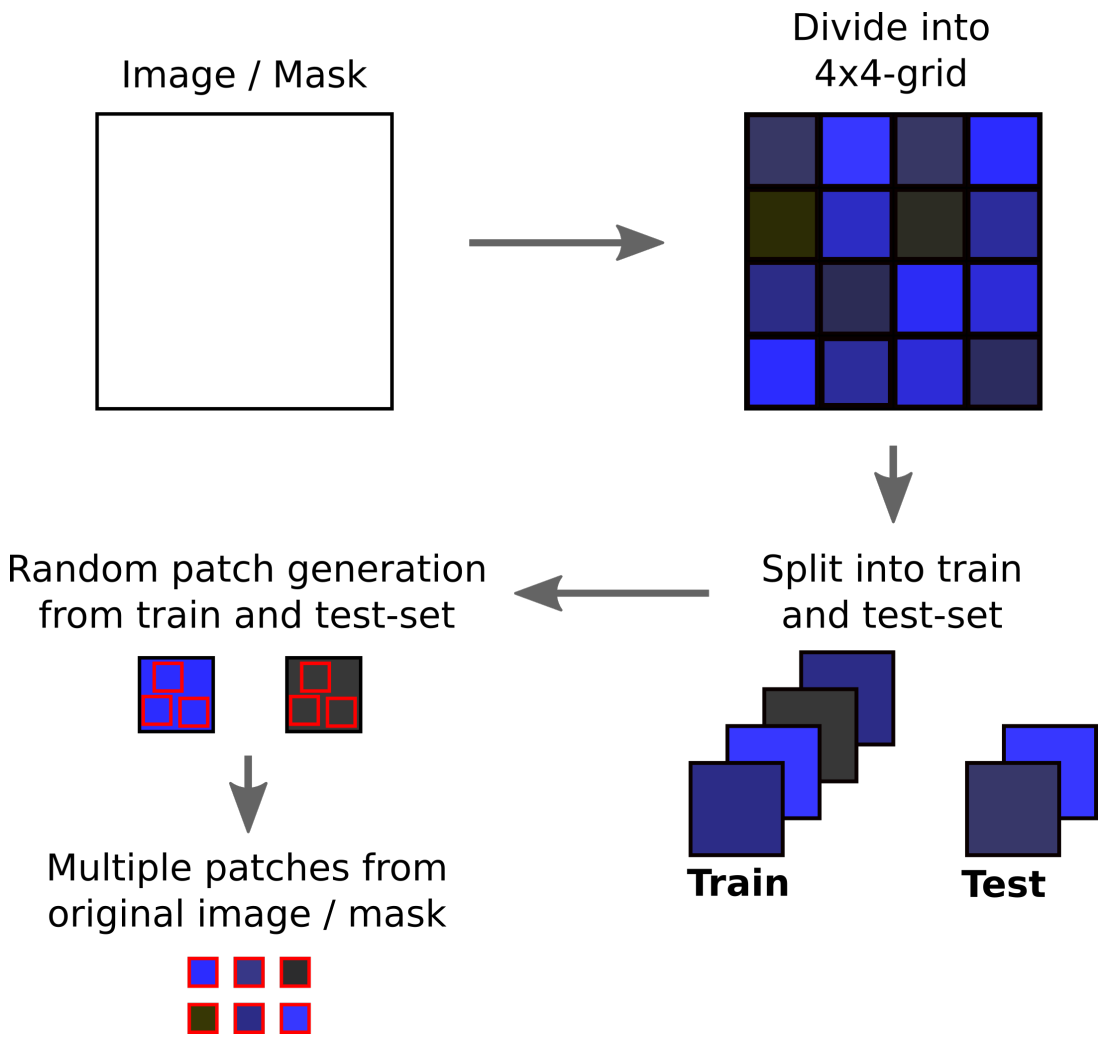
- State-of-the-art tools (ImageJ, Ilastik, SuperSegger etc.)

Problems:

- Methods don't work properly for their datasets
- No software development or Machine Learning expertise in research group
- Data analysis is bottleneck of research work

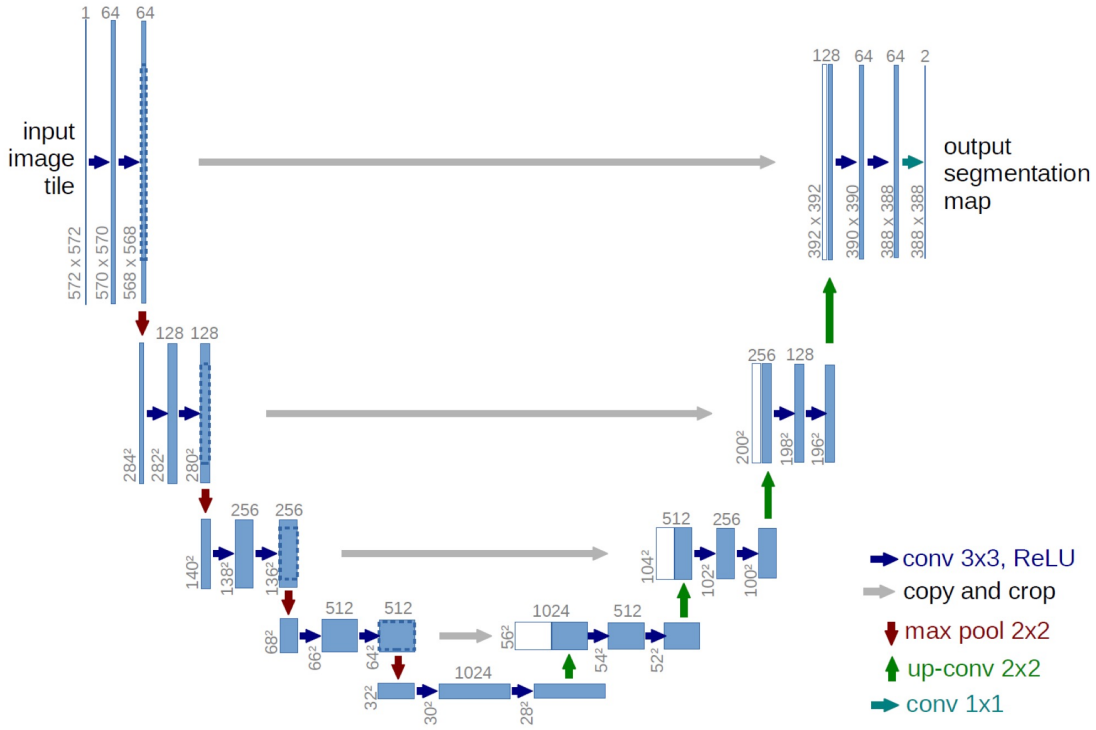
→ Contacted SIS for support

Dataset generation



- Genenration of training dataset with minimal manual effort
- Original image size: ~500 x 500 pixels
- Size of single cell: ~30 x 30 pixels
- → allows split of original image in multiple smaller patches
- Data augmentation of each pach
- Approach generates ~ 8000 training images based on one original image

Deep Neural Network for segmentation: U-Net



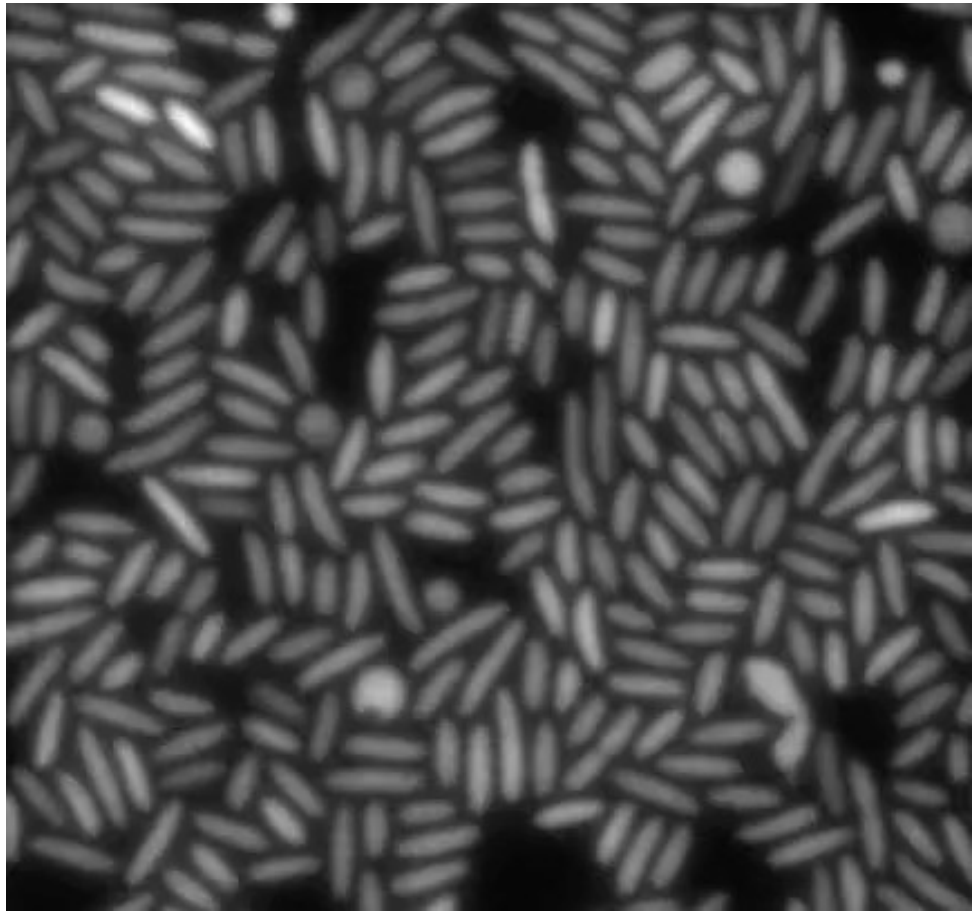
Ronneberger et al, MICCAI 2015

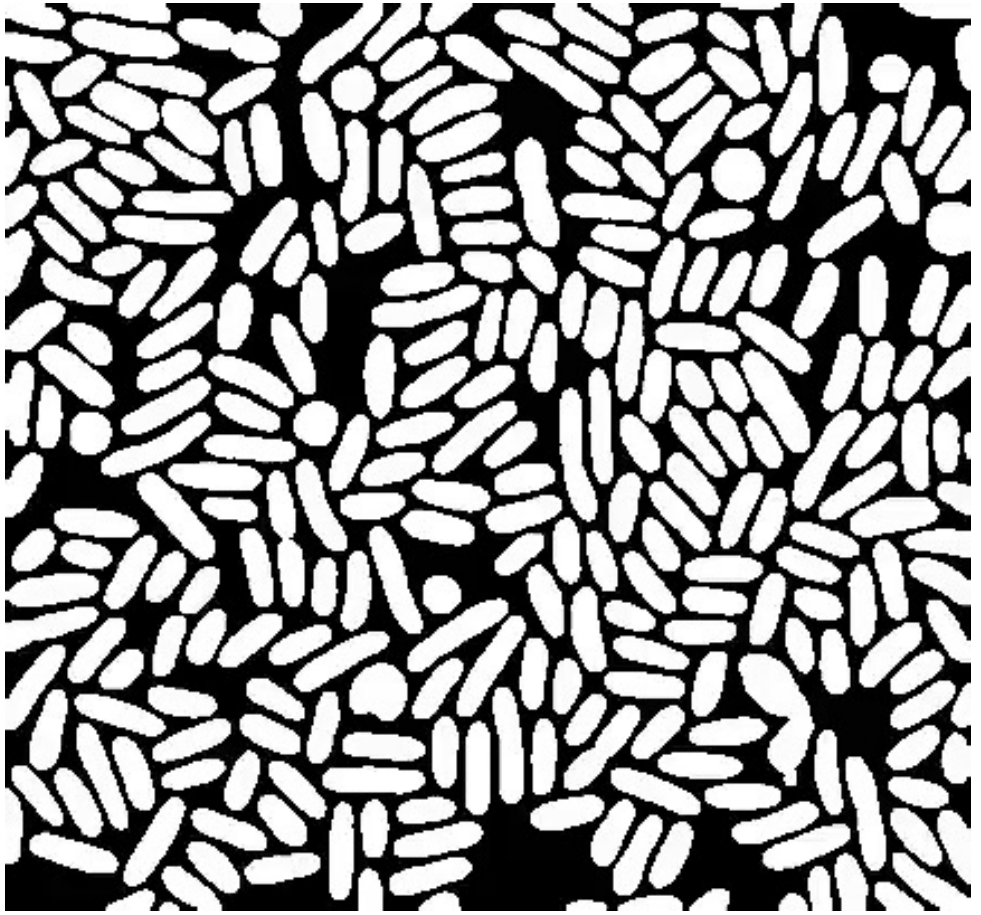
- **Convolution block**: Learning of cell shape
- **Downstream branch**: ‘what’-information
- **Upstream branch**: ‘where’-information

Reasons for using the U-Net:

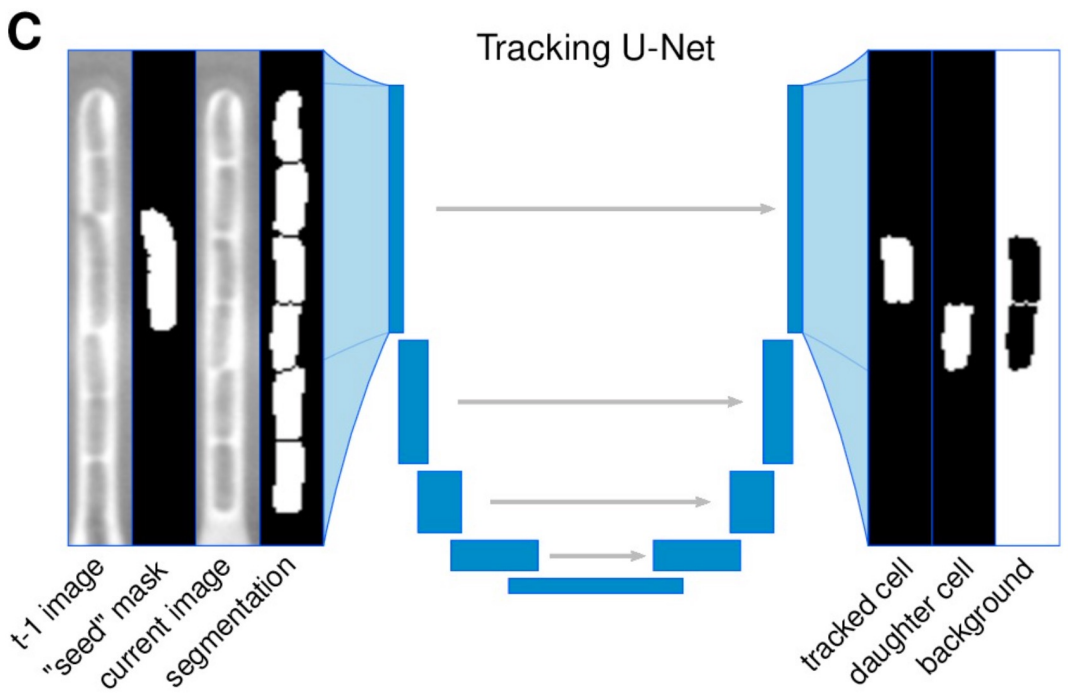
- State of the art machine learning model for segmentations (> 15 000 citations)
- Winner in several image segmentation competitions

Segmentation results





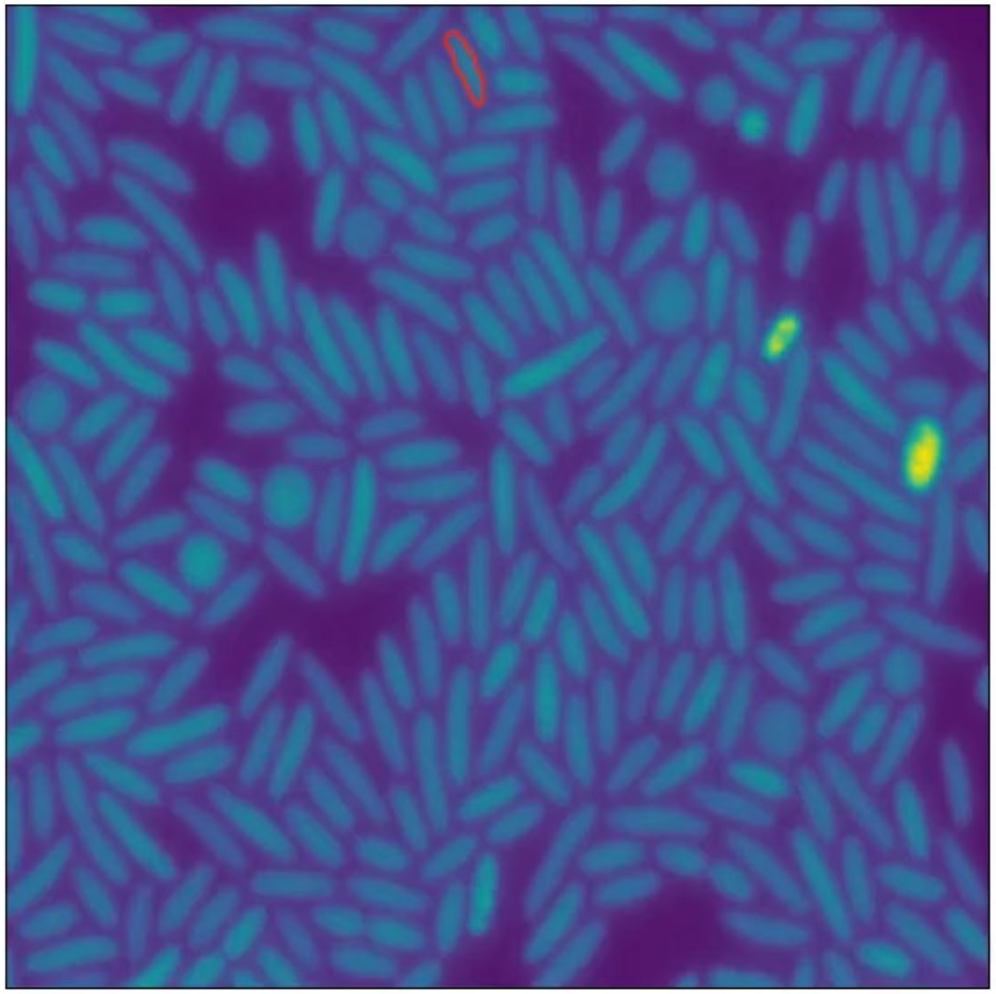
Deep Neural Network for tracking: U-Net



- Modified input and output compared to original U-Net
- Tracking is computed per time frame and per cell

Lugagne et al, Plos Computational Biology, 2020

Tracking results



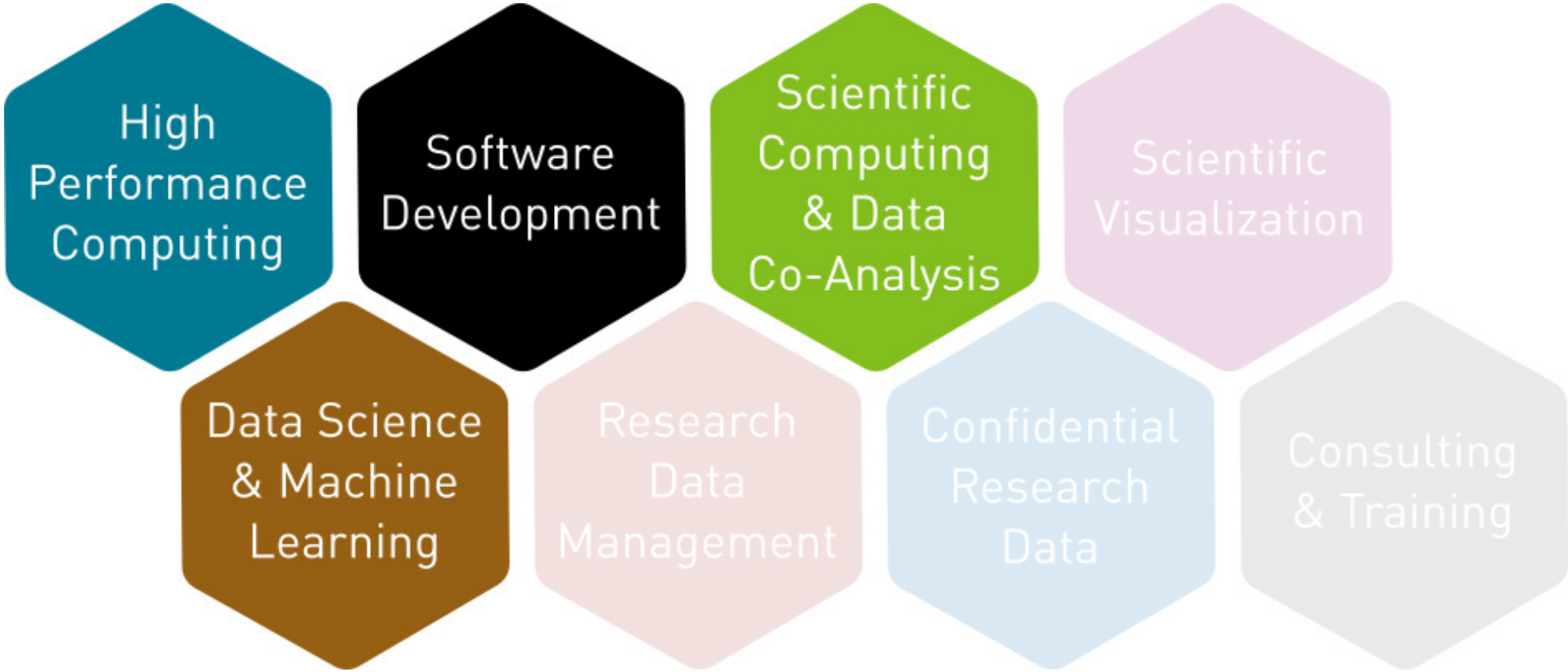
- Application of pretrained network
- First results look promising, but fine-tuning with own training-dataset is necessary

Requirements for upscaling

- Increasing interest from collaboration partners in pipeline

- Adaptations of pipeline are required:
  1. **Neural Networks:**
     - Building general models for different datasets and cell types
  2. **GUI:**
     - Manual correction of segmentation and tracking
  3. **Workflow managers:**
     - Enable modular usage of pipeline
     - Storage of metadata
  4. **HPC:**
     - Processing of multiple datasets at the same time

From prototype to production : role of SIS in upscaling



Benefits, challenges and risks of this working model

Benefits:

- Pool of experts which work on one solution
- Easy access to knowledge which is not available in research group
- Continuous exchange between researchers and SIS members

Challenges and risks:

- Find a common language (IT & Biology)
- Expectation management (AI can not perform miracles)
- Not every field/problem is suitable for the application of ML

ETH zürich

Thank you!

https://sis.id.ethz.ch/