AI-based High-throughput X-ray Imaging

Alexey Ershov

INSTITUTE FOR PHOTON SCIENCE AND SYNCHROTRON RADIATION (IPS)

Outline

- **Motivation**
  - Synchrotron-based X-ray Imaging
  - Methods for Life and Material Sciences
- **Components**
  - Deep Learning for Image segmentation
  - Simulation and Generative Modelling
  - Virtual Experiments
  - Neural Networks for Inverse problems
  - Evaluation of ML algorithms
- **Domain adaptation and Transfer Learning**
- **Unsupervised Learning**
- **Data and Experiment Lifecycle**
- **Outlook**

Motivation

Synchrotron-based X-Ray Imaging

Modern **synchrotron** light sources provide enormous photon flux densities.

Huge progress in detector technologies:

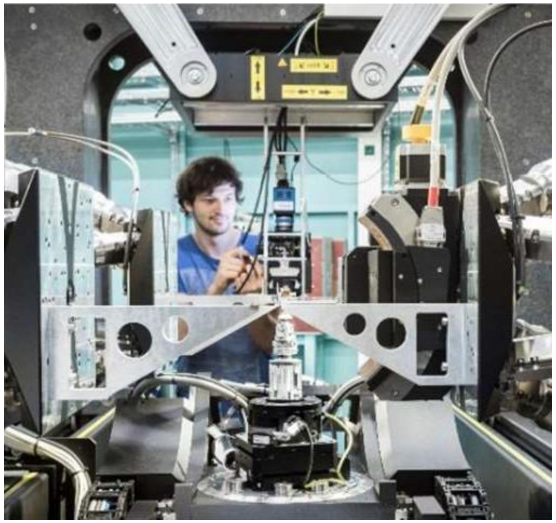
- High resolution
- Fast data acquisition
- Large storage capacity

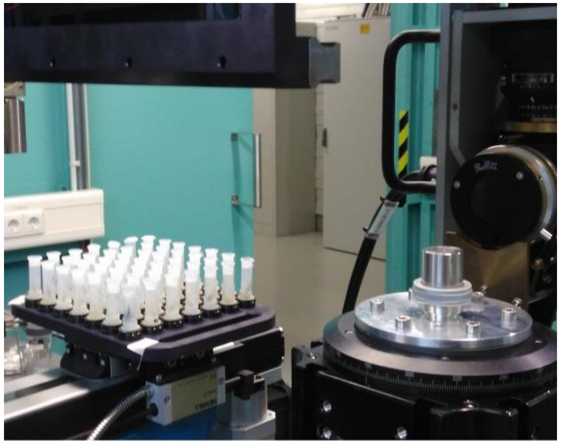
Methods:

- *In vivo* & *in situ* radiography and tomography
- **High-throughput** experiments with robotic sample changer

As a result:

- up to **10M** frames/s
- 20 tomograms/s
- 40 high resolutions (30 GB) samples/hour
- **hundreds** of samples per day!





robotic sample changer

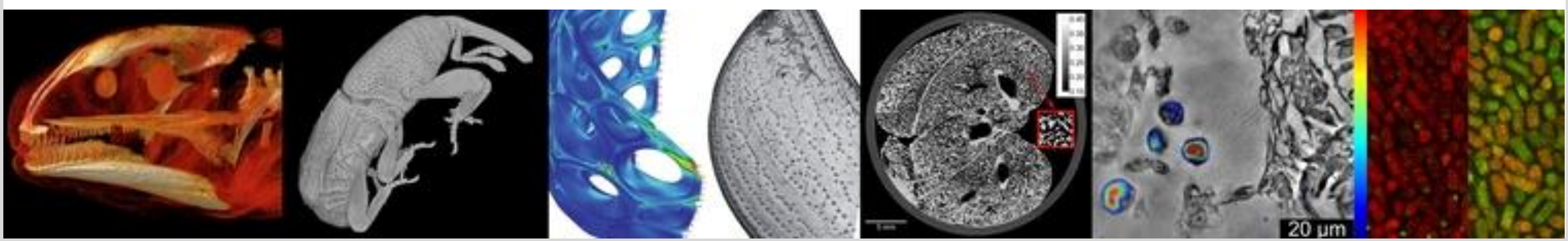
X-Ray Imaging for Life Sciences

Methods for hierarchical, correlated, *in vivo* imaging of organisms, tissues and cells:

- 3D phenotype atlases of model organisms
- Comparative 3D morphometrics of small animals
- 4D developmental studies of embryos
- 4D morphodynamics: mating, feeding, etc

Application fields:

- Developmental biology
- Genetics
- Medicine
- Zoology
- Palaeontology
- Environment
- Biotechnology
- Tissue Engineering
- Biomimetics



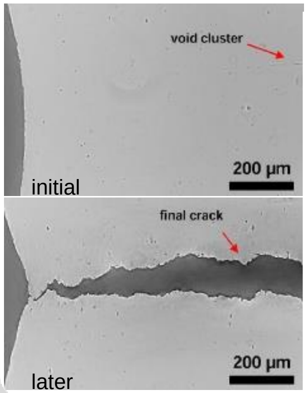
X-Ray Imaging for Material Sciences

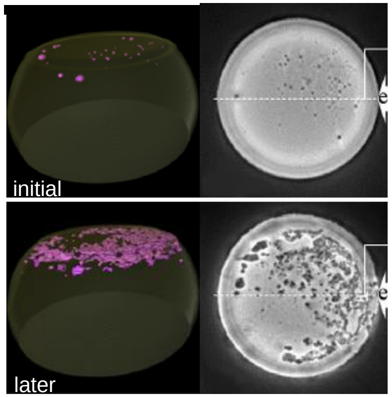
Methods for correlated, *in situ & operando* imaging of materials and processes

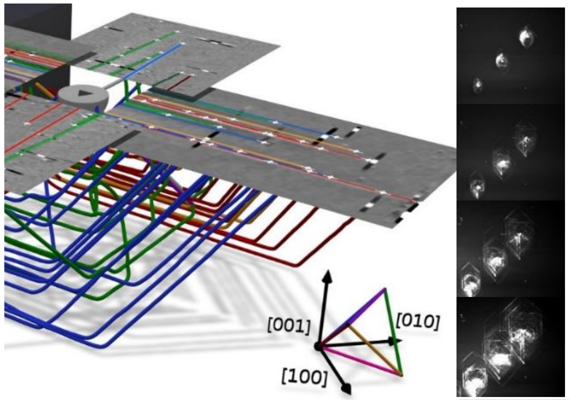
- Non-destructive 3D imaging
- Multi-contrast (absoption, phase-contrast, fluorescence)
- **Hierarchical** imaging (2D/3D screening and zooming into ROIs)

Application fields:

- Energy
- Nanotechnology
- Information Tech.
- Transport
- Fluid dynamics
- Aerospace

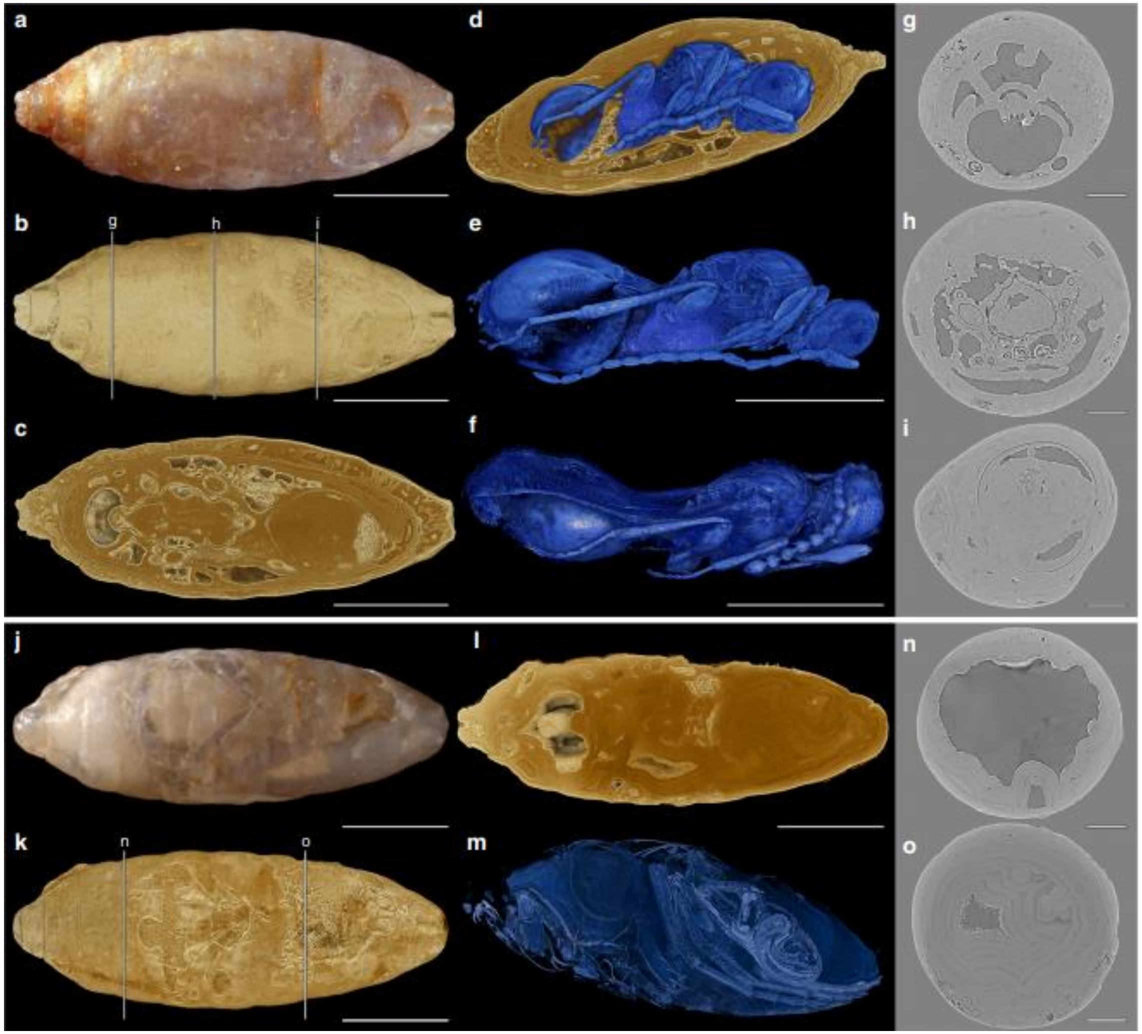






Deep Learning
for Image Segmentation

High Throughput Tomography of Fossils

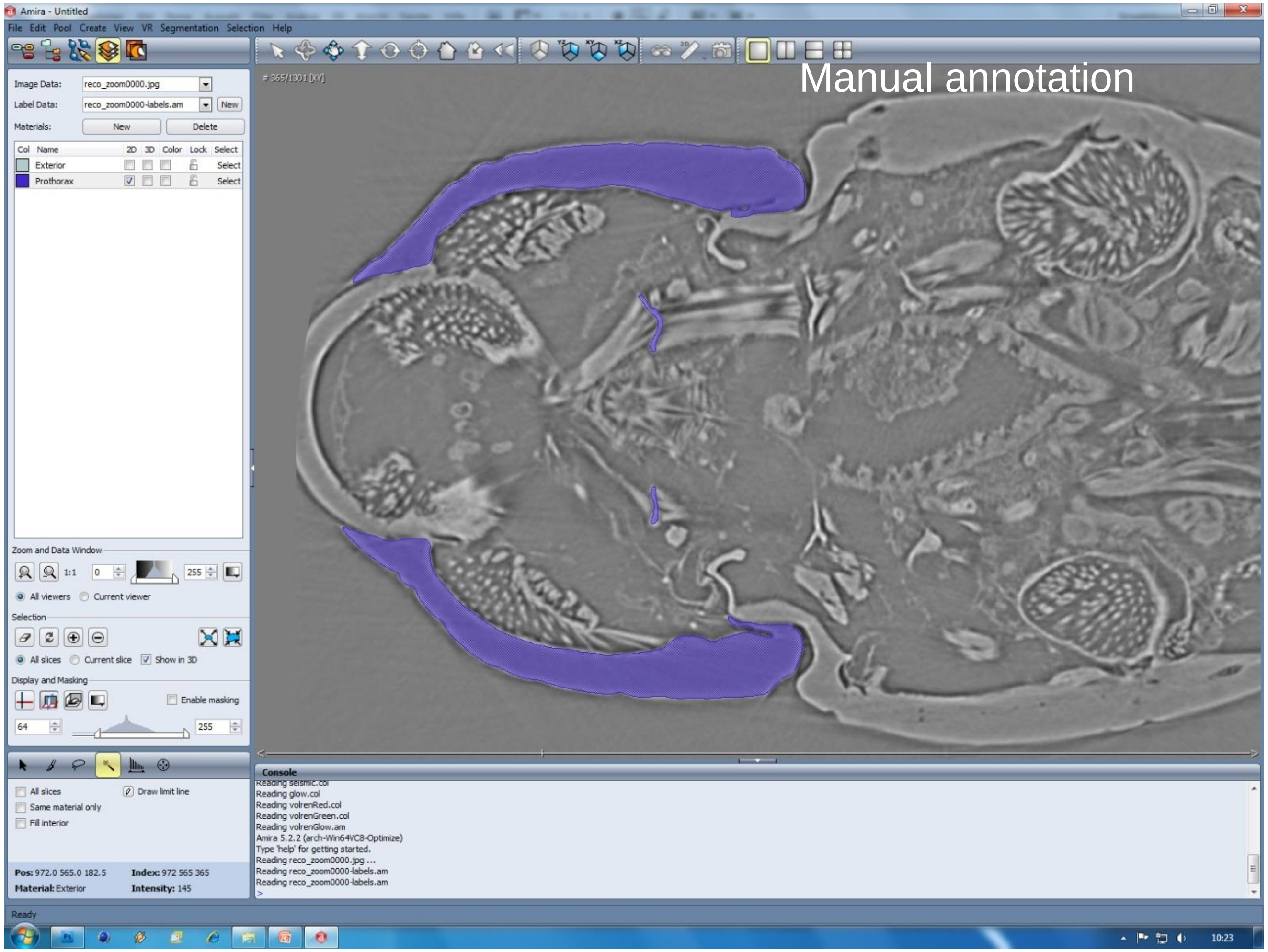


- **1510** fly pupae scanned in one week
- **55** parasitoid wasps found

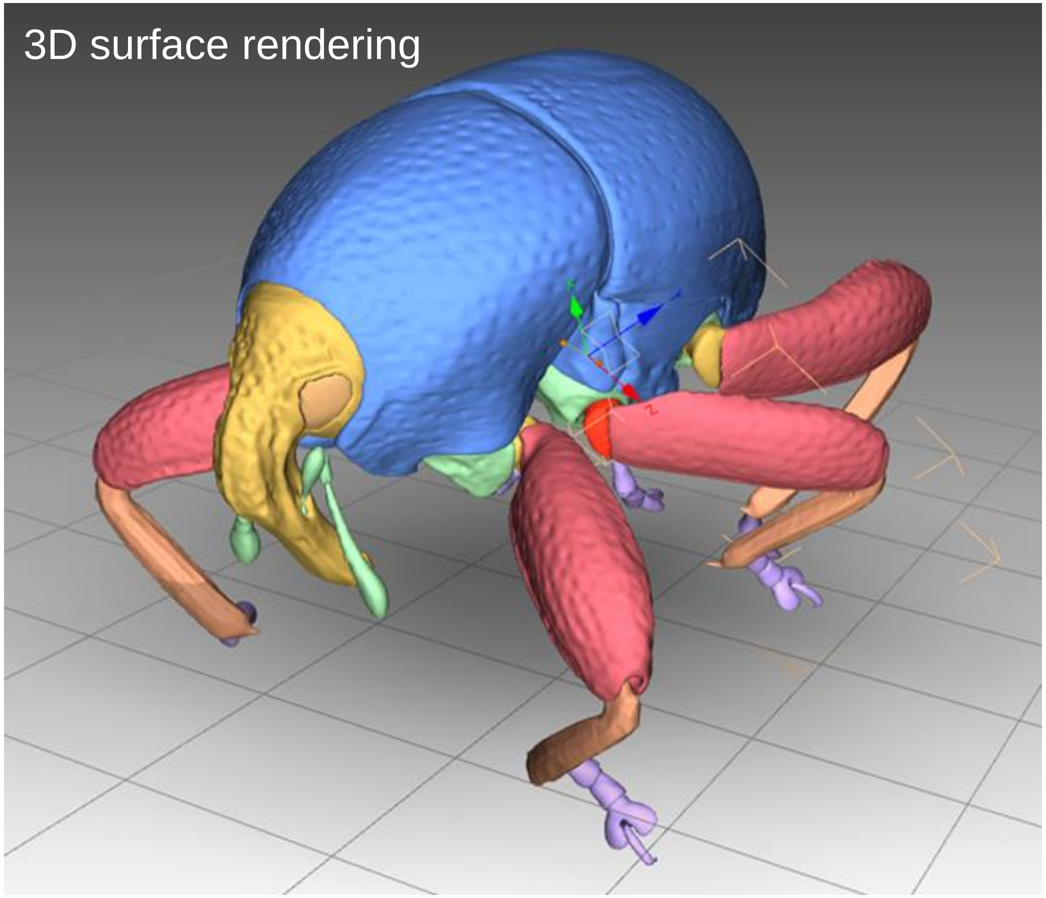
Van De Kamp et al. 2018: **Nature Communications**

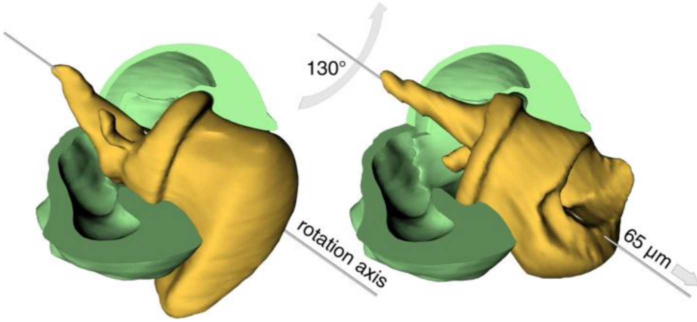
Segmentation: Manual approach

Aim: High quality **segmentation** of morphological structures in specimens.



Segmentation: Manual approach





Biological screw joint

Trigonopterus weevils

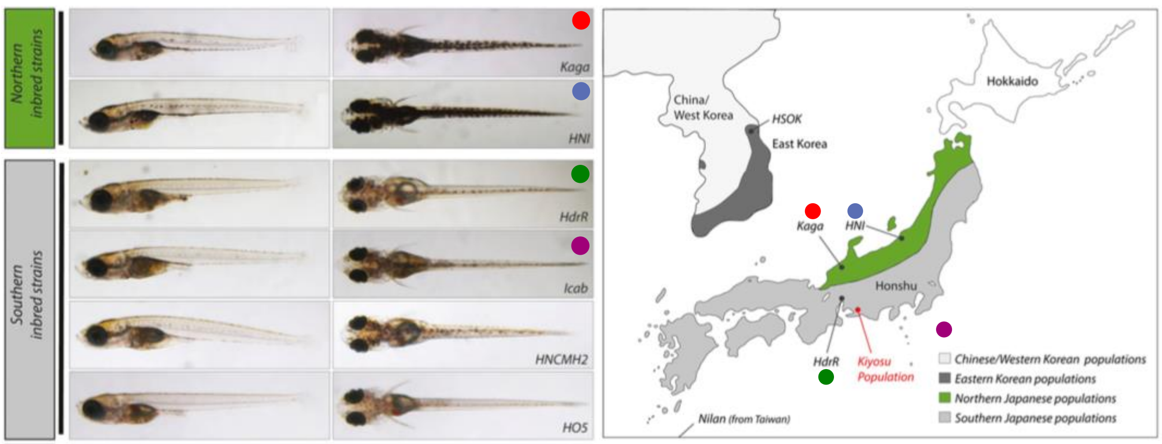
 Alexey Ershov, ershov@kit.edu

Institute for Photon Science and Synchrotron Radiation, Imaging Group

3D Morphometric Analysis

Model organisms provide insight into the working of other organisms.

- Aims:
  - Study of morphological data as basis for genotype-phenotype association
  - **Ultimate**: A smart way to decode the genome of a specimen



different phenotypes

Oryzias latipes (Medaka)

In collaboration with:

Centre for Organismal Studies Heidelberg

UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

EZRC

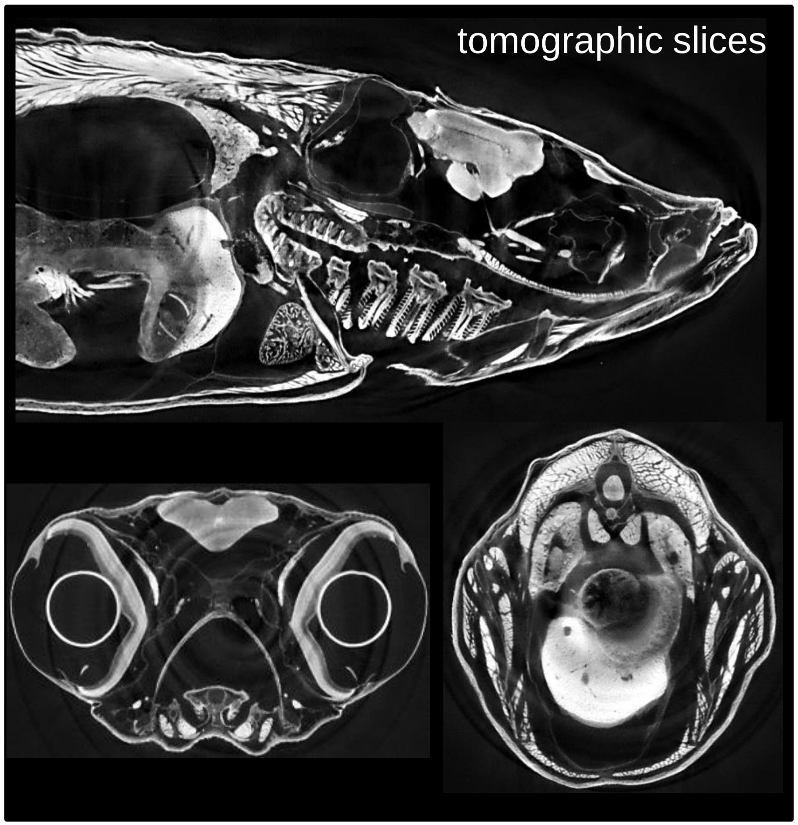
Alexey Ershov, ershov@kit.edu

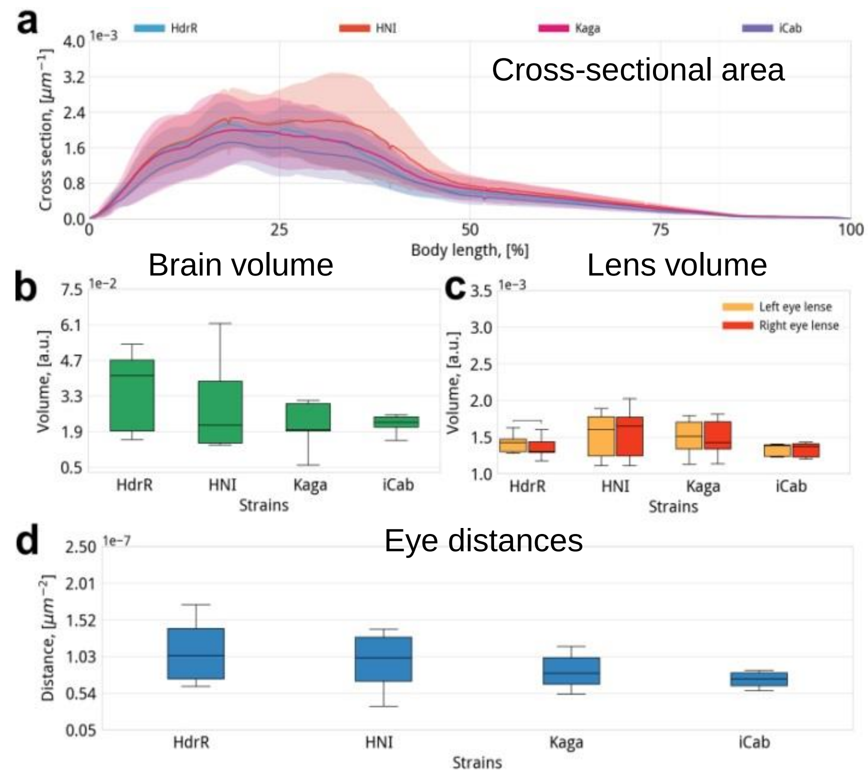
Institute for Photon Science and Synchrotron Radiation, Imaging Group

3D Comparative Studies

Model organism: Medaka fish, different inbred lines

Comparative study of inbred lines for the identification of line-specific characters





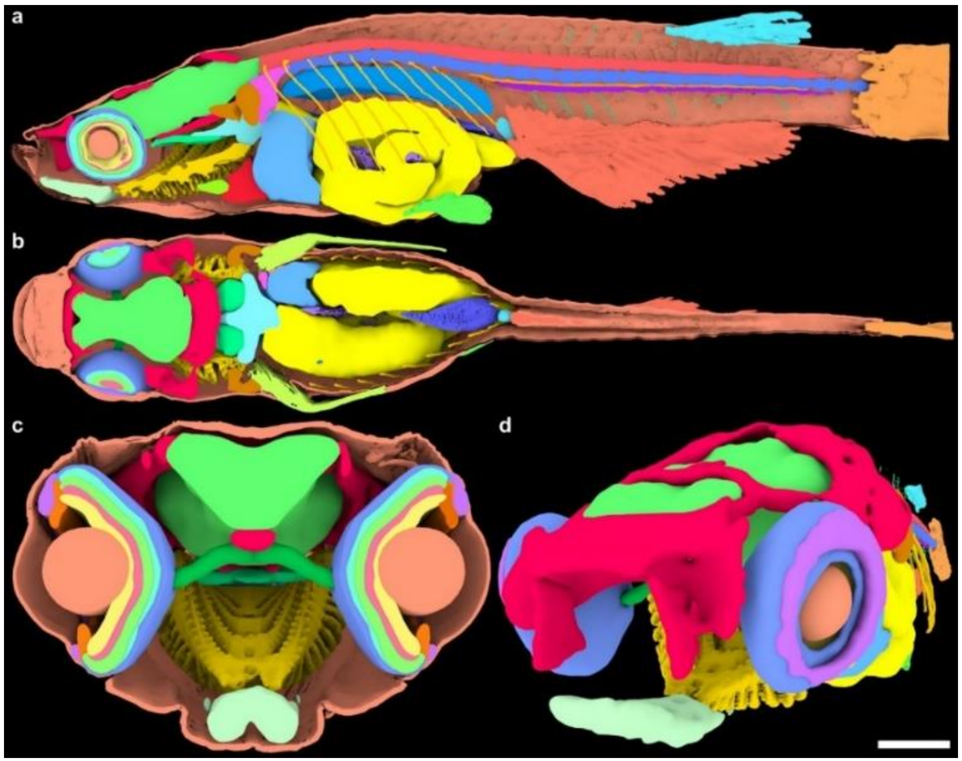
Statistical analysis

V. Weinhardt et al., SREP 2018

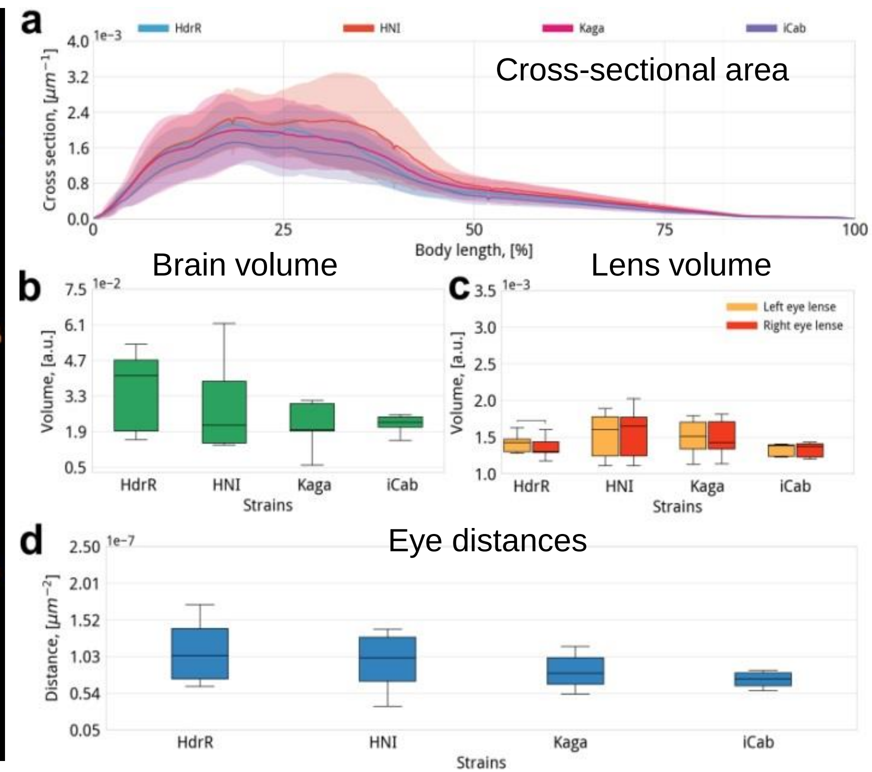
3D Comparative Studies

Model organism: Medaka fish, different inbred lines

Comparative study of inbred lines for the identification of line-specific characters



full body 3D digital atlas



Statistical analysis

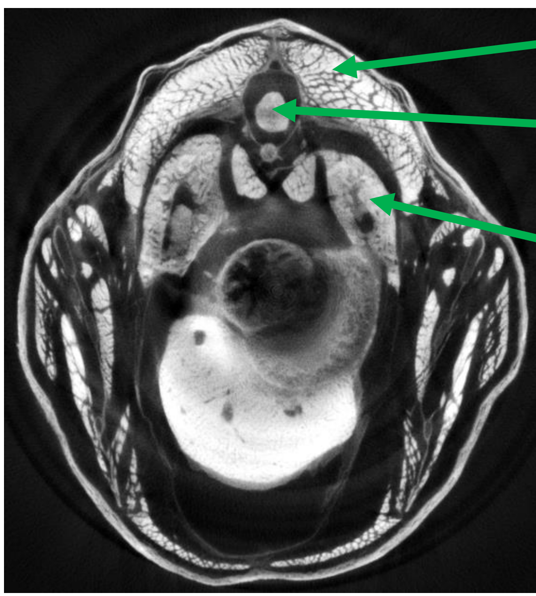
V. Weinhardt et al., SREP 2018

Challenges

Shape variation among specimens

Hearts

Contrast variations between organs



Muscles

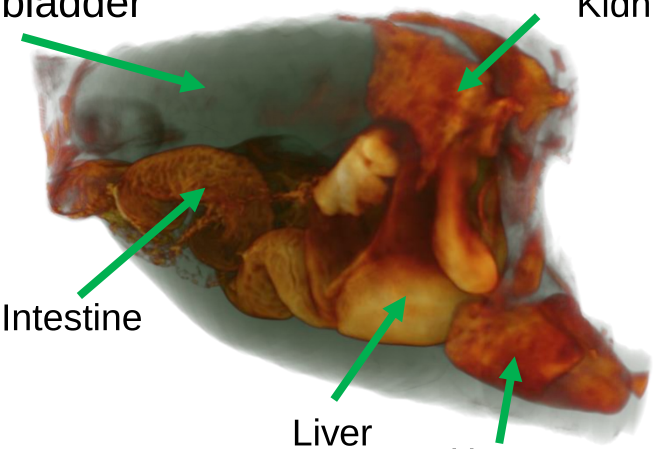
Spinal cord

Kidneys

Tightly **packed** organs.

Swim bladder

Kidneys

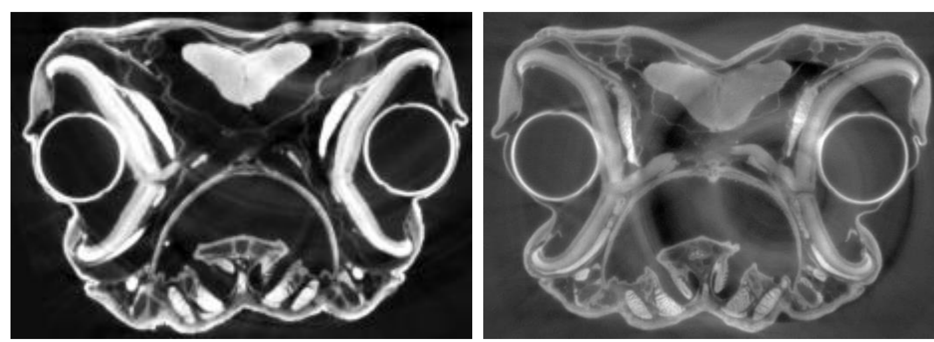


Intestine

Liver

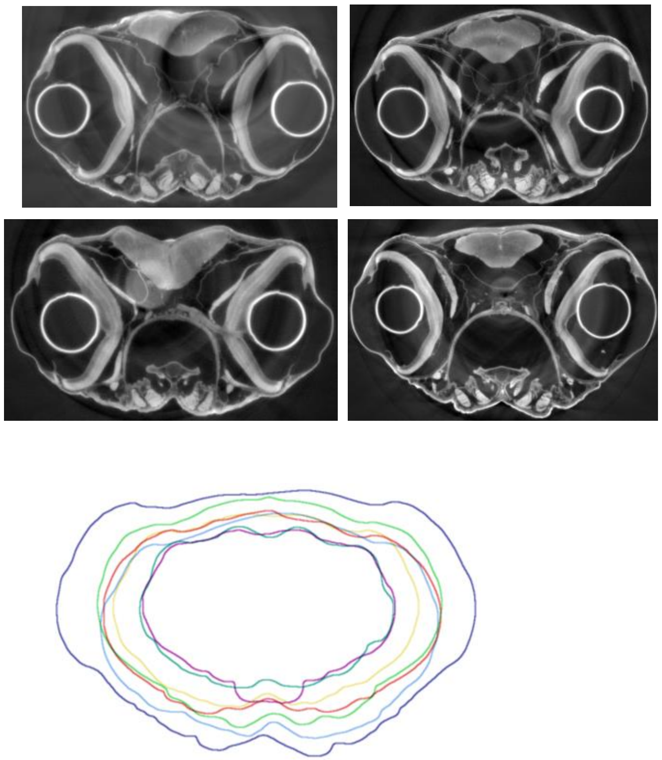
Heart

Contrast **variations** between **specimens**

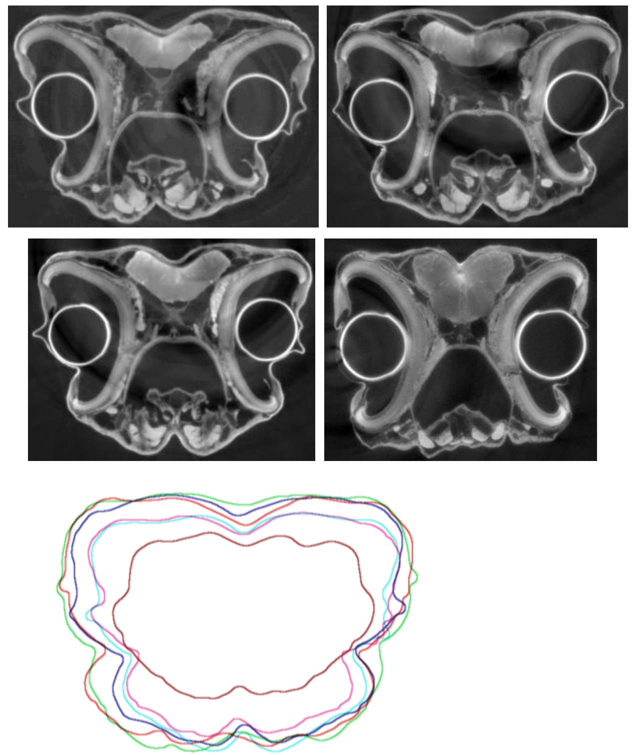


Challenges

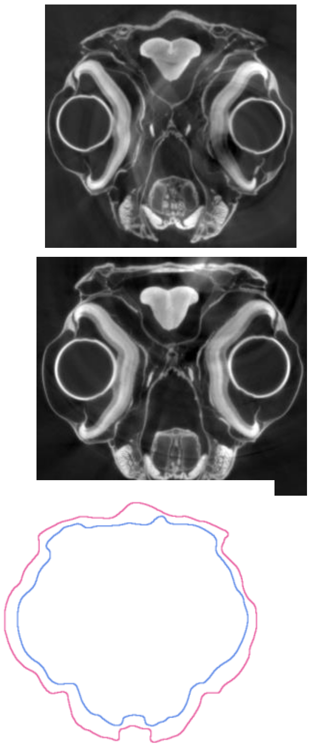
Large **morphological variability** which we want to measure with high precision.



fish line 1



fish line 2

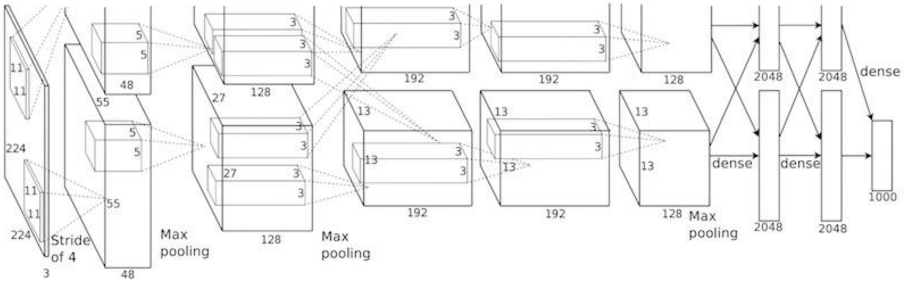


fish line 3

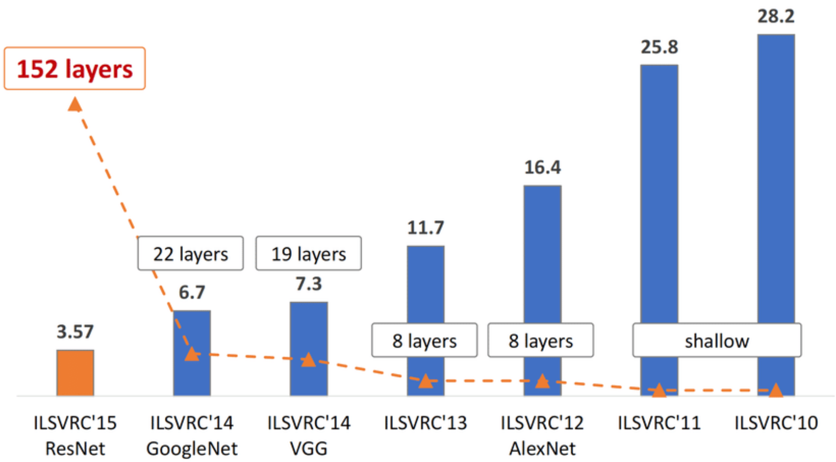
How to design **accurate**, **robust**, but also **general** segmentation procedure?

Computer Vision: Automated classification

AlexNet: Approach based on **deep neural networks**.

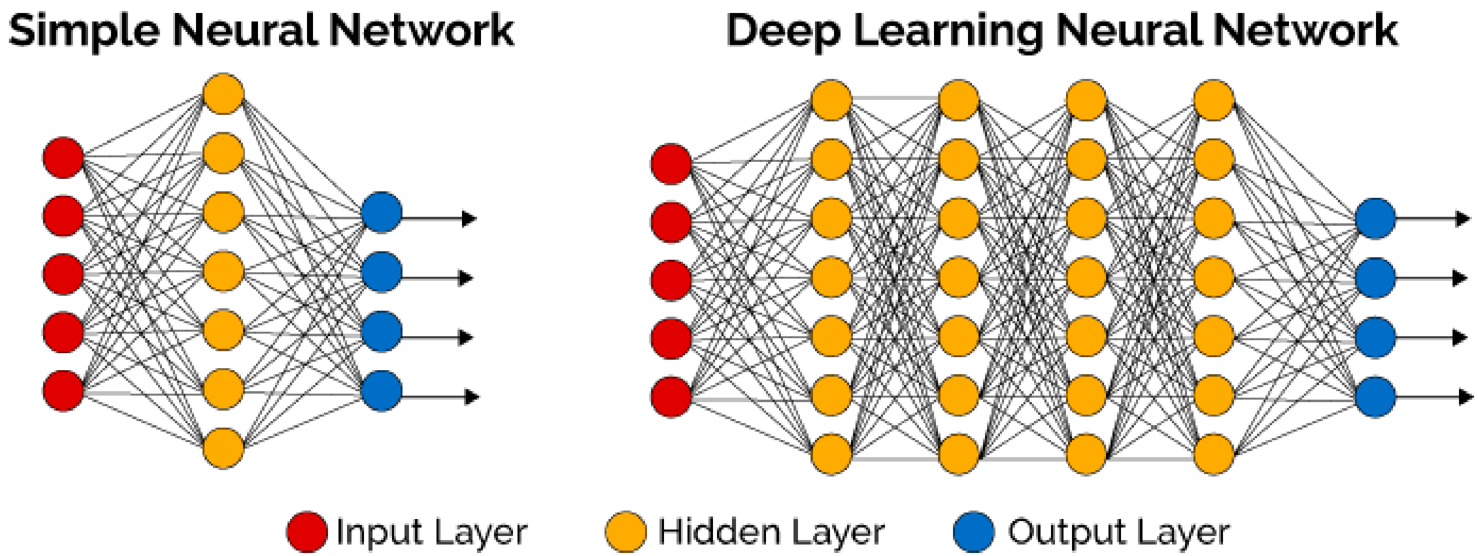


Krizhevsky et al., arXiv, 2012



Russakovsky et al., arXiv, 2014

Deep Learning



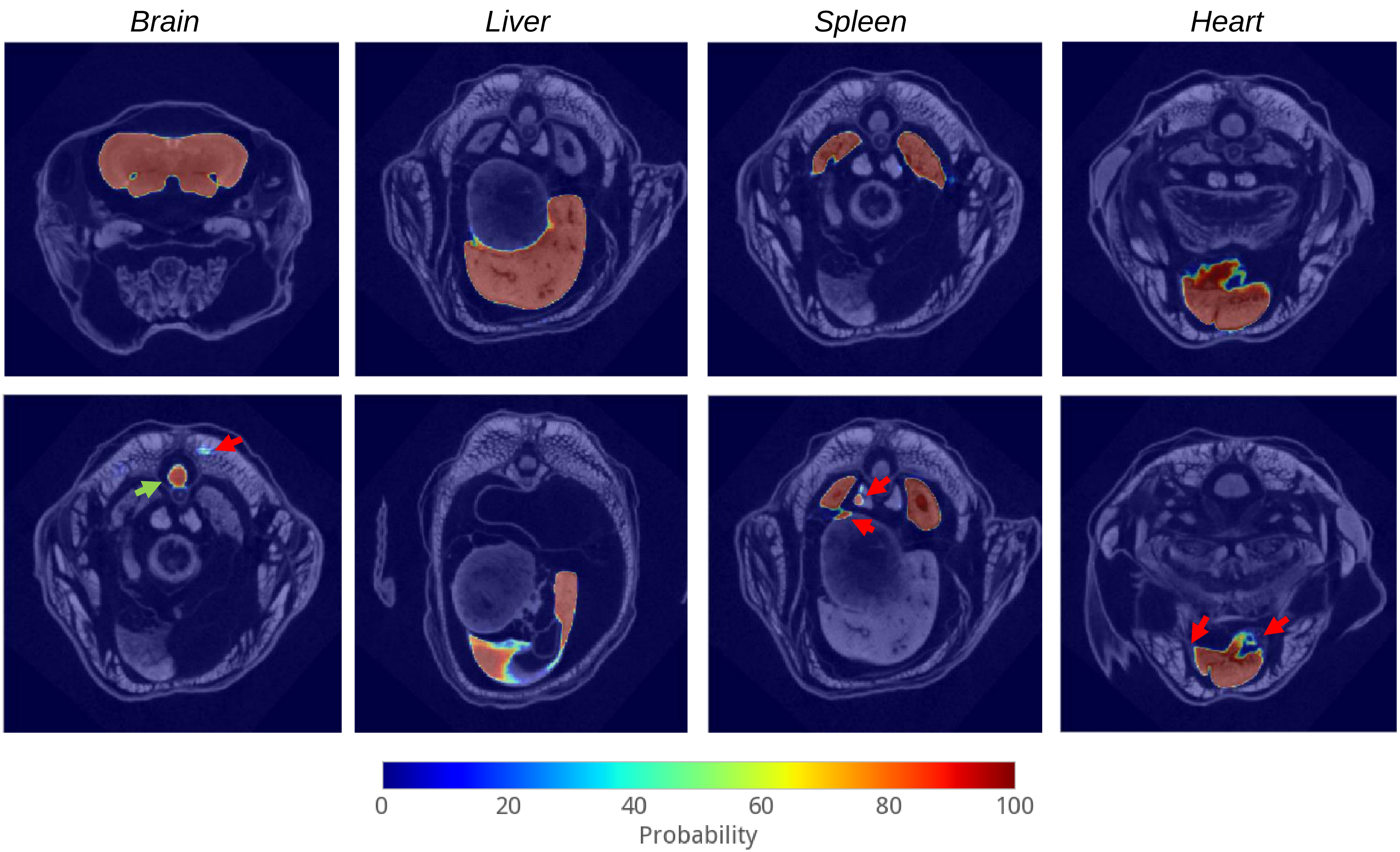
Learnable linear classifier

Deep, hierarchical, non-linear model

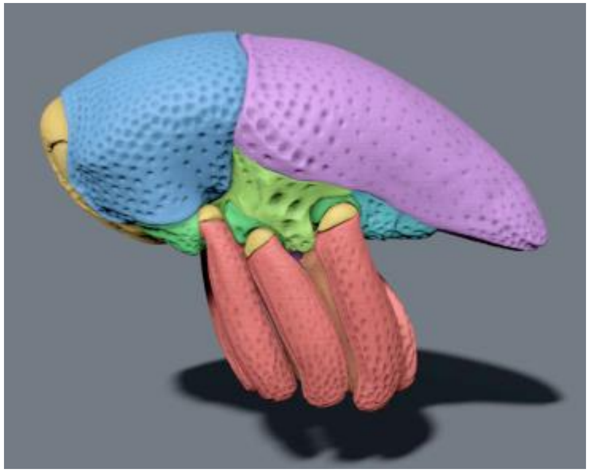
Idea:

- Train the model on a large amount of manually labeled data
- Optimize on each step to improve a certain metric (find best mapping)
- Apply on the new data

Deep Learning: Organ Segmentation



Processing Time and Costs



Full body segmentation, **1500** samples, **10** organs

	Manual	Neural Networks
Time: One dataset	77 h (215 slices + inter.)	(24 h + 2 min) * 10 ~ **240 h**
Time: All	77 * 1500 = **13 years**	(24 h * 10) + (2 min * 1500) ~ **30 days**
Computation cost	**a lot**	30 * 24 * **0.5** Euro / h = **360** Euro

Google Cloud Platform

Deep learning outperform classical methods on large datasets: both more precise and cost effective

Alexey Ershov, ershov@kit.edu

Institute for Photon Science and Synchrotron Radiation, Imaging Group

Challenges for Deep Learning

Deep learning methods provide very promising results.
But there are some open questions and challenges.

- Training ➡ **Ground truth** data
- Variation in data ➡ **Generalization** of NN Models
- Common data features ➡ Model Sharing and **Transfer Learning**
- Different conditions / experiments / methods ➡ **Domain adaptation**
- New applications / methods ➡ **Unsupervised Learning**

Alexey Ershov, ershov@kit.edu

Institute for Photon Science and Synchrotron Radiation, Imaging Group

Karlsruhe Institute of Technology

Simulation and Generative Modelling

syris: Virtual X-ray Imaging Lab

Karlsruhe Institute of Technology

- Aims
  - Design and **optimize experiments** and **imaging conditions**
  - Evaluate image processing **algorithms**, choose best parameters
  - **Train Machine Learning** methods on **synthetic** data
- Approach
  - Model complete experiment and data acquisition
  - Python API for easy usage
  - GPU accelerated computation

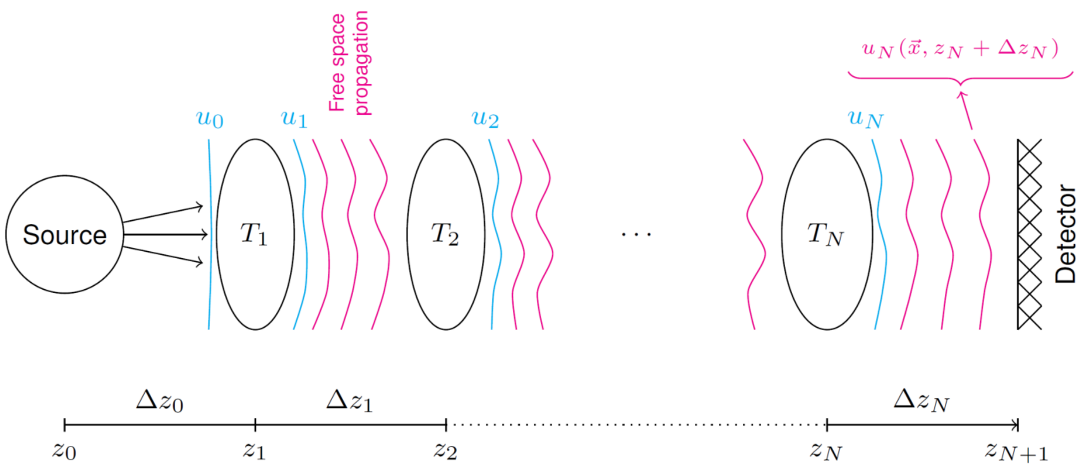


Image formation in ***syris***

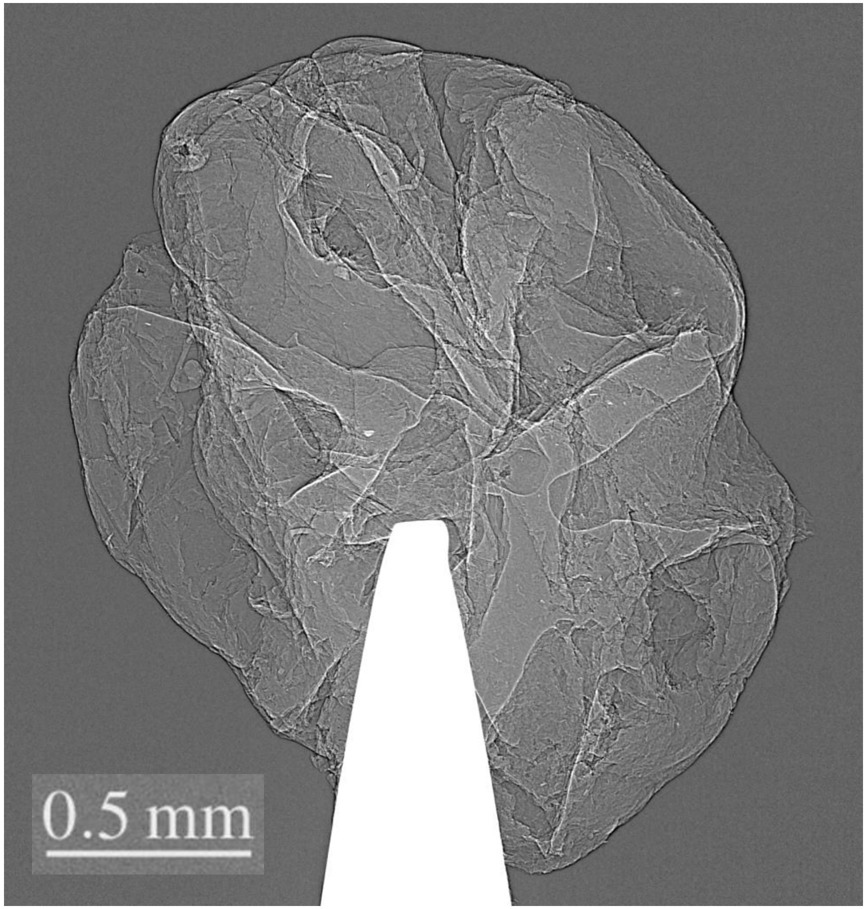
syris: Simulation Parameters

Category	Parameter	Value
Source	Storage ring energy	2.5 GeV
	Source type	Bending magnet
	Source size (FWHM)	–
	Magnetic field	1.5 T
	Electric current	106 mA
	Beam mode	White beam
	Energies	8 keV to 30 keV, 1 keV step
	Filters	0.7 mm aluminium
	Slit position	–
	Slit opening	–
Sample	Material	Wax on a steel needle
	Dimensions	1.5mm × 1.5 mm × 1.5 mm
	Source distance	–
	Detector distance	82 mm
Acquisition	Experiment	Tomography
	Number of projections	3000
	Data processing	Flat-field correction
	Reconstruction algorithm	Filtered back projection

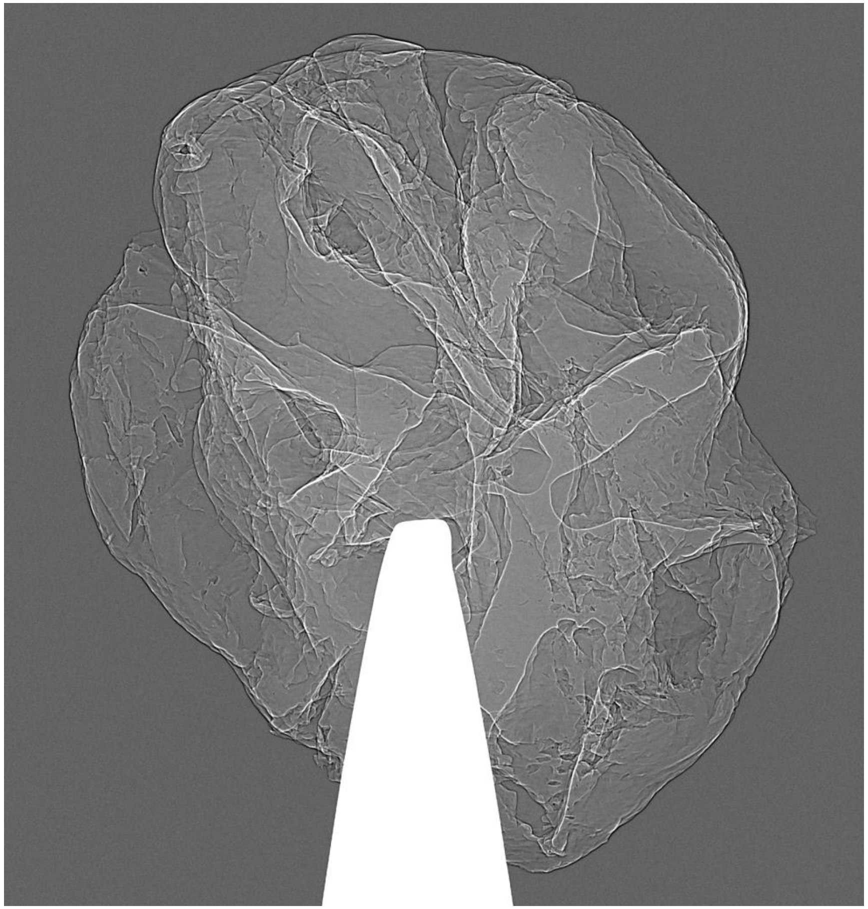
Detector system

Parameter	Value
Type	Indirect
Effective pixel size	1.22 µm
Objective lens	Mitutoyo Plan Apo
Magnification	10×
Numerical aperture	0.28
Scintillator	LSO:Tb (Lu_2SiO_5)
Density	7.4 g cm^{-1}
Thickness	12 µm
Light yield	36 photons keV^{-1}
Emission peak	545 nm
Camera	PCO.dimax
Sensor size	2016 × 2016
Sensor pixel size	11 µm
A/D factor	10 e^- per count
Quantum efficiency	49% at 525 nm
Dynamic range	1600:1
Frame rate	25 frames s^{-1}

syris: Simulation of a Complex Shape



Real data

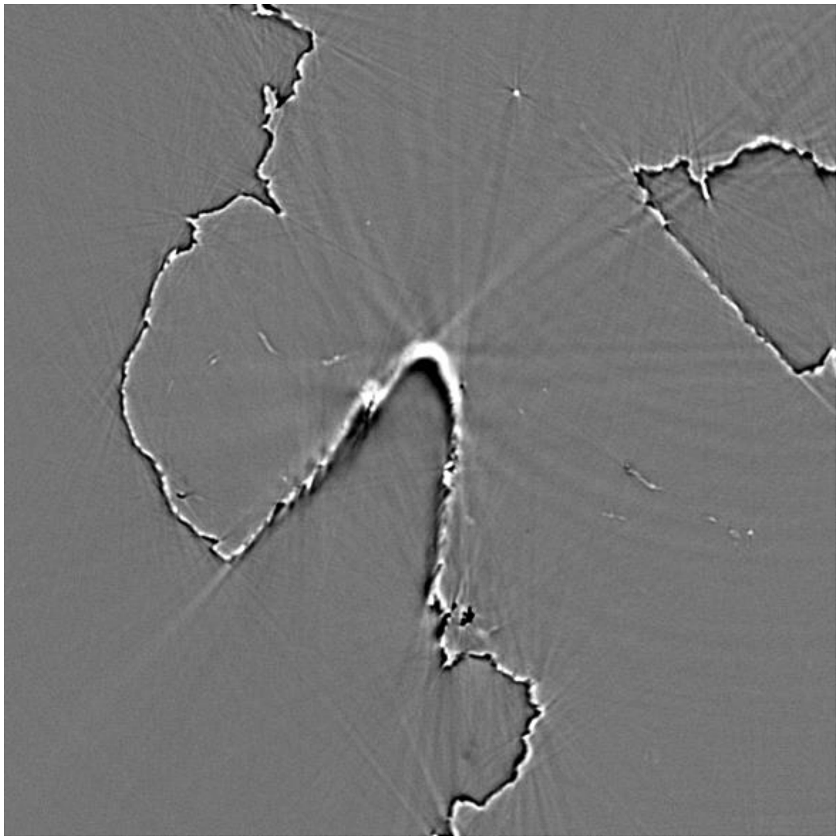


Simulation

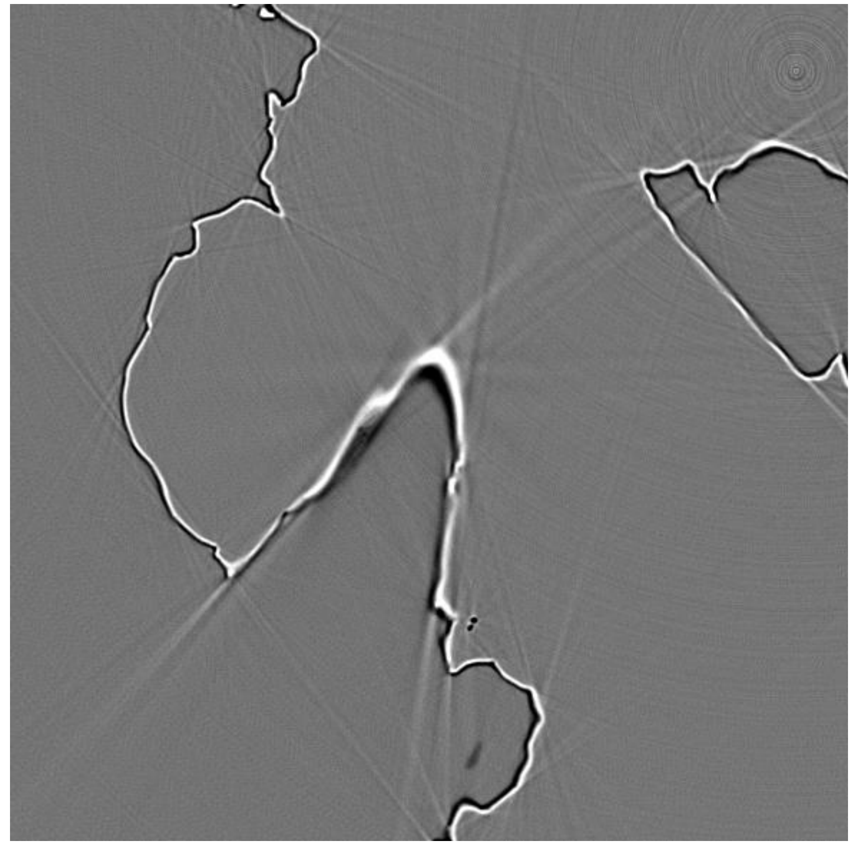
T. Farago et. al. ***syris***: *a flexible and efficient framework for X-ray imaging experiments simulation*, Journal of Synchrotron Radiation (2017)

syris: 3D Reconstruction

2D slice through the reconstructed volume



real data



simulated data

Use as **ground truth** data to train **Machine Learning** methods

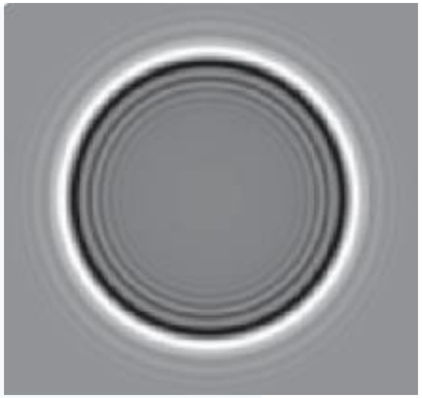
Institute for Photon Science and Synchrotron Radiation, Imaging Group

syris: Simulation example

A simple script to simulate a plastic ball in Python programming language.

```
syris.init()
energies = np.arange(10, 30) * q.keV
n = 1024
pixel_size = 0.4 * q.um
distance = 2 * q.m
material = make_henke('PMMA', energies)

sample = make_sphere(n, n / 4 * pixel_size, pixel_size, material=material)
image = propagate([sample], (n, n), energies, distance, pixel_size).get()
plt.imshow(image)
plt.show()
```



Open-source:

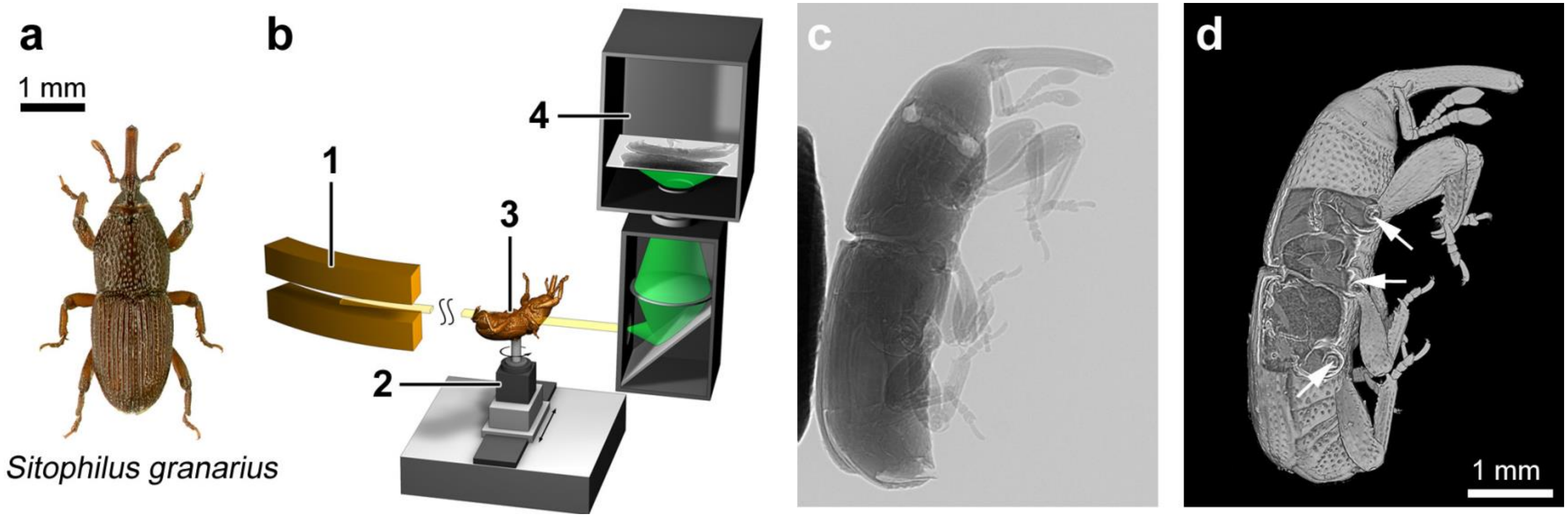
https://github.com/ufo-kit/syris

Virtual Experiments

Ultrafast *in vivo* 4D Cine-tomography

Application: Analyze the function of hip joints during defense mechanism in living insect.
Aim: Real-time study of structure **+** motion

Fast setup: **20** tomograms/s

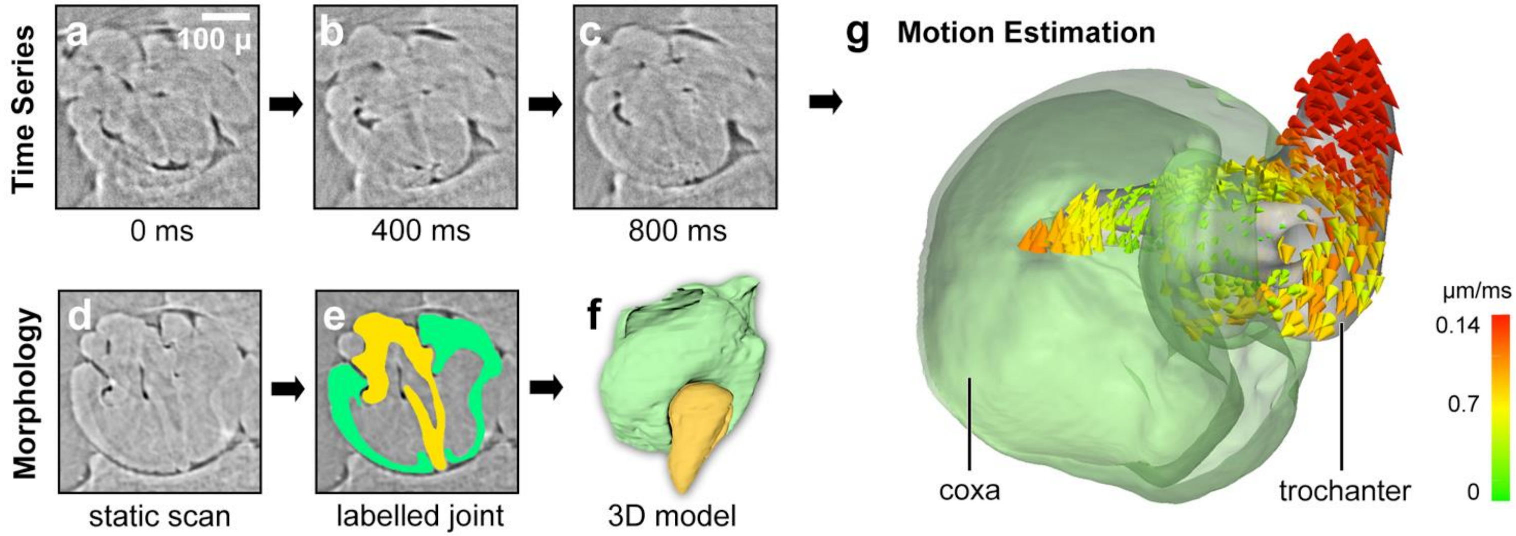


Experimental setup   Radiographic projection   Volume rendering

Source: beam energy E = 30 keV, photon flux = 10^{12}photons/s/mm^2

T. dos Santos Rolo et al. *In vivo X-ray 4D Cine-tomography for Tracking Morphological Dynamics*
Proceedings of the National Academy of Sciences 2014

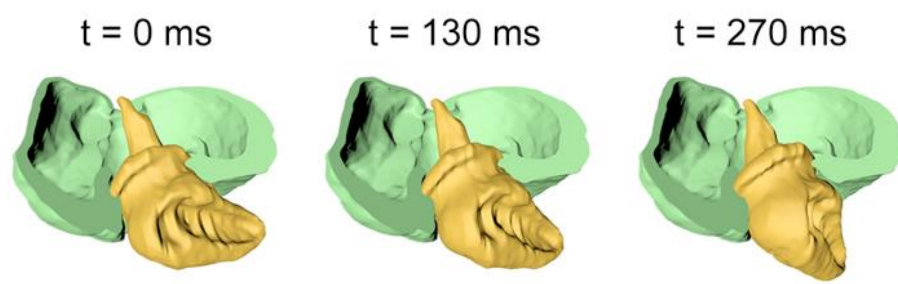
Motion Analysis and Kinematics



Challenge: How to design an experiment and find optimal conditions?

- We want:
  - **High**-resolution 3D data
  - **Ultra** fast acquisition
  - **Good** image quality
- But:
  - **Low** dose, longevity of a sample
  - **No** image artifacts

Reconstructing **Morphodynamics**



Alexey Ershov, ershov@kit.edu

Institute for Photon Science and Synchrotron Radiation, Imaging Group

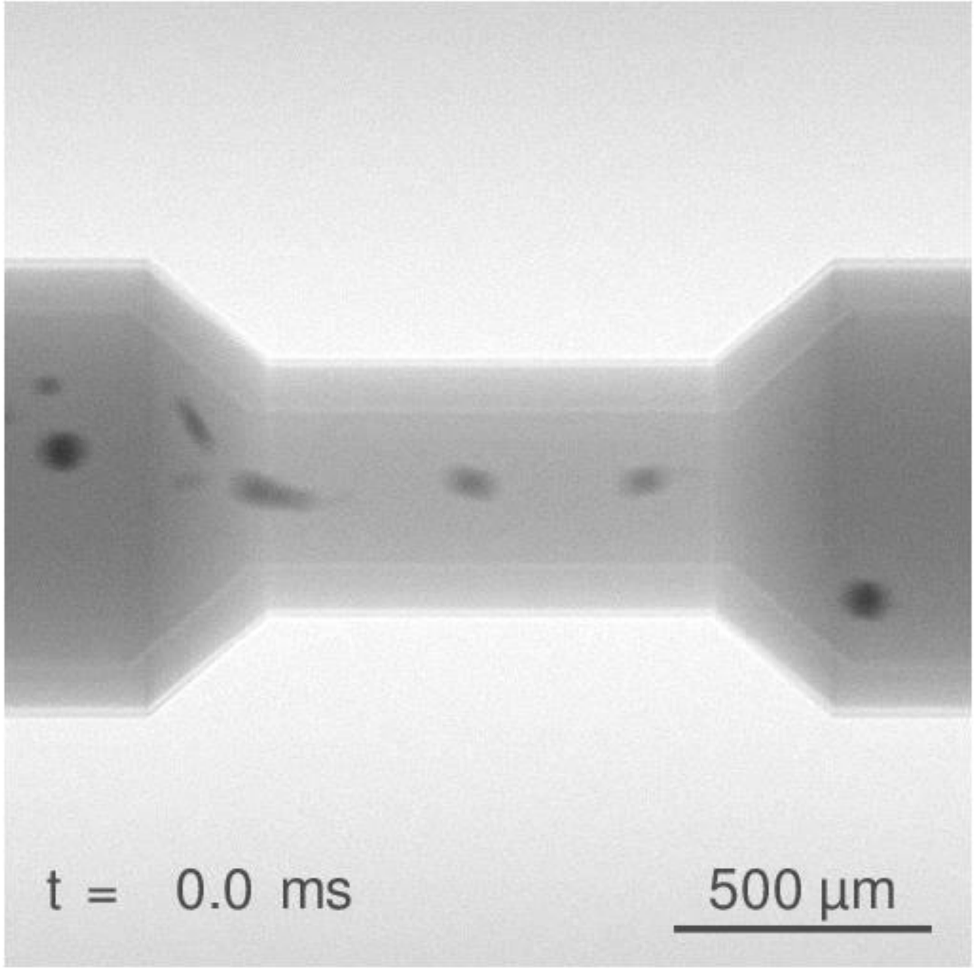
syris: Virtual *in-situ* Experiment

- *syris* accounts for:
  - Motion modeling
  - Noise
  - Beam variations
- Extendable for:
  - Elastic deformations
  - Artifacts modeling

Example:

Simulation of liquid droplets

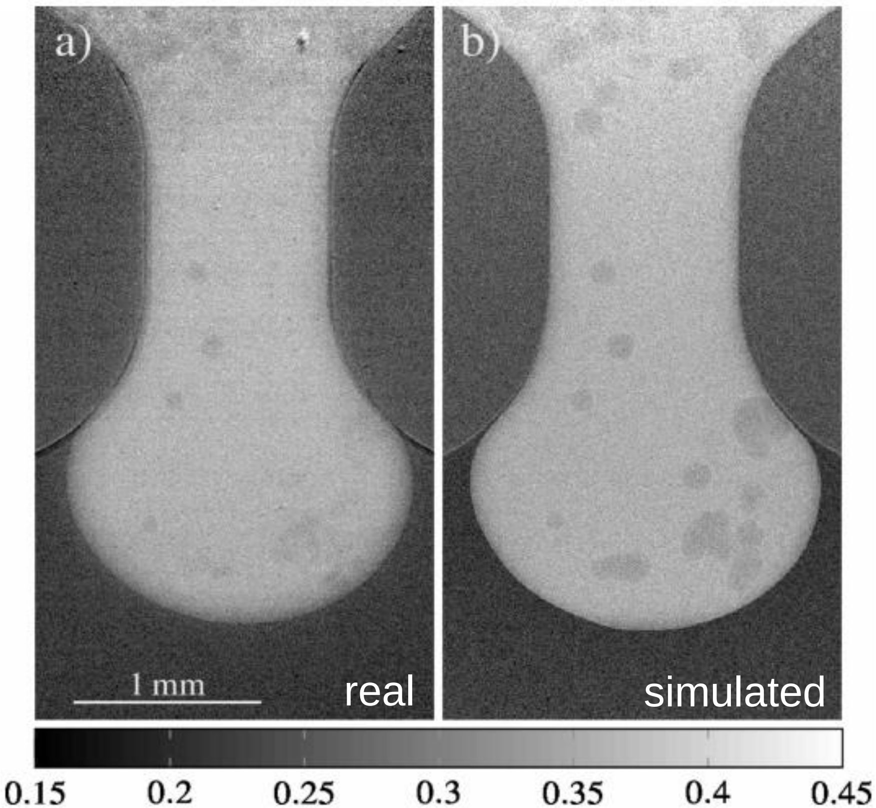
- Camera: 5 000 frames/s
- Speed: 20 pixels/frame



Alexey Ershov, ershov@kit.edu

Institute for Photon Science and Synchrotron Radiation, Imaging Group

Evaluation of data processing methods



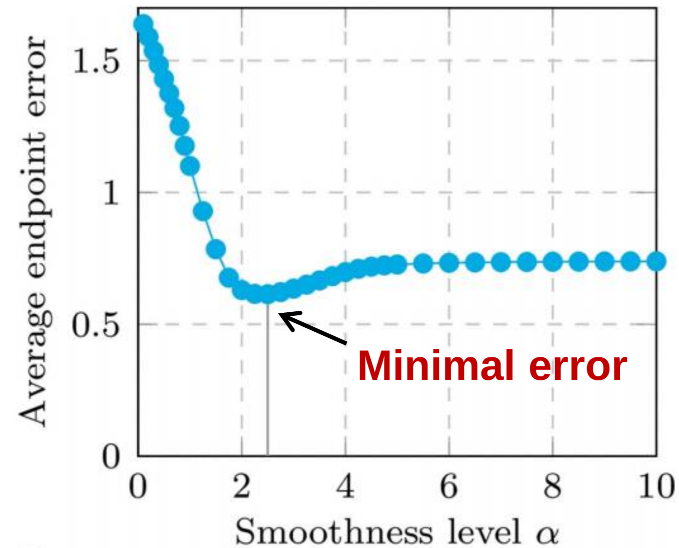


Figure 11
Motion estimation accuracy of algorithm M_4 as a function of the smoothness parameter α.

Computational model:

$$E_{\mathrm{HS}}(u, v) = \int_{\Omega_2} (I_x u + I_y v + I_t)^2 + \alpha(|\nabla_2 u|^2 + |\nabla_2 v|^2) \mathrm{d}x\mathrm{d}y$$

Optimization parameter

- **Optimize** parameters on **simulated** data
- **Apply** on real data
- **Verify** correctness and estimate **confidence**

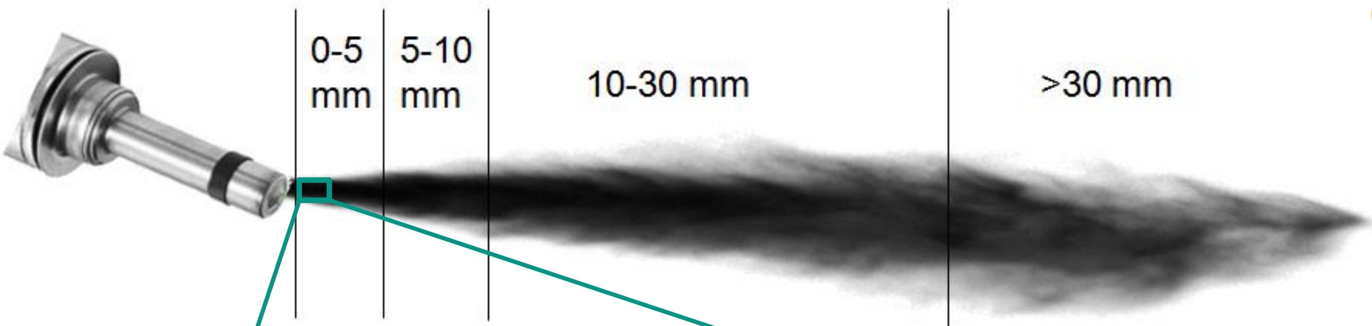
Institute for Photon Science and Synchrotron Radiation, Imaging Group

Neural networks
for Inverse problems

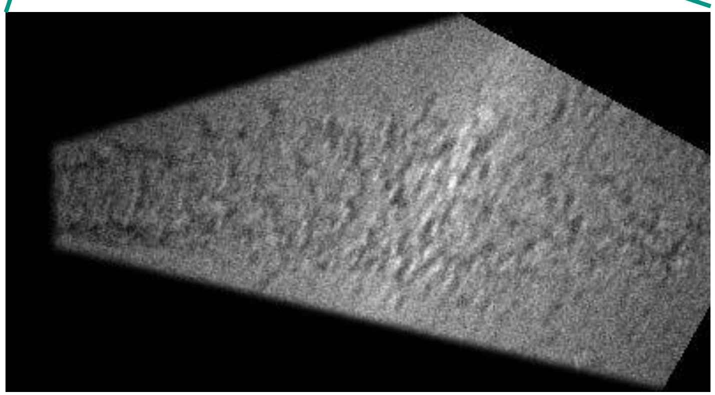
Alexey Ershov, ershov@kit.edu

Institute for Photon Science and Synchrotron Radiation, Imaging Group

Fast Radiography of Fuel Sprays



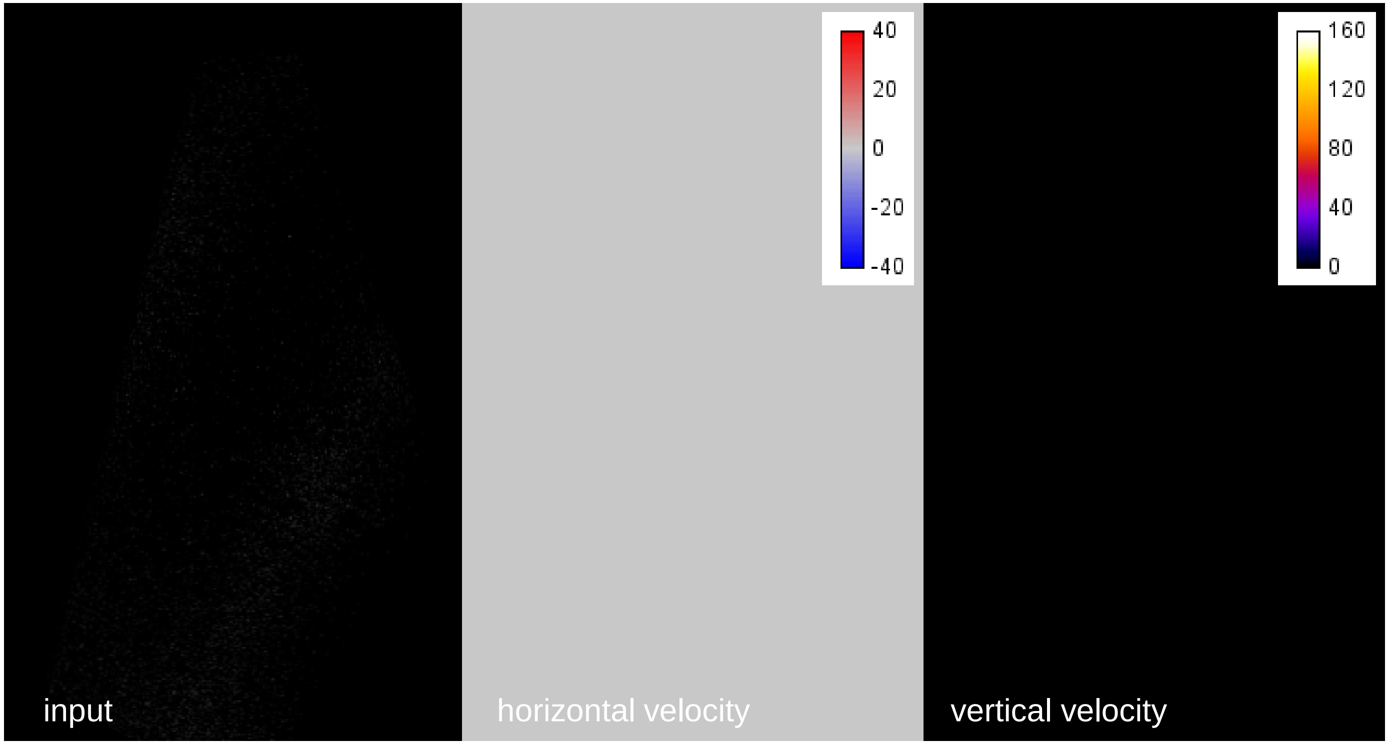
Shadowgraphy



X-ray radiography

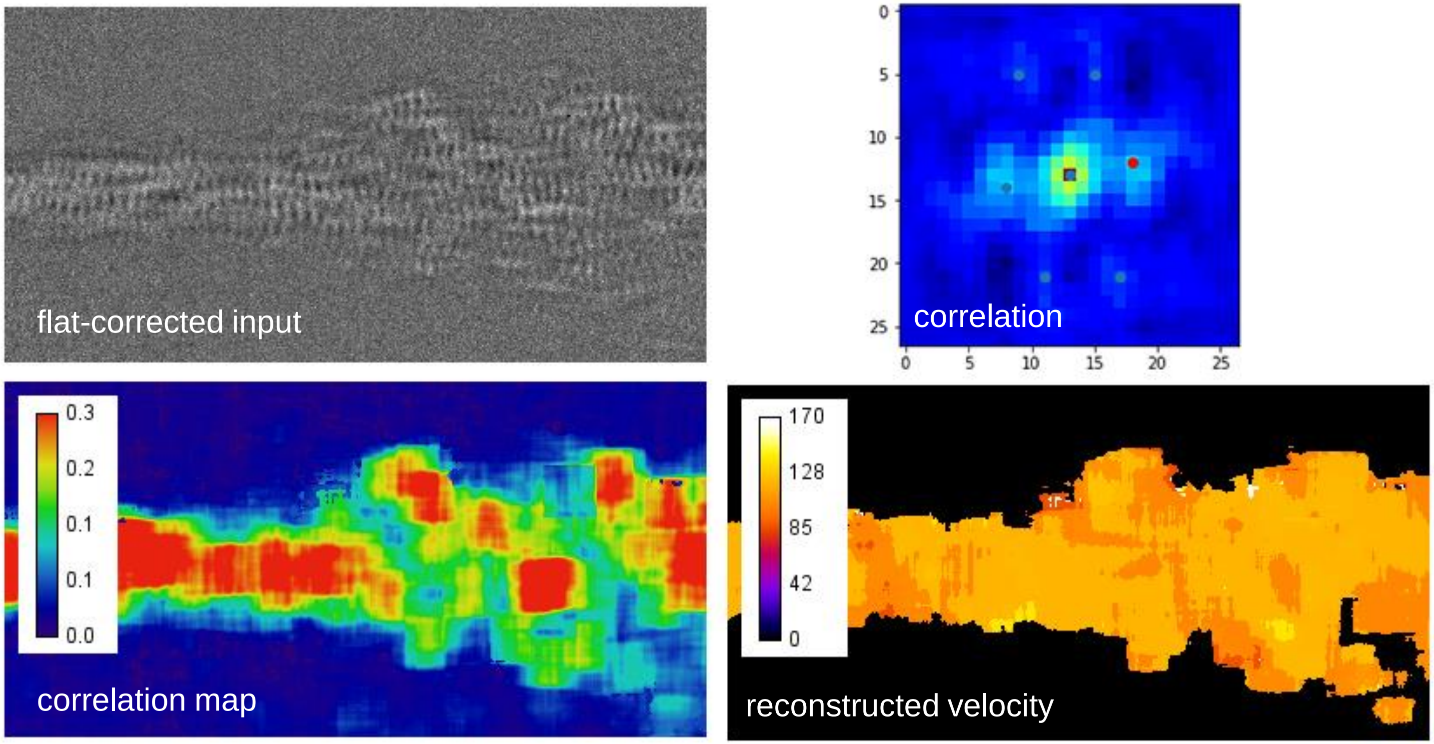
Project goal: reduce particle emission of injector-based motors by designing novel injector geometries. Highly accurate models are need

Fuel Sprays: Continuous Process



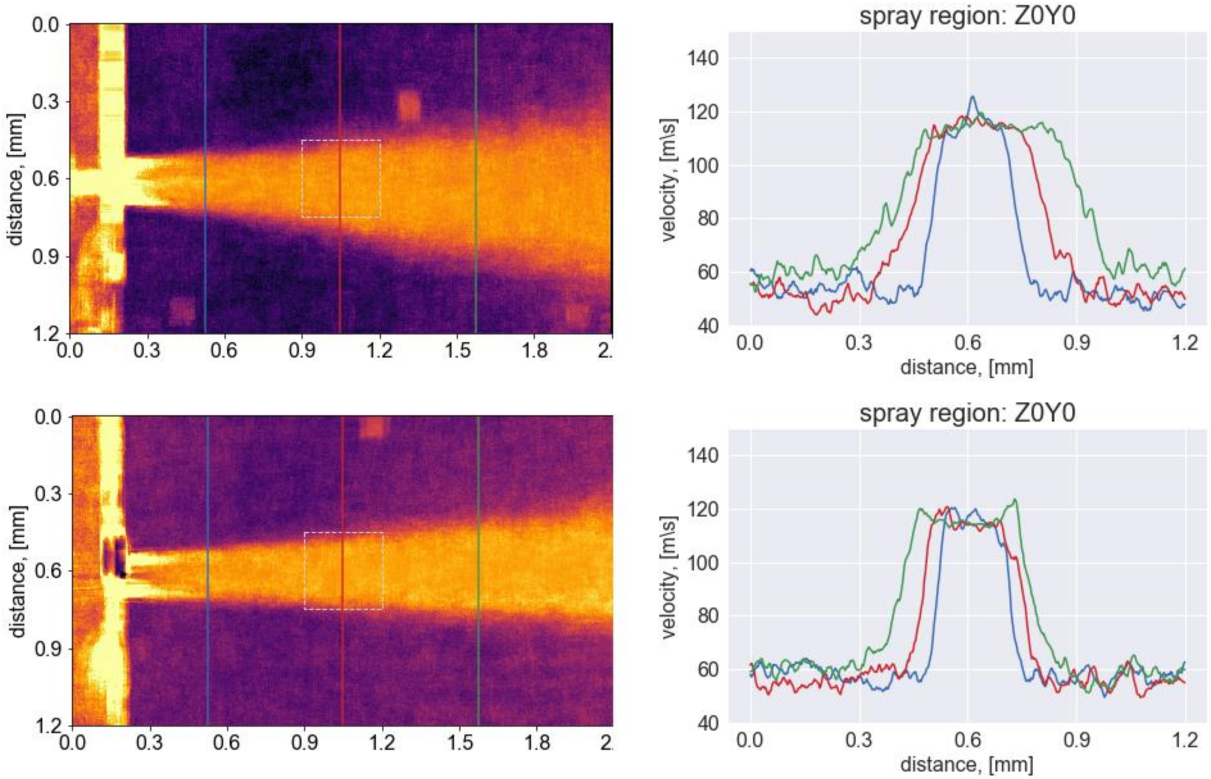
- **5 million** frames/s at ID19, ESRF
- Phase contrast
- One frame illuminated by a single photon bunch (30 -100 ps)
- **Continuous** process, **Short** sequence (128 image frames)

Fuel Sprays: Stochastic Process



- **Reduce** frame rate. Multi-exposure images
- **Discrete**, Stochastic process
- Long sequence, **whole** spraying event
- Auto correlation analysis

Fuel Sprays: Stochastic Process



- Only 2D projections, **no 3D** structure
- **Idea**: Use **Deep Learning** to reconstruct **3D shape and temporal structure** from a series of 2D projections.
- Use computational fluid dynamics (CFD) simulations for training
- Neural Networks as **approximation** to solve Inverse problems

Evaluation and Benchmarking of algorithms

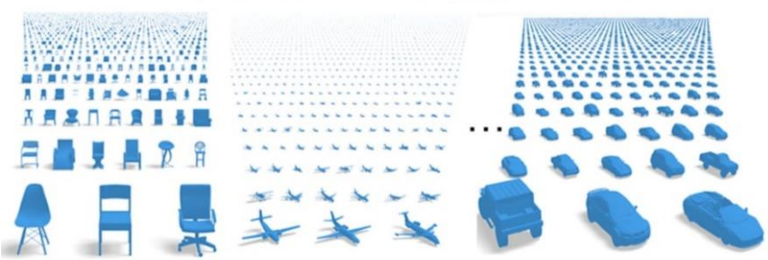
Large-scale Open Databases

Huge progress in Deep Learning is due to the establishment of **large**, **open** databases for training and benchmarking of ML algorithms.

- **ImageNet**: Visual Recognition Challenge evaluates algorithms for object detection and image classification. 20000 categories, 14M annotated images

- **ShapeNet**: large-scale dataset of 3D shape 55 object categories, 51,300 unique 3D models

and many others



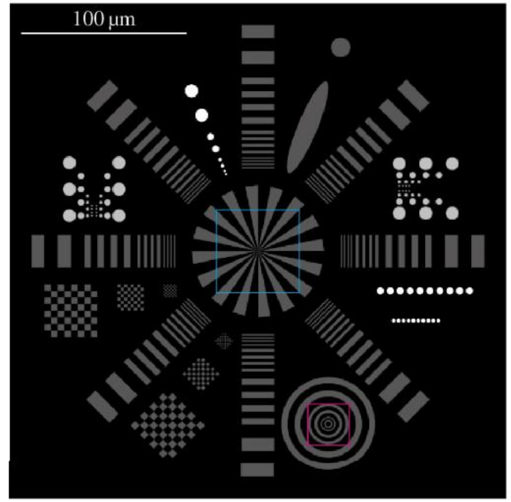
Institute for Photon Science and Synchrotron Radiation, Imaging Group

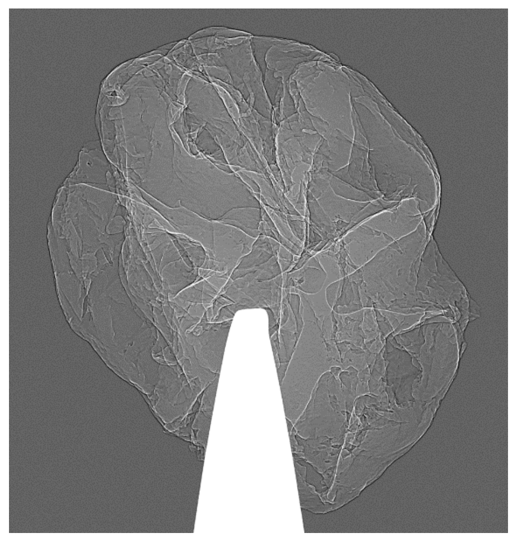
Database for X-ray Data

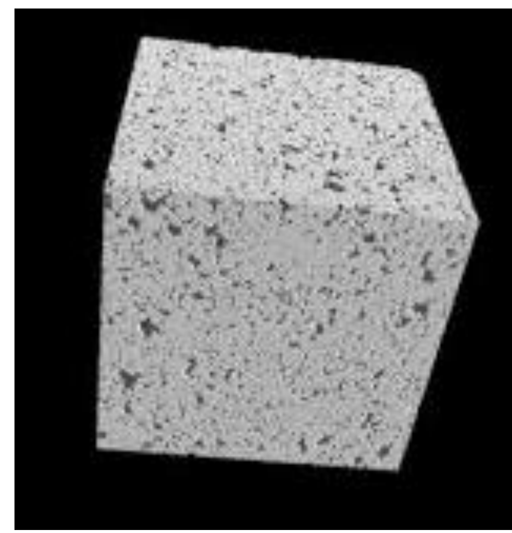
Data types:
- Real experimental data
  - Annotations
- Synthetic data
- Physical phantoms

X-ray sources:
- Synchrotrons
- Free-electron lasers
- X-ray tubes
- Liquid jets

Methods:
- 2D radiography
- 3D tomography
- 4D (3D+t) tomography







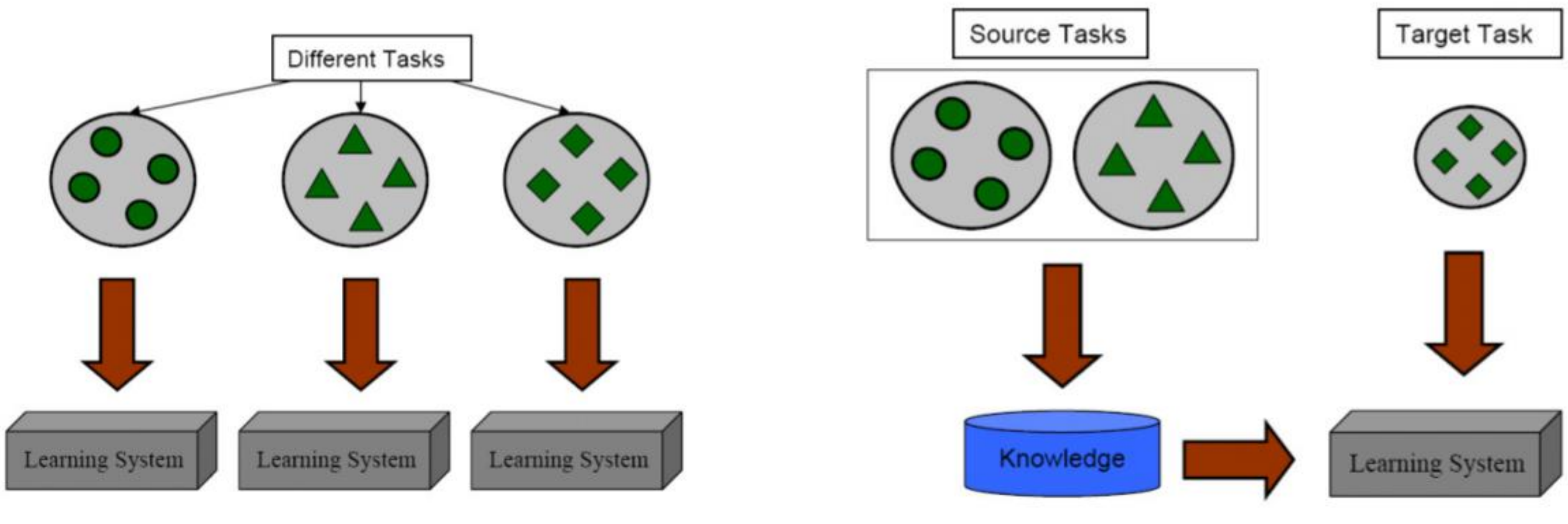
Transfer Learning and Domain adaptation

Transfer Learning / Knowledge Transfer

Problems

- Features of training and test (application) data are changing
- Lack of massive amount of labeled training data
- Expensive to perform re-training

Question: How to use previously trained models on new data?

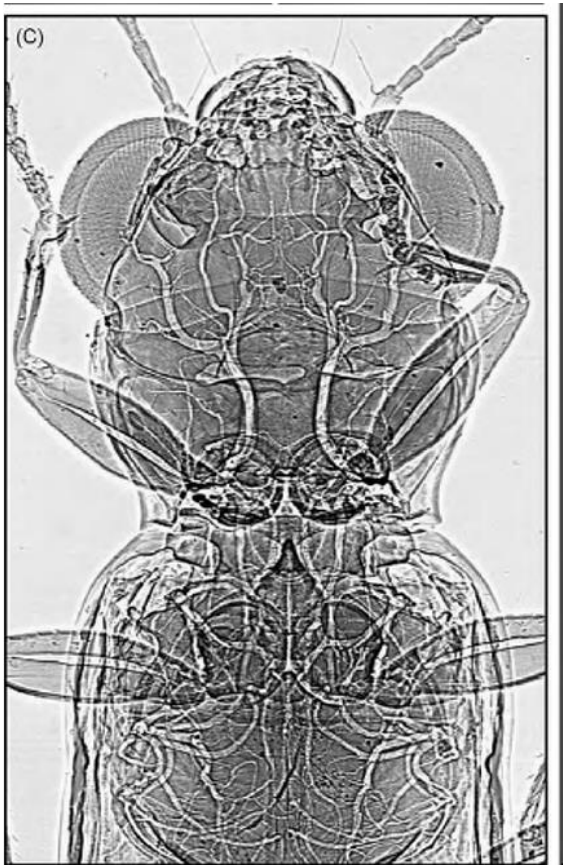


(a) Traditional Machine Learning (b) Transfer Learning

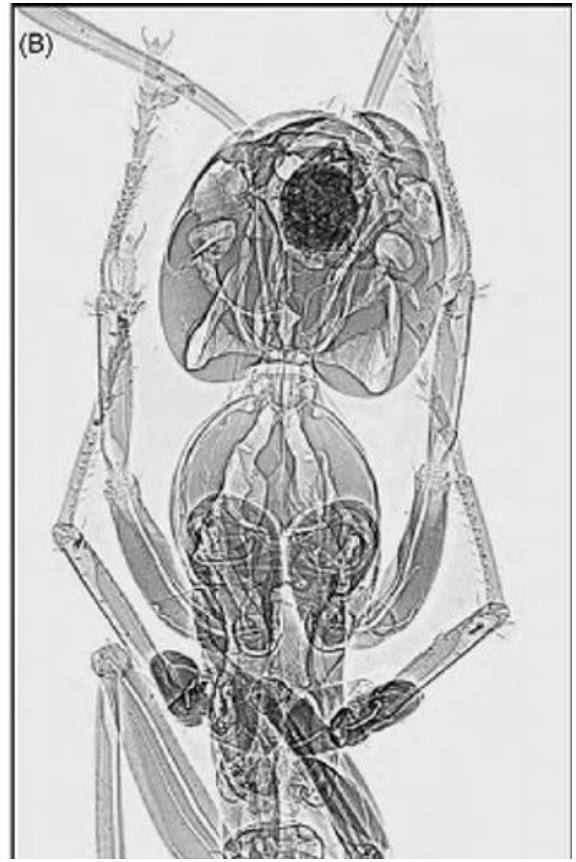
Alexey Ershov, ershov@kit.edu
Institute for Photon Science and Synchrotron Radiation, Imaging Group

Transfer Learning

Different input distribution: e.g. different specimens



Ant worker
(*Camponotus pennsylvanicus*)

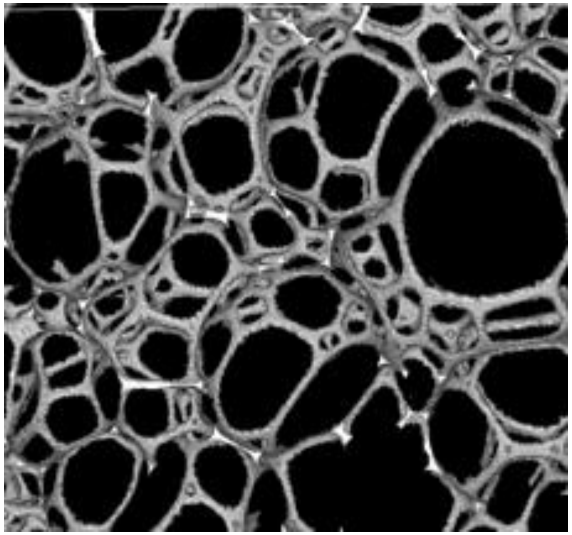


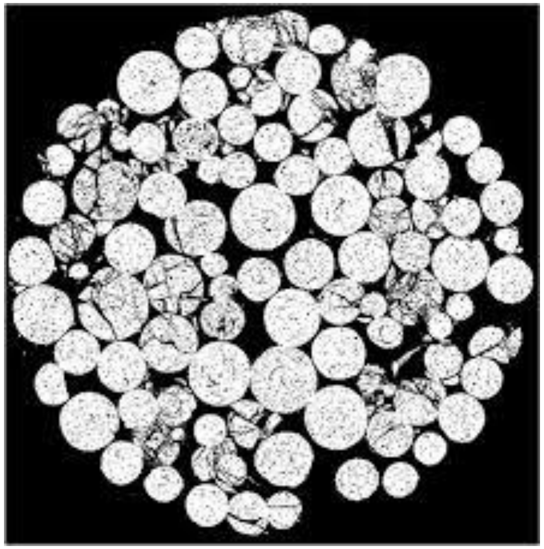
Carabid beetle
(*Notiophilus spp.*)

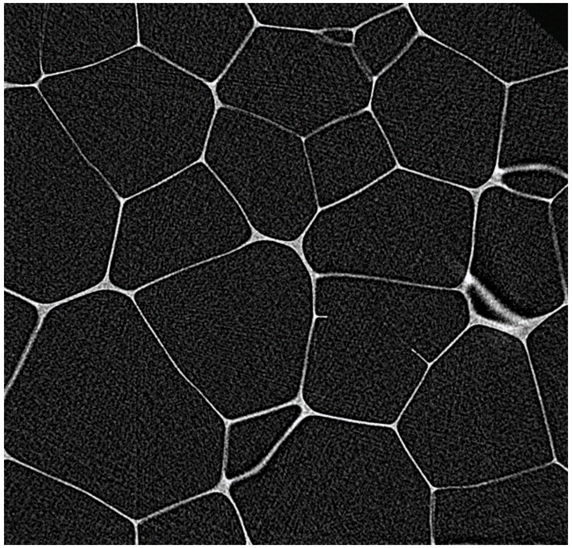
Socha, et al, 2010

Transfer Learning

Pre-training of Neural Networks on similar structures (features).
Example: cellular materials.







Alexey Ershov, ershov@kit.edu

Institute for Photon Science and Synchrotron Radiation, Imaging Group

Domain Adaptation

Question: Can we learn from other domains (experiments, methods)?

Problems

- Labeled data exists only in the **source** domain and absent in the **target**
- Data **distribution** is different, so we cannot directly apply ML methods.

Idea: Learn representation (model or features) which is both semantically meaningful and **domain invariant**.

Application examples:

- Train on synthetic data ⟺ Apply on real data
- 2D ⟺ 3D ⟺ 4D
- High resolution ⟺ Low resolution
- X-ray Tomography ⟺ MRI imaging ⟺ Ultrasound

Unsupervised Learning

Unsupervised Learning

All Machine learning techniques could be classified as:

- Fully supervised (Deep learning)
- Semi-supervised (Generative Models, Transfer Learning, Domain Adaptation)
- **Unsupervised**: Learn from unseen data or data without predefined labels

Different methods of Unsupervised Learning:

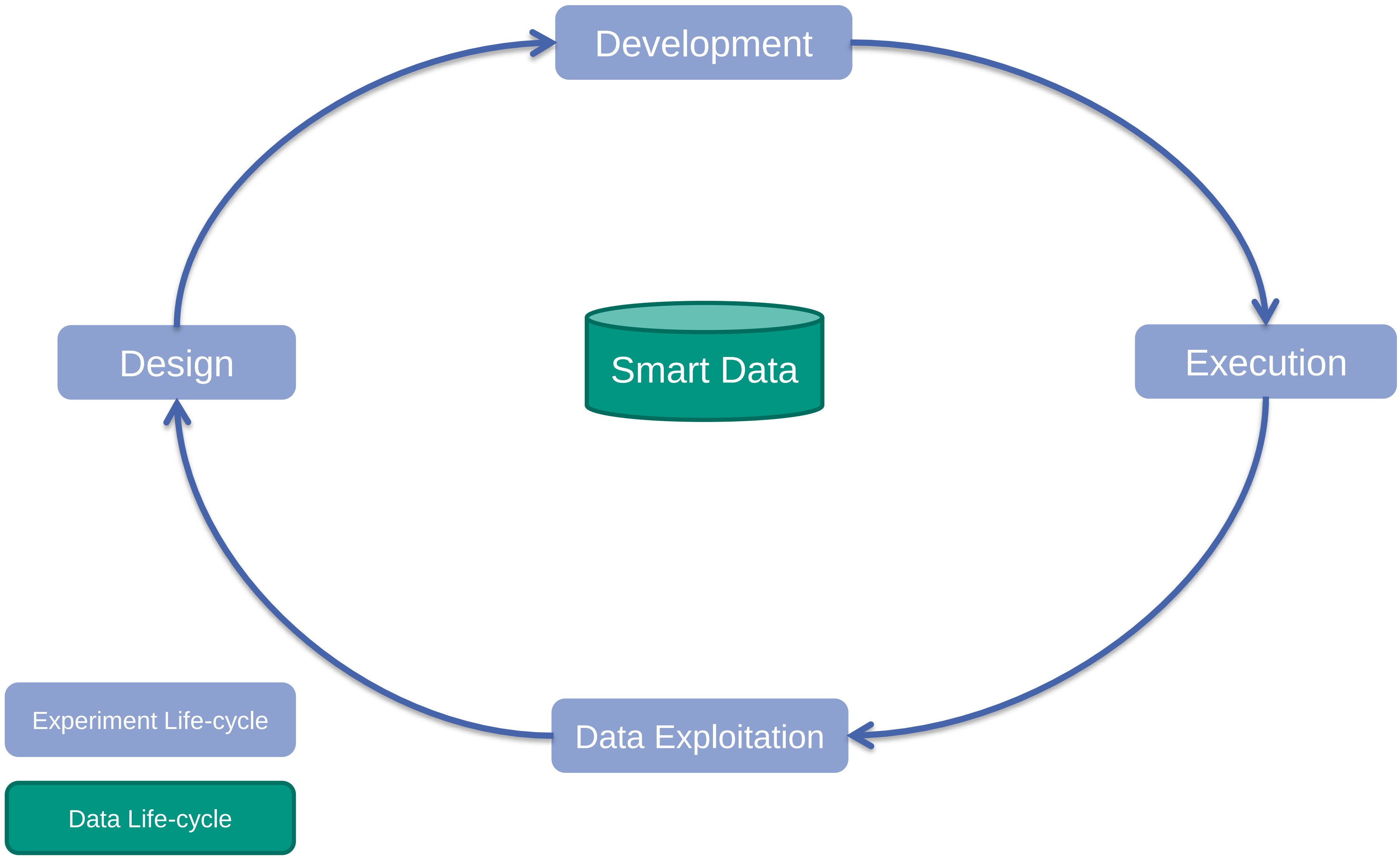
- Automated Machine Learning
- Anomaly Detection
- Model searching / Model Query
- Self-learning system, Meta-learning, Learning to learn

Data and Experiment Lifecycle

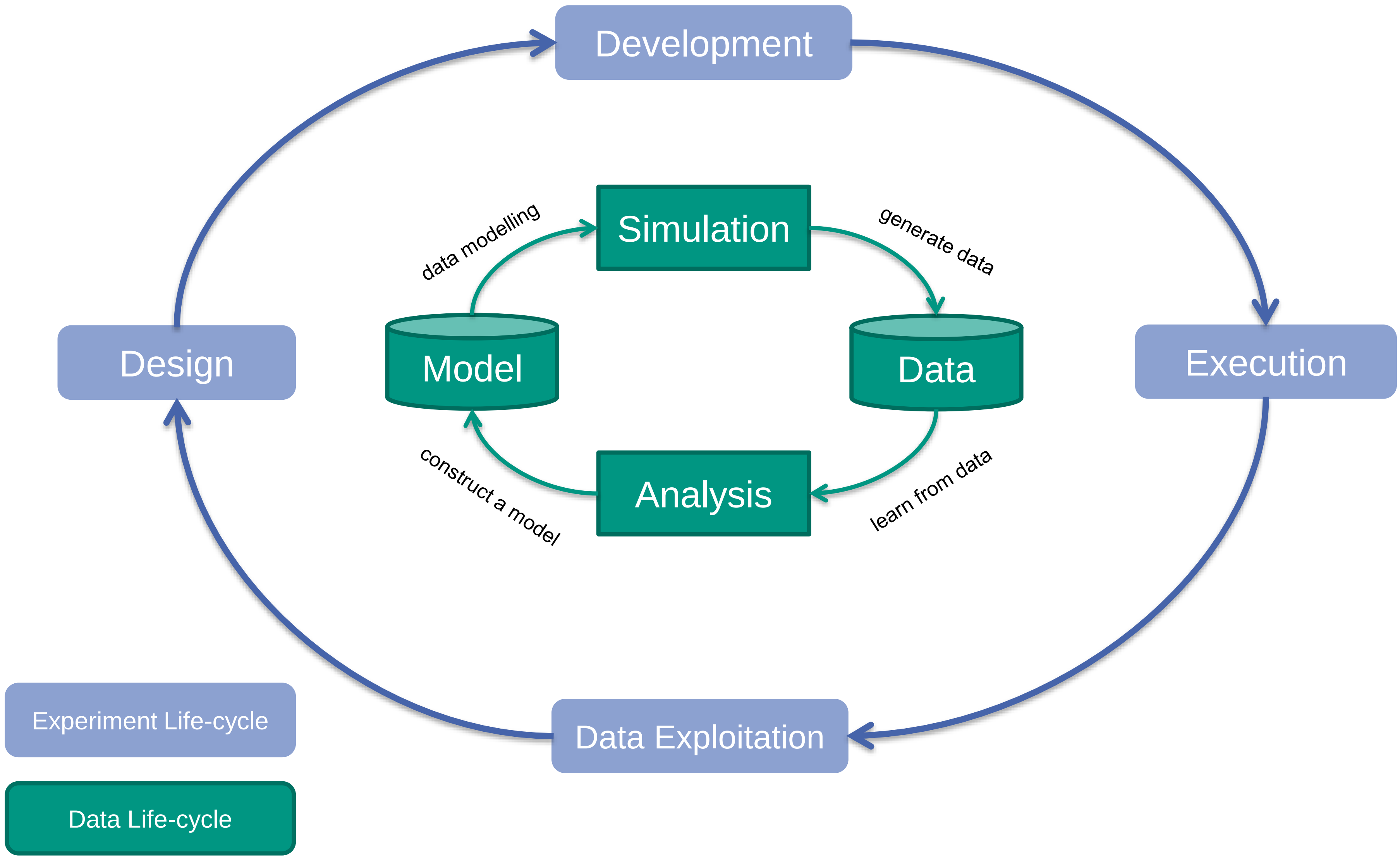
Alexey Ershov, ershov@kit.edu

Institute for Photon Science and Synchrotron Radiation, Imaging Group

Experiment and Data Lifecycles

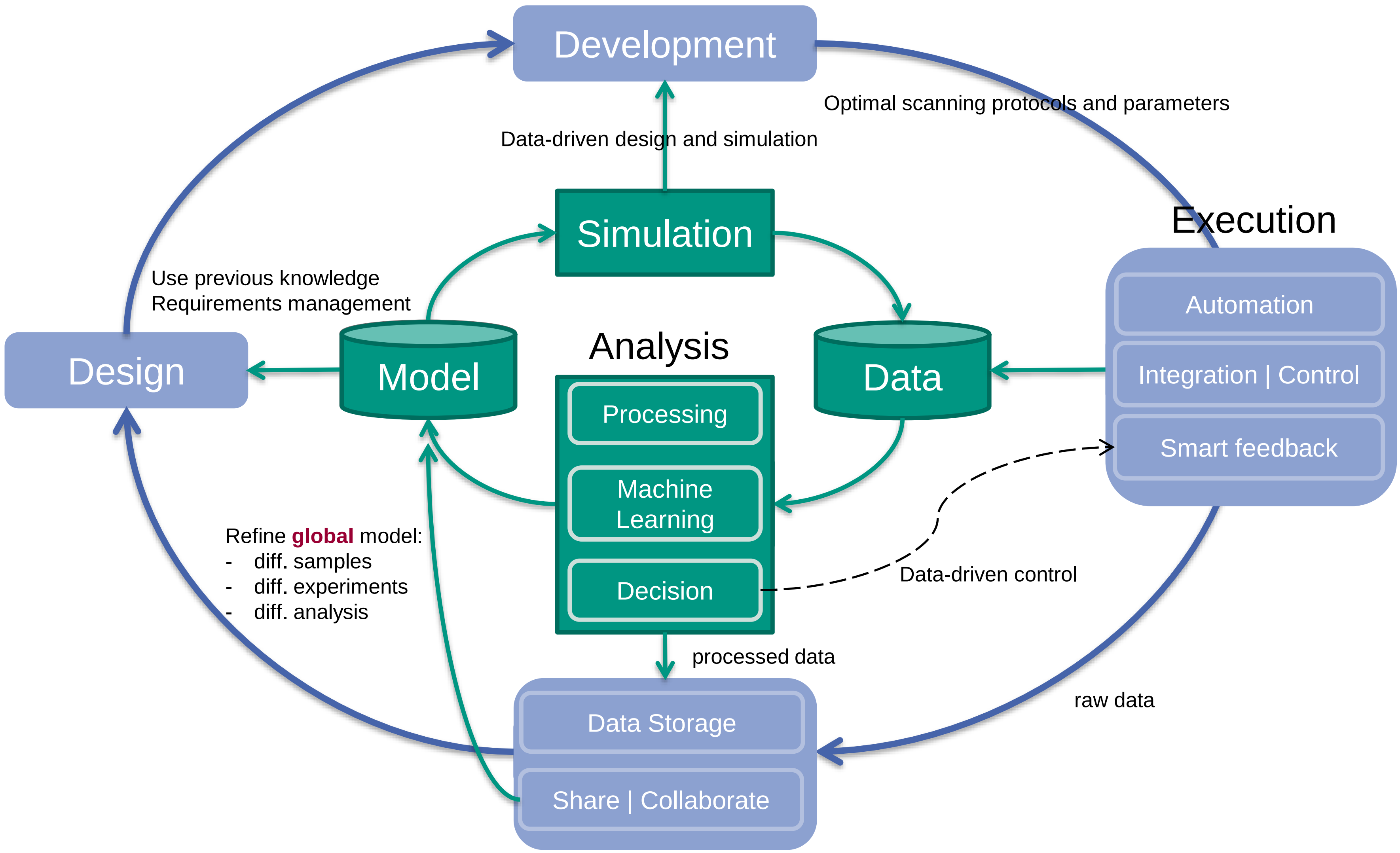


Experiment and Data Lifecycles

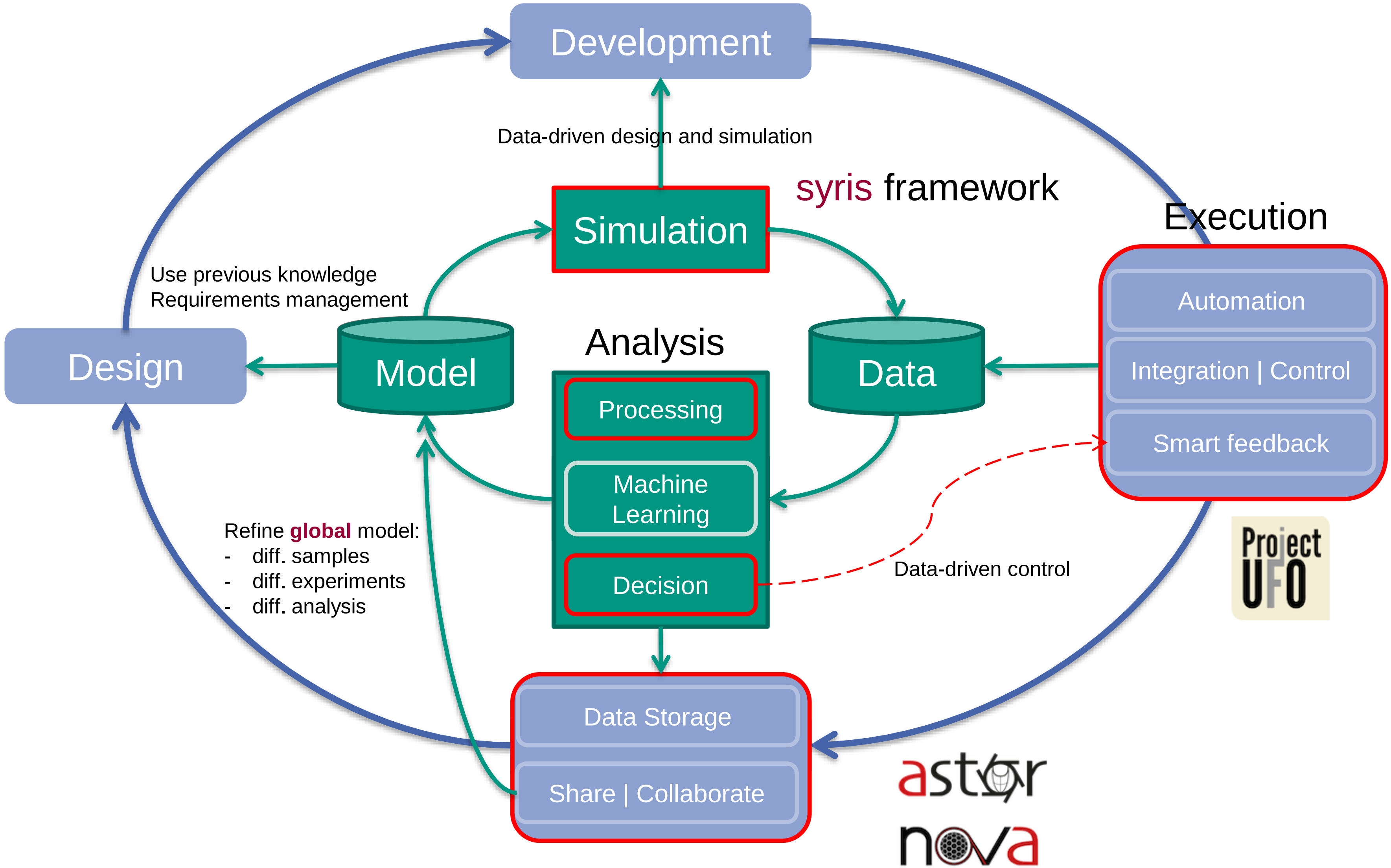


Institute for Photon Science and Synchrotron Radiation, Imaging Group

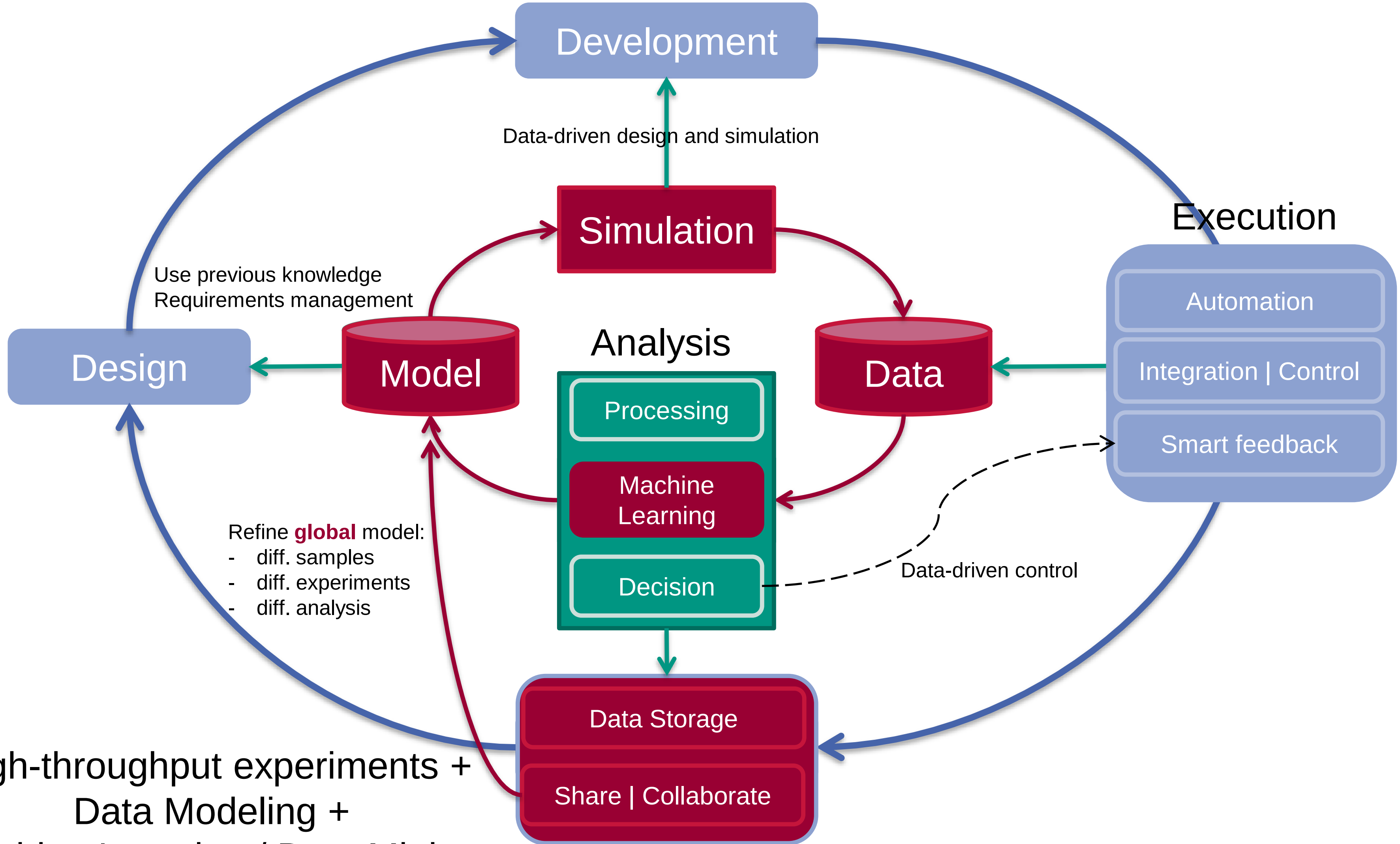
Experiment and Data Lifecycles



Experiment and Data Lifecycles



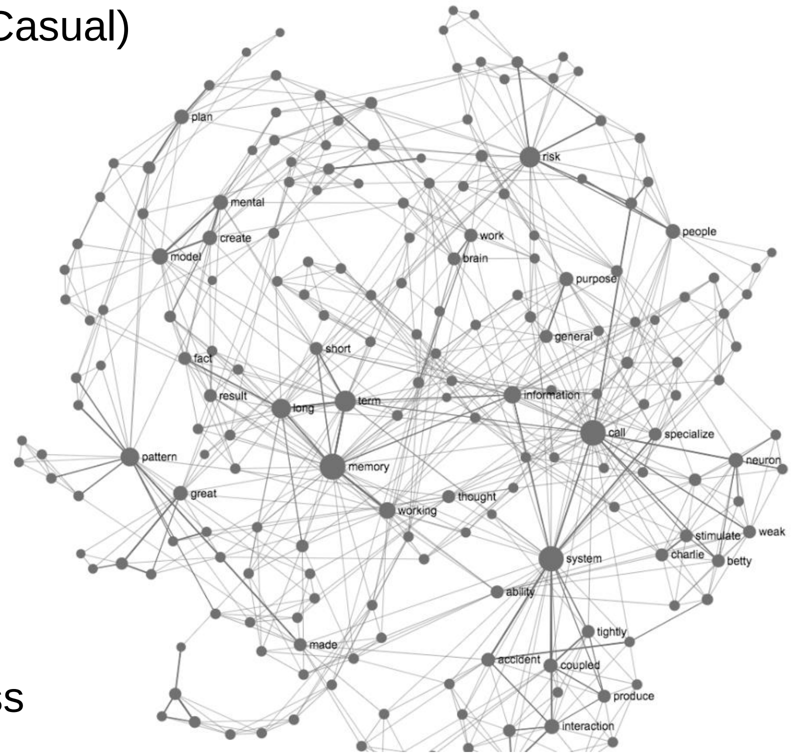
Crucial components for Digital Transformation



High-throughput experiments +
Data Modeling +
Machine Learning / Data Mining

Universal Knowledge Base

Contains:
- Descriptive models (Mathematical, Statistical, Casual)
- Experiments
- Simulation (generative models)
- Raw data, Simulated data
- Processing workflows and data views / states
- Machine learning algorithms
- Evaluations



Properties of Models:
- Abstract
- Unified
- Global
- Flexible
- Extensible
- Versioning (states)

Everything here is:
- Unified / Connected
- Online / Open Access
- Can be verified
- Can be build upon
- Explored (search)
- Modified
- Learned from

Summary

- Step-by-step approach for adoption of Machine Learning methods
- **Deep Learning** for high-throughput and most common tasks
  - Domain knowledge is crucial. Experts should learn to apply technology. Interdisciplinary research
  - Teaching within community: Courses, Workshops, Tutorials
- **Simulation**:
  - Best way to train Machine Lerning
  - Unified forward simulation framework for each community
- **Databases**: Large scale open databases, managed by consortium
- **Smart Data**: Models, Data, Algorithms -> Universal Knowledge Base
- Machine Learning:
  - **Transfer Learning and Domain Adaptation** – Extend ML to new data and applications
  - **Unsupervised Learning / Self-learning –** Ultimate goal

Thank you for your attention!