Helmholtz AI Consulting for matter research at HZDR

Peter Steinbach, Helene Hoffmann, Neda Sultova, Mahnoor Tanveer, David Pape, Steve Schmerler, Sebastian Starke

HZDR / ML@HZDR, Dec 6, 2021

Helmholtz AI

What is Helmholtz AI?

- initiative by President of the Helmholtz Association, Prof. Otmar D. Wiestler

What is Helmholtz AI?

- initiative by President of the Helmholtz Association, Prof. Otmar D. Wiestler
- running over 7 years, 2019 - 2026

What is Helmholtz AI?

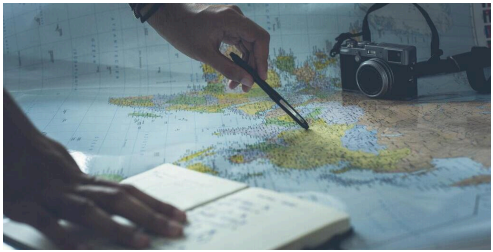
- initiative by President of the Helmholtz Association, Prof. Otmar D. Wiestler
- running over 7 years, 2019 - 2026
- 12 M€ per year
  (people + projects)

What is Helmholtz AI?

- initiative by President of the Helmholtz Association, Prof. Otmar D. Wiestler
- running over 7 years, 2019 - 2026
- 12 M€ per year (people + projects)
- central installation in Munich (universities and Helmholtz center)

Two Funding Lines

Helmholtz AI Projects



unsplash.com:Glenn Carstens-Peters

- current call open until Dec 1, 2021
- max. 3 years, max. 200k € (must be matched)

Helmholtz AI Vouchers

unsplash.com:Dominik Scythe

- voucher submissions open anytime
- get in touch first: consultant-helmholtz.ai@hzdr.de

Helmholtz AI Local Unit For Matter At HZDR

Figure: Nico Hoffmann, YIG Lead

Figure: Peter Steinbach, Consultant Lead

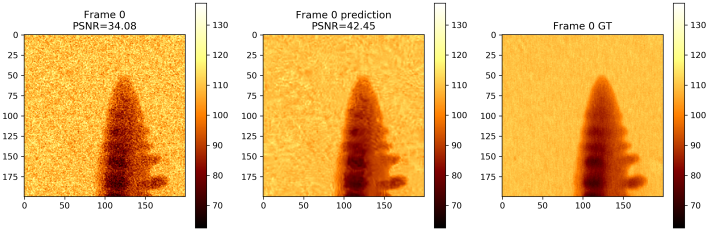
Helmholtz AI Consultant Team at HZDR

- **reproducible automated** (ML) pipelines
- **inverse problems** & generative modelling
- (image) **denoising**
- **anomaly detection**
- regression & pattern recognition (object localisation, **image segmentation**)
- aspects of trustworthy ML (**uncertainties**, **robustness** and interpretability)

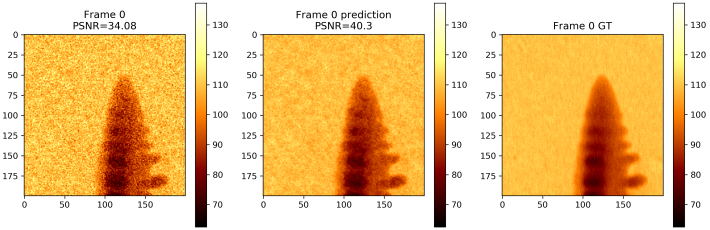
Past and Present Vouchers

Denoising Radiograms

BM3D state-of-the-art image denoising



PN2V/PPN2V semi-supervised/unsupervised image denoising



Anomaly Detection at EUXFEL

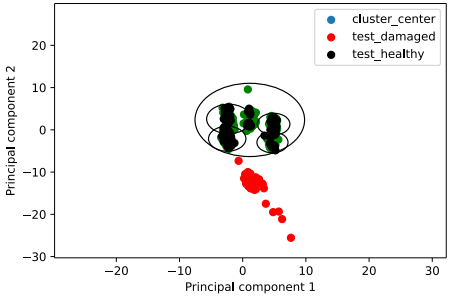
Feature extraction:

- requirement: low dimensional representation of time-courses
- here: used `tsfresh`-package
  - simple features: mean, min, max, ...
  - more sophisticated features: fft-coefficients, entropy, absolute energy, ...

Clustering:

- Principal Component Analysis
- kNN-Clustering

Results:



Object Segmentation in bubbly flows

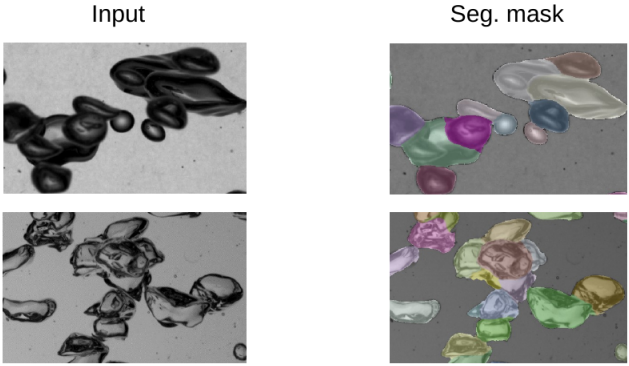
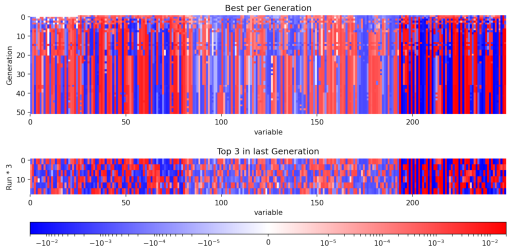


Figure: Using StarDist for object segmentation

- Experimental investigations of bubbly flows
  - optimize industrial process (energy/chemical/metallurgical engineering)
- Challenges:
  - Identify/segment overlapping bubbles
  - Different image conditions (image quality, bubble size/shape, bubble number density...)

COSY simulation: Evolutionary algorithm optimization



goal: improve simulation with 1500 input parameters
(simulation output $\hat{\mathbb{Y}}$: 1 Orbit Response Matrix (ORM) with 3149 entries)

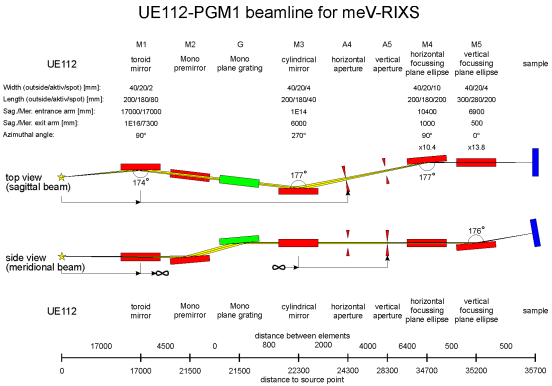
- `deap` framework `https://github.com/DEAP/deap`, uses "population" of possible x'
- **many params $x_i'^*$ hit their allowed range limits**
- Suggestion:
  - repeat runs: very similar C^* but different x'^* (similar to neural network optimization!)
  - our suggestion: improve population initialization, run monitoring (convergence behavior), ...
  - possible: loss landscape analysis (many optima, tune EA exploration behavior based on that)

Tooling landscape for ML ops ...

Figure: taken from medium blog post

- multiple open-source tools exist (dvc, mlflow, ...)
- which one to choose?
- can it be deployed on academic HPC infrastructure?
- outcome:
  - example workflows
  - extensive final report weighing several aspects

Inverting a beamline simulation at BESSY …



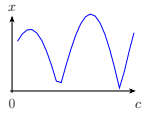
$$\vec{\vartheta} = \begin{pmatrix} 42. \\ 12. \\ 1. \\ -0.8 \\ 2. \\ 320 \end{pmatrix}$$

simulation

$$\vec{x} = f_{sim}(\vec{\vartheta})$$

inversion?

$$\vec{\vartheta} = f_{sim}^{-1}(\vec{x})$$

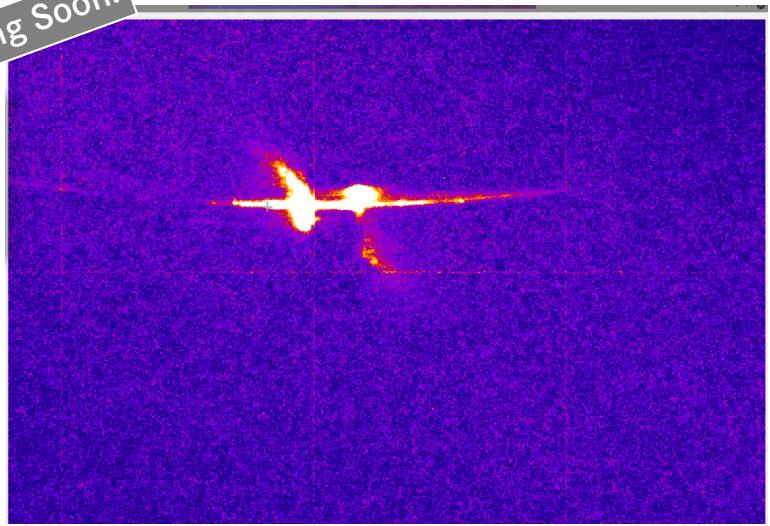


$$\begin{pmatrix} .50 \\ .60 \\ .67 \\ \cdots \\ \cdots \\ \cdots \\ .40 \\ .58 \\ .75 \end{pmatrix} = \vec{x}$$

- **goal**: given a beamline profile (knife-edge scan), which beam control properties would result in this profile

Future: Background Estimation in HIBEF based SAXS data

Coming Soon!



courtesy of Michal Smid (FWKT / HZDR)

Future: Anomaly Detection at Belle2 with KIT

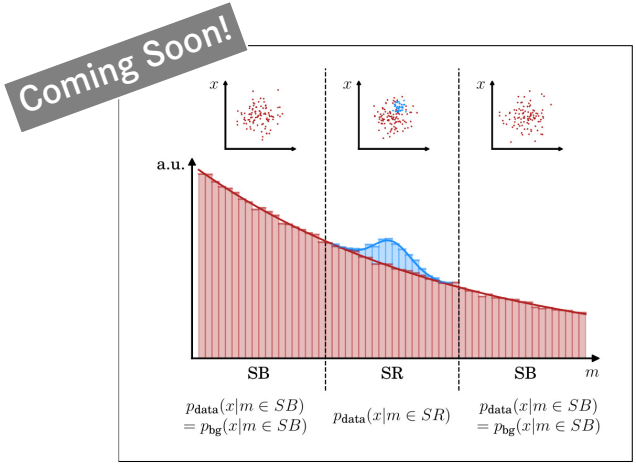


FIG. 1. Schematic view of the bump hunt. The signal (blue) is localized in the signal region (SR). The background (red) is estimated from a sideband region (SB).

from [1]

Lessons Learned

Machine Learning needs a clear goal!

- narrow AI can aid in many tasks in science and society

Machine Learning needs a clear goal!

- narrow AI can aid in many tasks in science and society
- needs mediation between data produces, consumers and ML consultants (communication)

Machine Learning needs a clear goal!

- narrow AI can aid in many tasks in science and society
- needs mediation between data produces, consumers and ML consultants (communication)
  - learn from others

Machine Learning needs a clear goal!

- narrow AI can aid in many tasks in science and society
- needs mediation between data produces, consumers and ML consultants (communication)
  - learn from others
  - get (professional) perspectives

Machine Learning needs a clear goal!

- narrow AI can aid in many tasks in science and society
- needs mediation between data produces, consumers and ML consultants (communication)
  - learn from others
  - get (professional) perspectives
  - enter unfamiliar domains

Machine Learning needs a clear goal!

- narrow AI can aid in many tasks in science and society
- needs mediation between data produces, consumers and ML consultants (communication)
  - learn from others
  - get (professional) perspectives
  - enter unfamiliar domains
- academic consulting at scale is something quite new in (Helmholtz) science

Machine Learning needs a clear goal!

- narrow AI can aid in many tasks in science and society
- needs mediation between data produces, consumers and ML consultants (communication)
  - learn from others
  - get (professional) perspectives
  - enter unfamiliar domains
- academic consulting at scale is something quite new in (Helmholtz) science
- reproducibility, training and openness as inroads to trustworthiness

FAIR as a basis for ML (credits David Pape / HZDR)

Findable

- use a public repository
- obtain unique global ID
- enrich metadata

Accessible

- nobody to ask
- automated retrieval: data and metadata can be obtained by a freely implemented protocol

Interoperable

- document based on standards (SI, datacite, ...)
- use established machine-readable formats (yaml, json, hdf5, tiff, ...)

Reusable

- Choose a license!
- data meets community standards (description, i/o libraries, ...)

Automate and document the above as soon as possible!

Summary

Conclusion

- Helmholtz AI open for projects

Conclusion

- Helmholtz AI open for projects
- already learned a lot about ML consulting projects

Conclusion

- Helmholtz AI open for projects
- already learned a lot about ML consulting projects
- ML needs a concrete goal (expectations, testability, trustworthiness)

Conclusion

- Helmholtz AI open for projects
- already learned a lot about ML consulting projects
- ML needs a concrete goal (expectations, testability, trustworthiness)
- ML works well on a FAIR dataset
- ML consultants established and happy to collaborate/support

Conclusion

- Helmholtz AI open for projects
- already learned a lot about ML consulting projects
- ML needs a concrete goal (expectations, testability, trustworthiness)
- ML works well on a FAIR dataset
- ML consultants established and happy to collaborate/support

Questions, Comments, Feedback or Concerns are highly welcome!

References

References I

[1] Anna Hallin, Joshua Isaacson, Gregor Kasieczka, Claudius Krause, Benjamin Nachman, Tobias Quadfasel, Matthias Schlaffer, David Shih, and Manuel Sommerhalder. Classifying anomalies through outer density estimation (cathode). *arXiv preprint arXiv:2109.00546*, 2021.