Analytics4NN: Accelerating Neural Architecture Search through Modeling and High-Performance Computing Techniques

Georgia Channing*, Ria Patel*, Paula Olaya*, Ariel Keller Rorabaugh*, Silvina Caino-Lores*,
Osamu Miyashita[ǂ], Catherine Schuman*, Florence Tama[ǂ†], and **Michela Taufer***
**University of Tennessee, Knoxville, USA*
[ǂ]Center for Computer Science, RIKEN, Kobe, Japan
[†]Nagoya University, Nagoya, Japan

Joint Laboratory for Extreme-Scale Computing

GLOBAL COMPUTING LAB

RIKEN

THE UNIVERSITY OF TENNESSEE KNOXVILLE

NSF

IBM

Supported by NSF award #2103845, OLCF CSC427 allocation, and IBM Global University Program

Analytics for Neural Networks (A4NN)

Design efficient workflows to reduce the **time and resources** required to generate accurate and efficient neural network architectures

Analytics for Neural Networks

University of Tennessee


Georgia Channing


Paula Olaya


Silvina Caino-Lores


Ria Patel


Ariel Rorabaugh


Seoyoung An


Michela Taufer

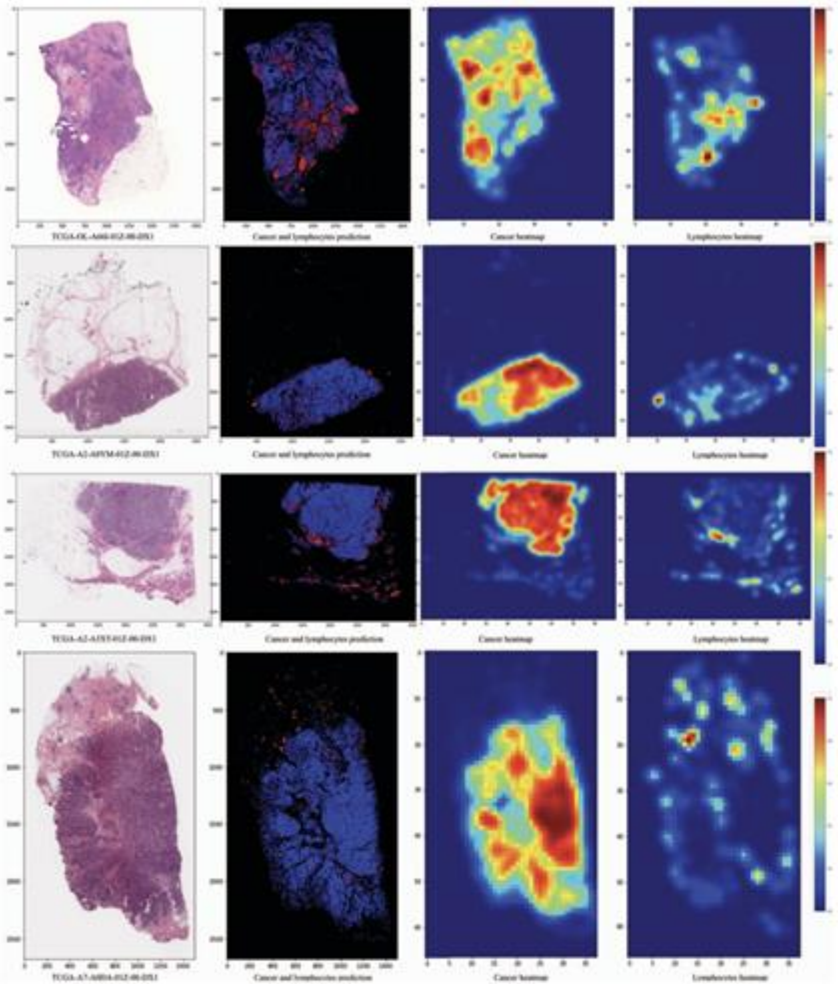

Catherine Schuman

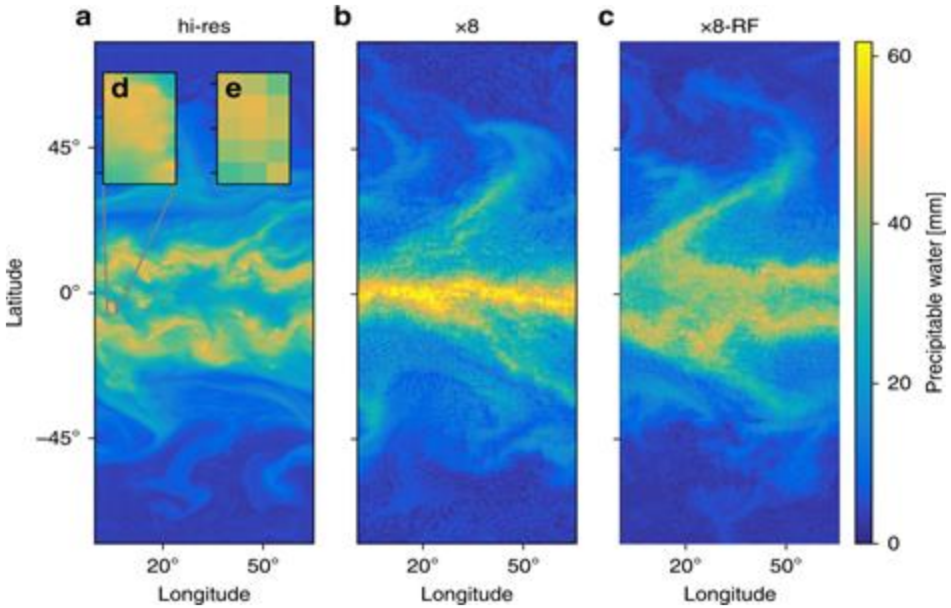
RIKEN

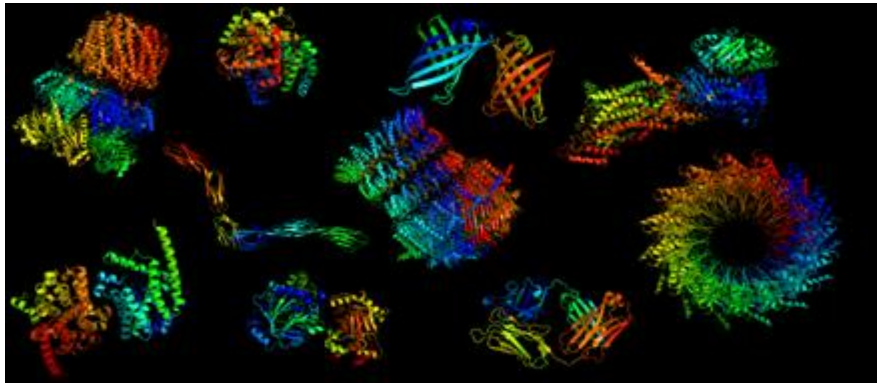

Osamu Miyashita

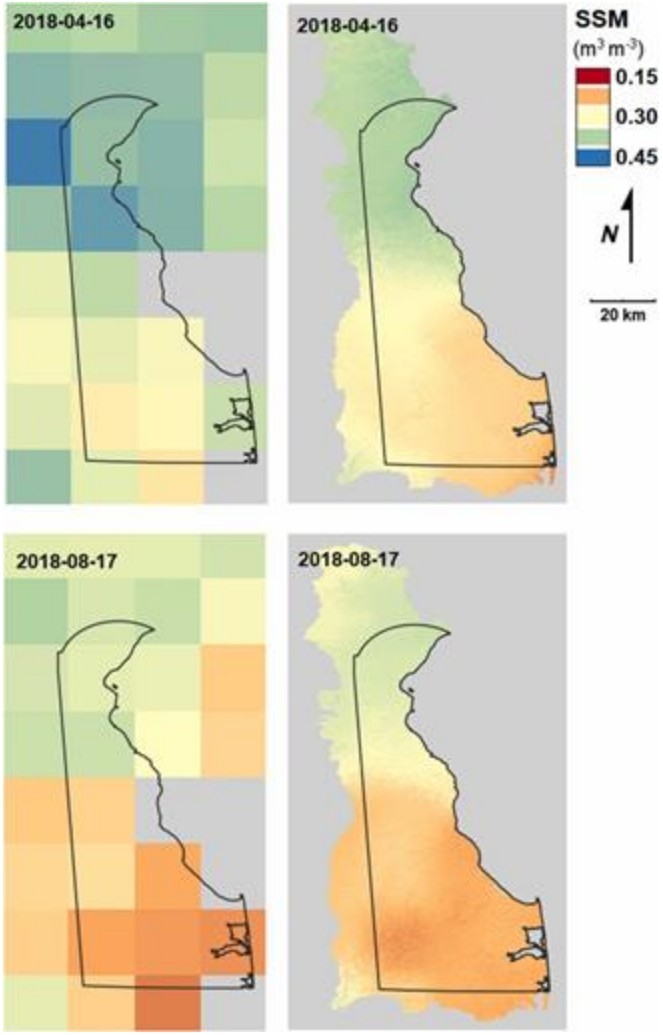

Florence Tama

An Explosion of Scientific Data from Experiments



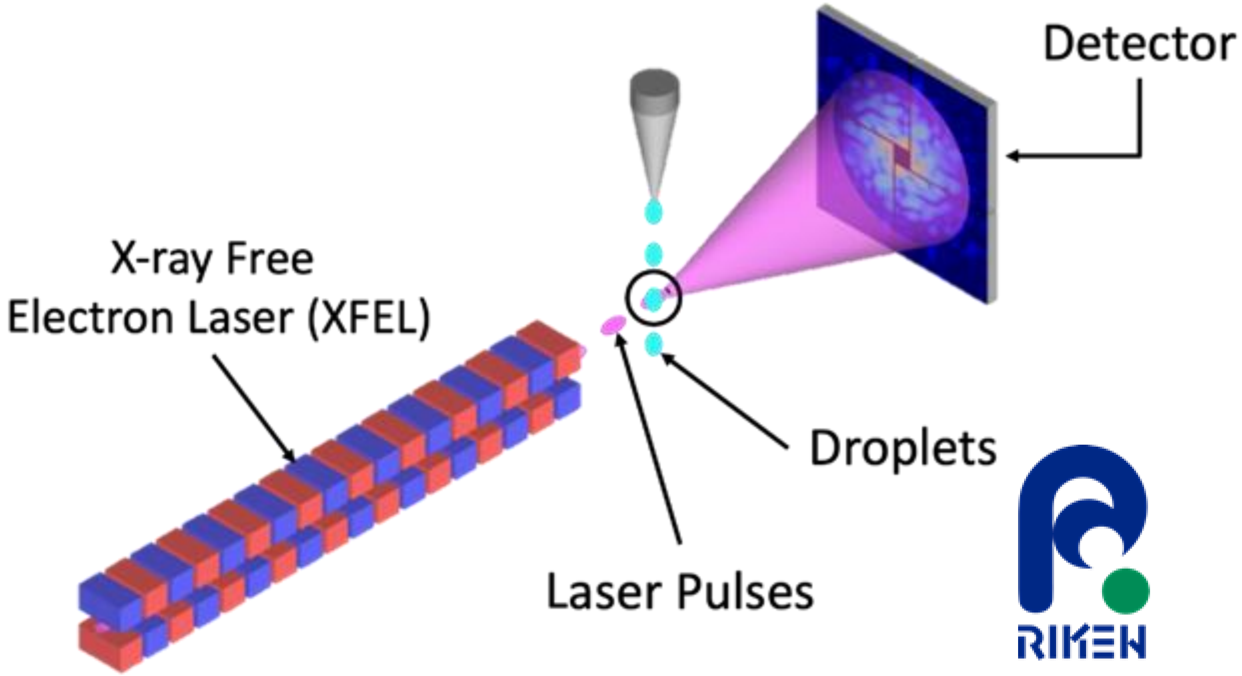


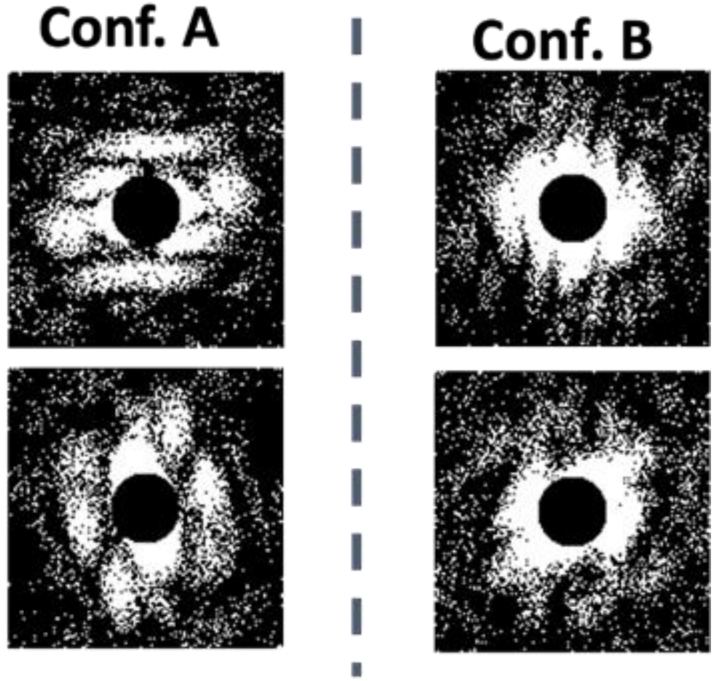




Generating Data through XFEL Experiments

X-ray Free Electron Laser (XFEL) beams create 2D protein diffraction (PD) patterns that reveal properties of the 3D protein structure

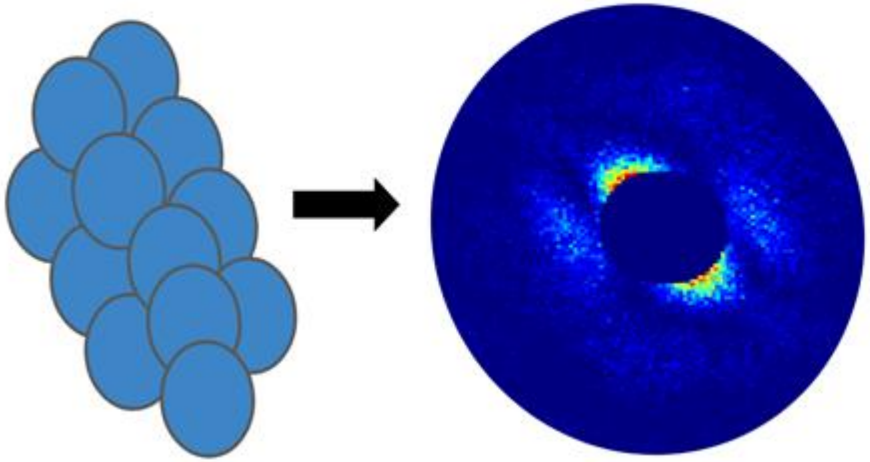




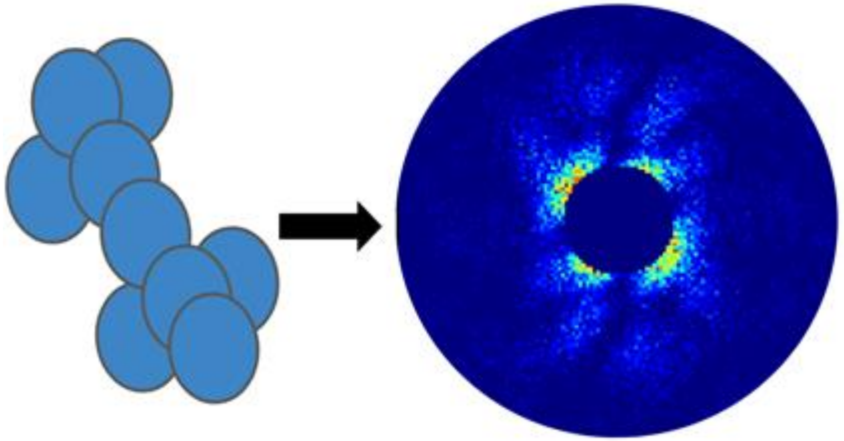
The experiments generate large image datasets which embed protein structures

Extracting Conformational Structural Properties

Conformation is the **shape adopted by a protein** and is caused by the rotation of the protein atoms around one or more single bonds



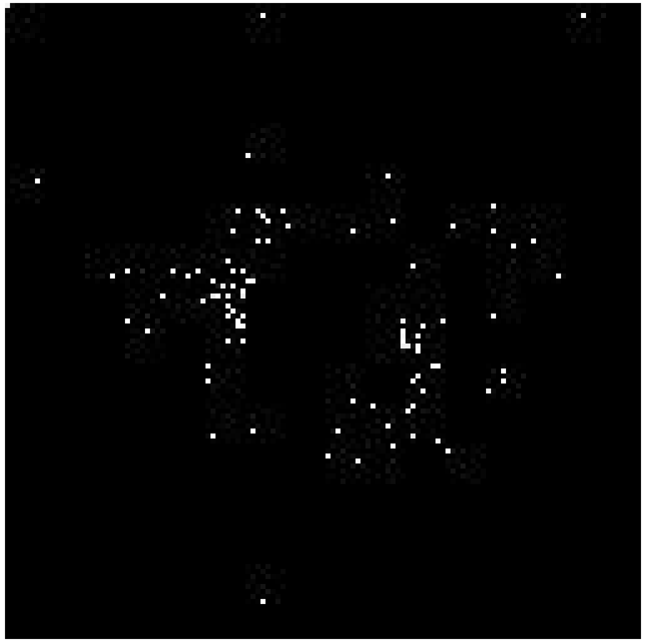
Conformation A
$\Phi, \theta, \Psi = 24^{o}, 151^{o}, 346^{o}$



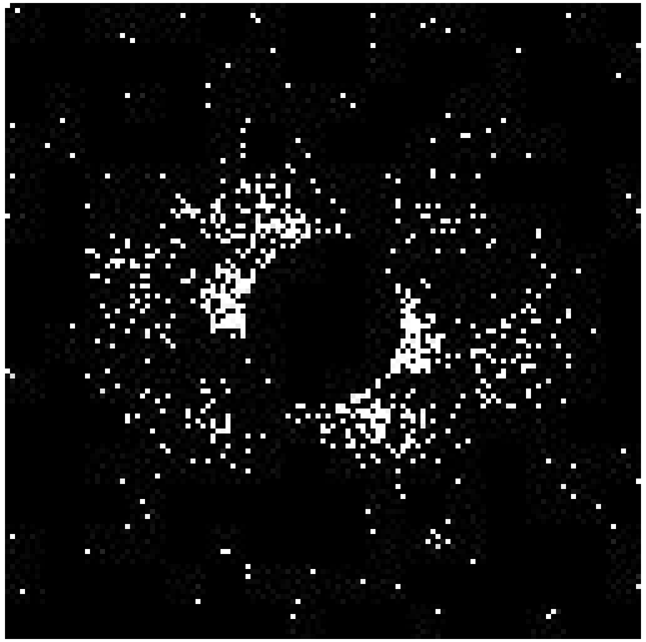
Conformation B
$\Phi, \theta, \Psi = 34^{o}, 139^{o}, 106^{o}$

Different Beam Intensity and Noise in XFEL Images

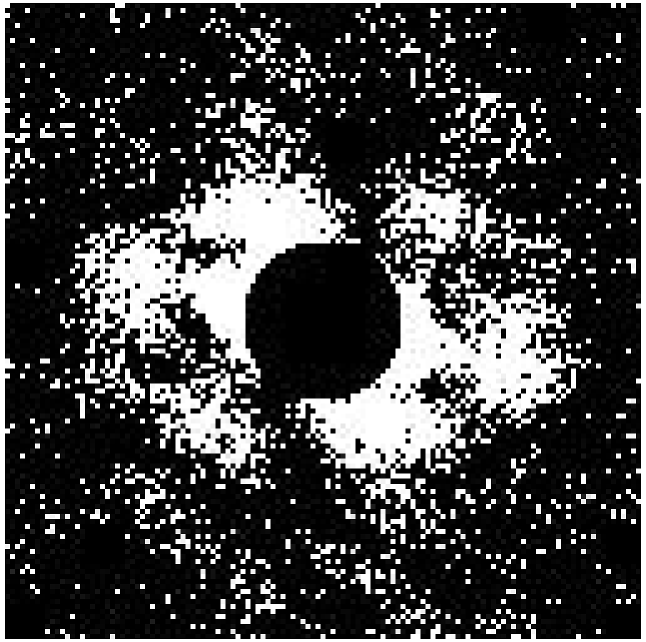
- Images have different granularity depending on the beam intensity
- Different intensities embed different amounts of noise



Low Beam Intensity
1 x 10^{14} photons/μm
/pulse



Medium Beam Intensity
1 x 10^{15} photons/μm
/pulse

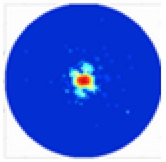


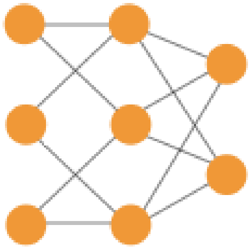
High Beam Intensity
1 x 10^{16} photons/μm
/pulse

Using Neural Networks and Neural Architecture Search

Neural Networks (NN) can extract information from scientific data

Protein images





Protein properties:

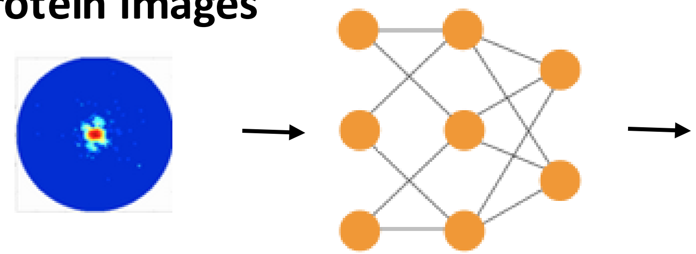
- Protein type
- Orientation
- Structure

Olaya *et al*, "Identifying Structural Properties of Proteins from X- ray Free Electron Laser Diffraction Patterns". *In Proceedings of the 18th IEEE International Conference on e-Science (eScience), pages 1–10, Salt Lake City, Utah, USA, October 2022. IEEE Computer Society*.

Using Neural Networks and Neural Architecture Search

Neural Networks (NN) can extract information from scientific data

Protein images



Protein properties:

- Protein type
- Orientation
- Structure

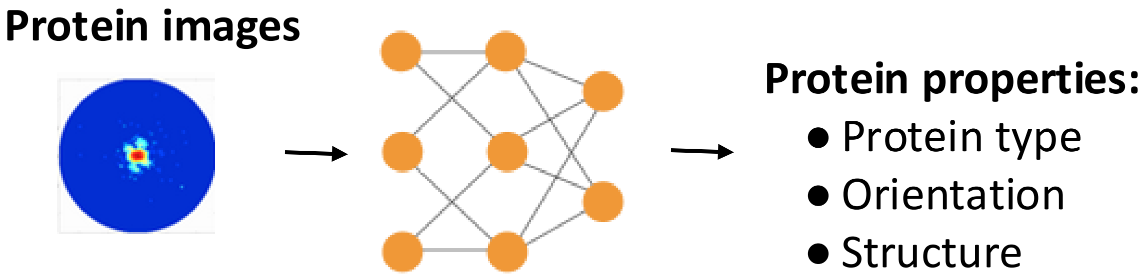
Custom NNs are needed for each dataset

Olaya *et al*, "Identifying Structural Properties of Proteins from X- ray Free Electron Laser Diffraction Patterns". *In Proceedings of the 18th IEEE International Conference on e-Science (eScience), pages 1–10, Salt Lake City, Utah, USA, October 2022. IEEE Computer Society*.

THE UNIVERSITY OF TENNESSEE KNOXVILLE

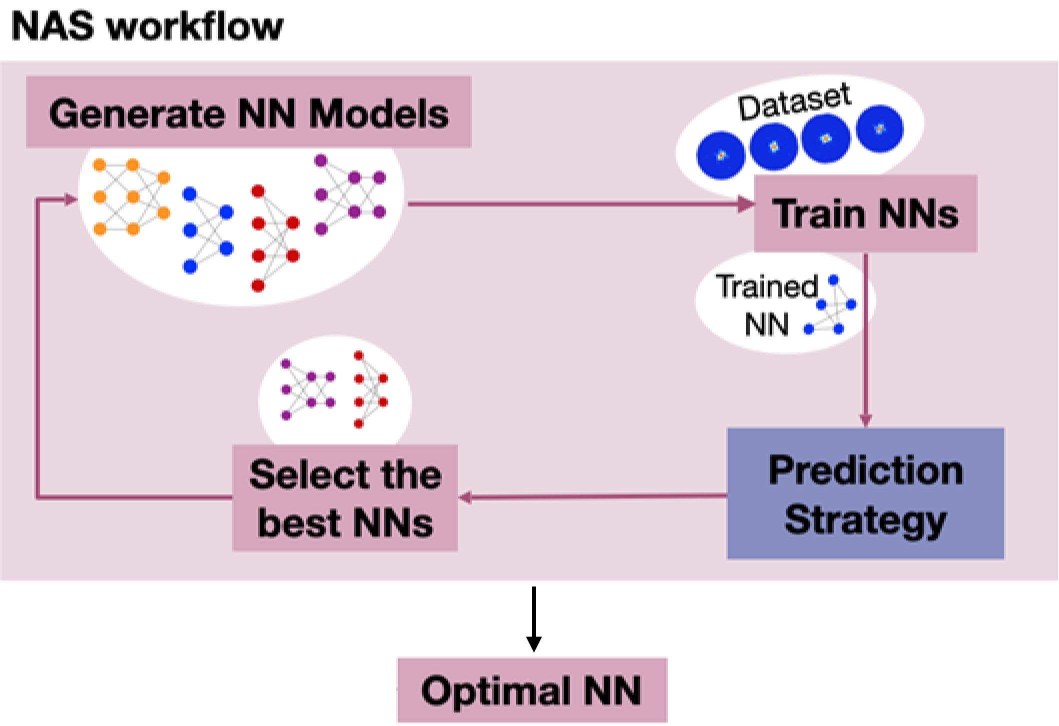
Using Neural Networks and Neural Architecture Search

Neural Networks (NN) can extract information from scientific data



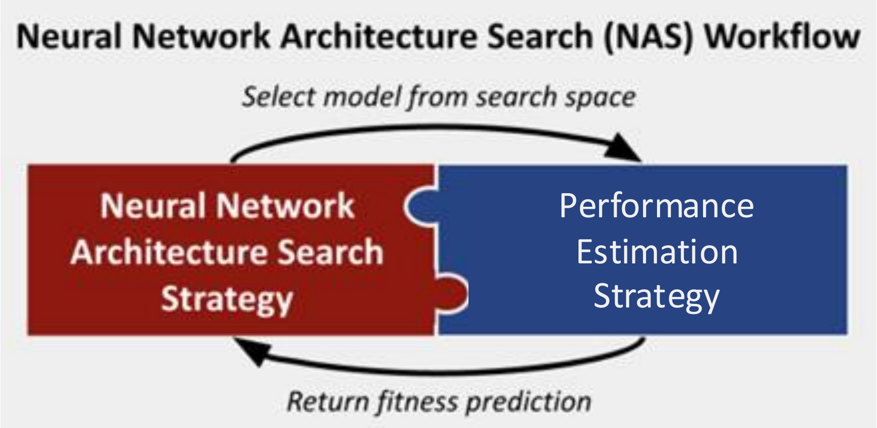
Custom NNs are needed for each dataset

Neural Architecture Search (NAS) can automatically find an optimal NN for a given dataset.



Monolithic Implementations

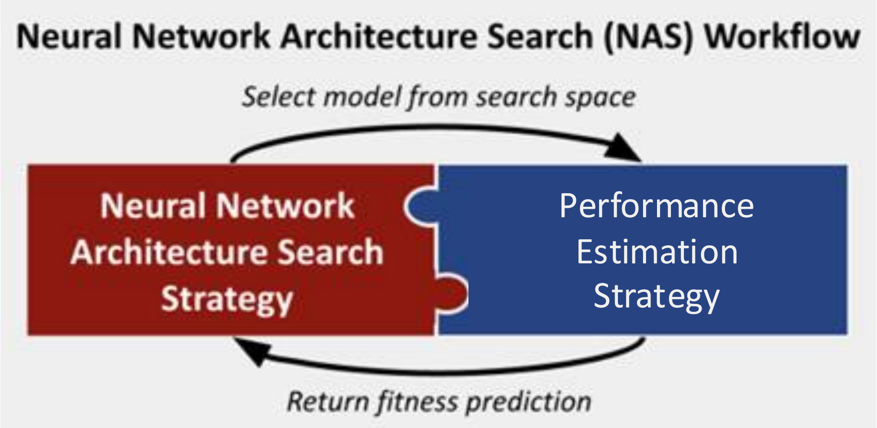
- NAS algorithms and their implementations are ***monolithic*** and couple the search and estimation strategies
  → ***Reduces the possibility*** *of modular optimizations*[1]



[1] C. *White* et al. *Neural Architecture Search: Insights from 1000 Papers*. arXiv:2301.08727. 2023.

Enormous Energy Consumption

- NAS algorithms and their implementations are ***monolithic*** and couple the search and estimation strategies
  → ***Reduces the possibility*** *of modular optimizations*[1]

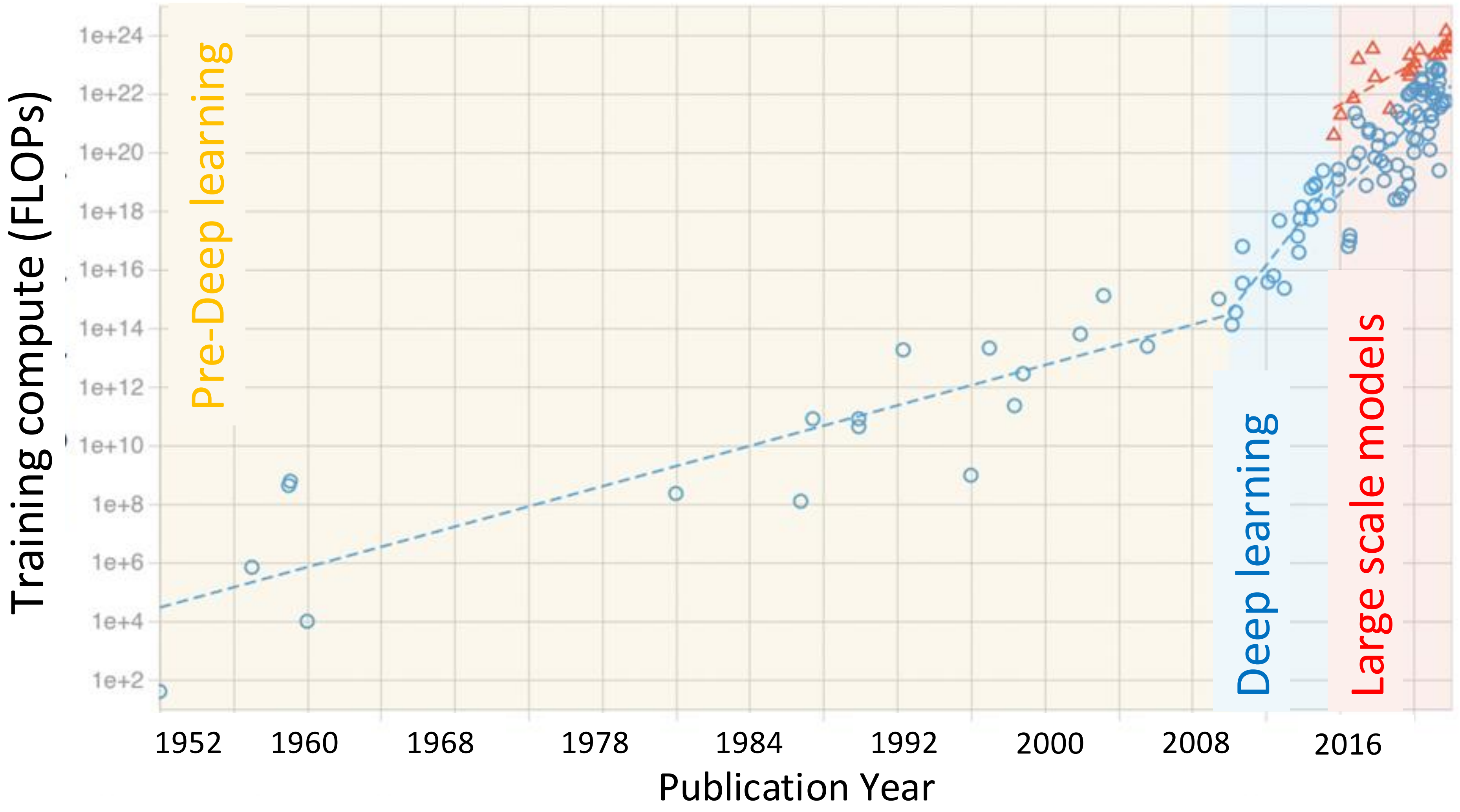


- NAS workflows consume ***enormous amounts of energy and time*** by training non-optimal networks for long training periods
  → ***Limits the accessibility of NAS*** *for researchers with compute limitations*[1]

[1] C. *White* et al. *Neural Architecture Search: Insights from 1000 Papers.* arXiv:2301.08727. 2023.

Ever-Growing Energy Consumption

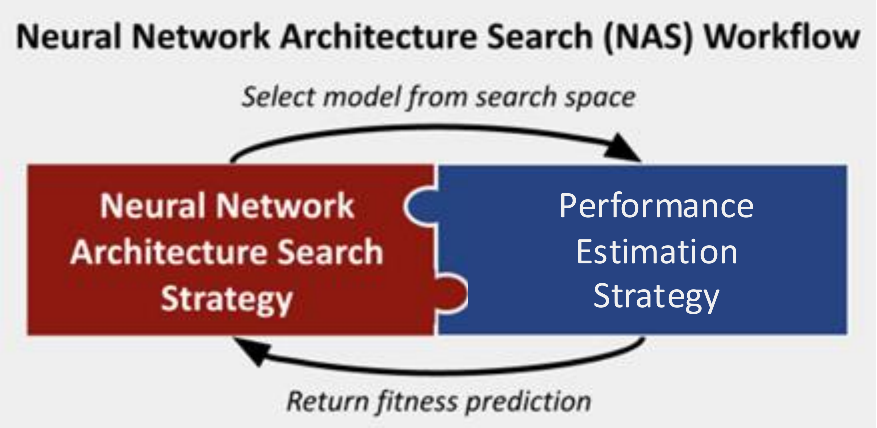
Training compute (FLOPs) of milestones Machine Learning system over time



Sevilla, Jaime & Heim, Lennart & Ho, Anson & Besiroglu, Tamay & Hobbhahn, Marius & Villalobos, Pablo. (2022). Compute Trends Across Three Eras of Machine Learning.

THE UNIVERSITY OF TENNESSEE KNOXVILLE

Obscured NN Evolution and Metadata



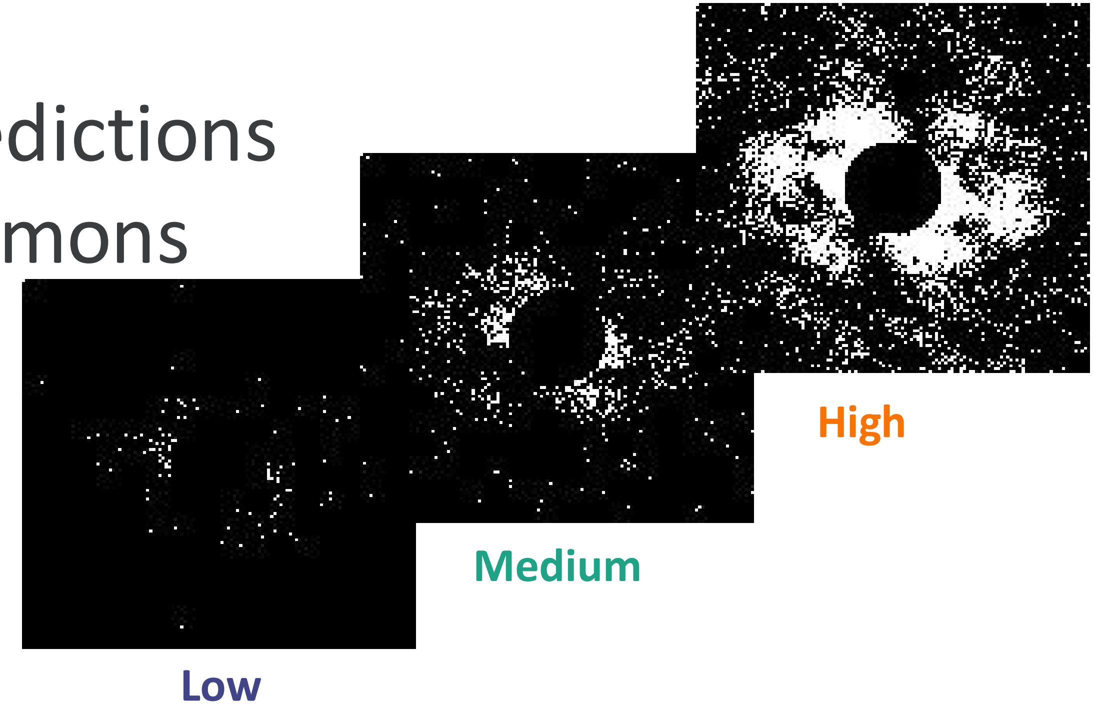
- NAS algorithms and their implementations are ***monolithic*** and couple the search and estimation strategies
  → ***Reduces the possibility*** *of modular optimizations*[1]
- NAS workflows consume ***enormous amounts of energy and time*** by training non-optimal networks for long training periods
  → ***Limits the accessibility of NAS*** *for researchers with compute limitations*[1]
- Search strategies ***obscure the evolution*** of NN architecture and their learning histories
  → ***Hinders the explainability*** *of resulting NNs*[1]

[1] C. *White* et al. *Neural Architecture Search: Insights from 1000 Papers*. arXiv:2301.08727. 2023.

Analytics for Neural Networks (A4NN)

We propose to augment NAS with our A4NN workflow:

- Transform NAS implementations from monolithic software tools into a flexible, modular workflow
- Generate adaptable NN fitness predictions
- Build an open-access NN data commons
- Assess A4NN with a dataset of simulated XFEL Images



Analytics for Neural Networks (A4NN)

We propose to augment NAS with our A4NN workflow:

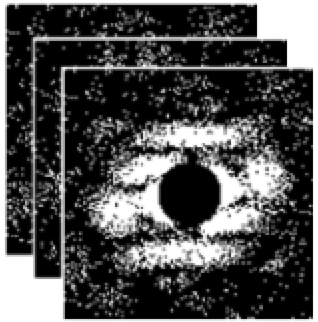
- Transform NAS implementations from monolithic software tools into a flexible, modular workflow
- Generate adaptable NN fitness predictions
- Build an open-access NN data commons
- Assess A4NN with a dataset of simulated XFEL Images

Low

The A4NN Workflow: Input Data

Conformations from protein diffraction patterns generated through simulations of XFEL experiments

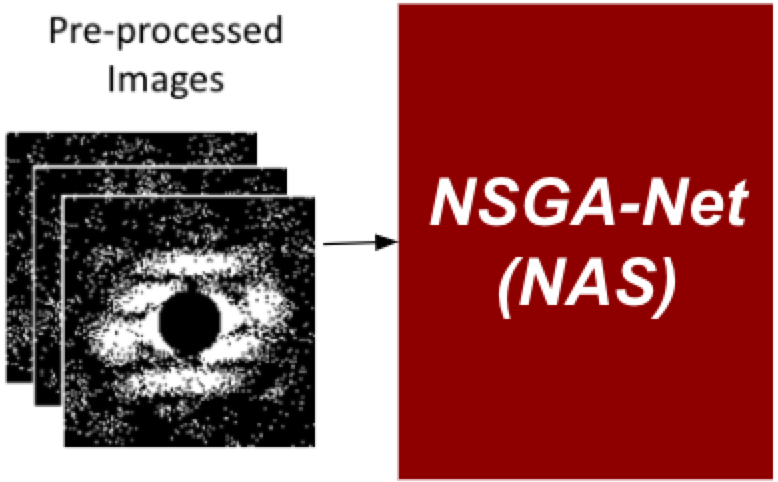
Pre-processed Images



Balanced conformation classes for each beam intensity → 80/20 train-test split of 63,508/15,876 images

The A4NN Workflow: NSGA-Net (NAS)

Select a NAS to train NNs from a specified search space

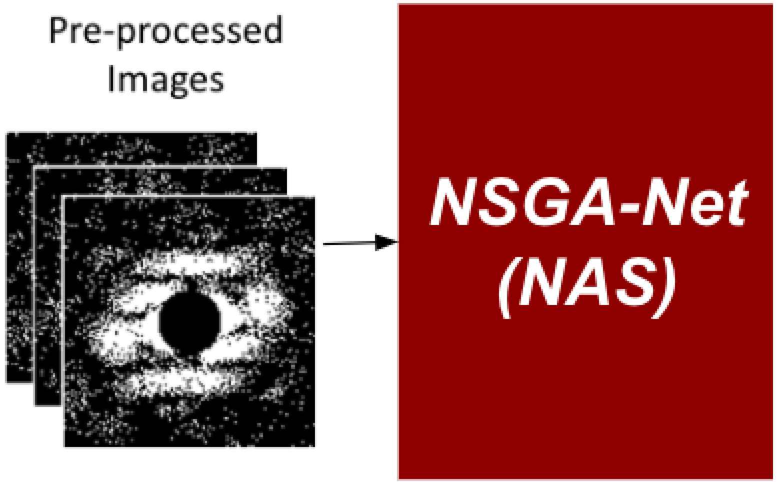


NAS	Type	Open source
EvoCNN [64]	evolutionary	✓
MENNDL [77]	evolutionary	×
NAS for image reconstruction [70]	evolutionary	×
psoCNN [20]	particle swarm	✓
Hierarchical representation [40]	evolutionary	×
NSGA-Net [42]	evolutionary	✓
Large-scale evolution [54]	evolutionary	×
Genetic CNN [73]	evolutionary	×
NASNet [79]	reinforcement learning	×
Auto-Keras [28]	bayesian	✓

*NSGA-Net optimizes for ***minimal FLOPS***

The A4NN Workflow: NSGA-Net (NAS)

Select a NAS to train NNs from a specified search space



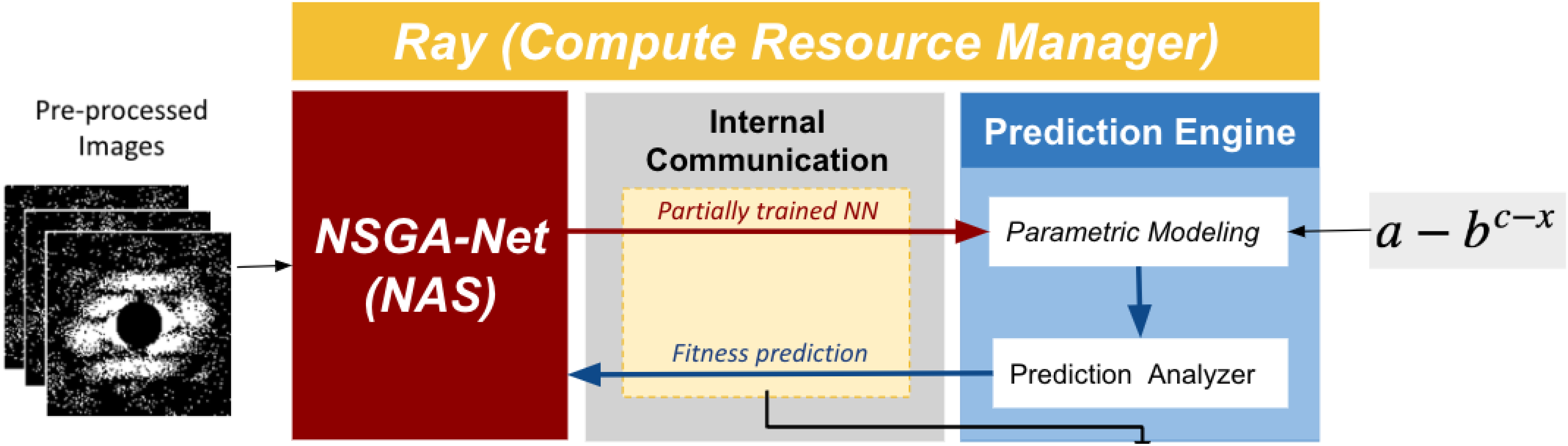
NAS	Type	Open source
EvoCNN [64]	evolutionary	✓
MENNDL [77]	evolutionary	×
NAS for image reconstruction [70]	evolutionary	×
psoCNN [20]	particle swarm	✓
Hierarchical representation [40]	evolutionary	×
NSGA-Net [42]	evolutionary	✓
Large-scale evolution [54]	evolutionary	×
Genetic CNN [73]	evolutionary	×
NASNet [79]	reinforcement learning	×
Auto-Keras [28]	bayesian	✓

*NSGA-Net optimizes for ***minimal FLOPS***

THE UNIVERSITY OF TENNESSEE KNOXVILLE

The A4NN Workflow: Prediction Engine

Pull partially trained NNs and curate a parametric function to **model the NNs fitness learning curves**

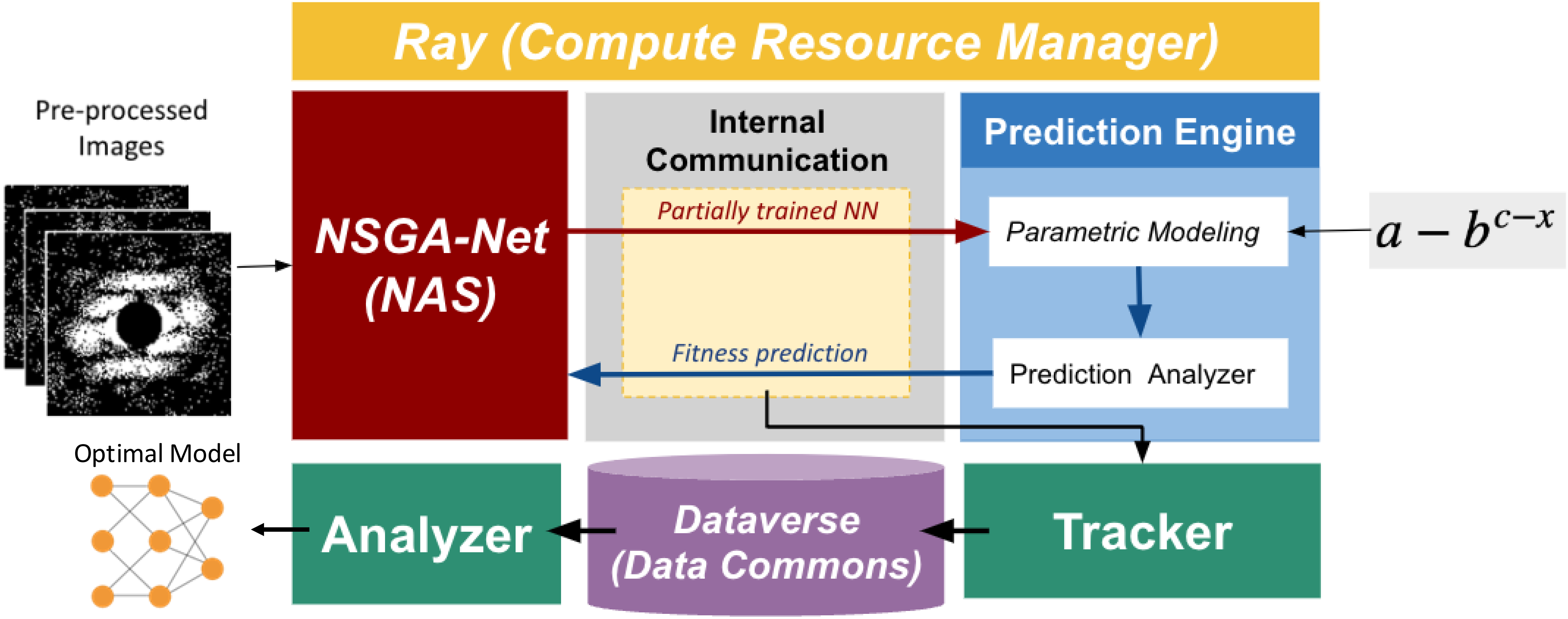


*NSGA-Net optimizes for ***minimal FLOPS***

THE UNIVERSITY OF TENNESSEE KNOXVILLE

The A4NN Workflow: NN Data Commons

Record the NN's behavior throughout training **for reproducible and explainable**

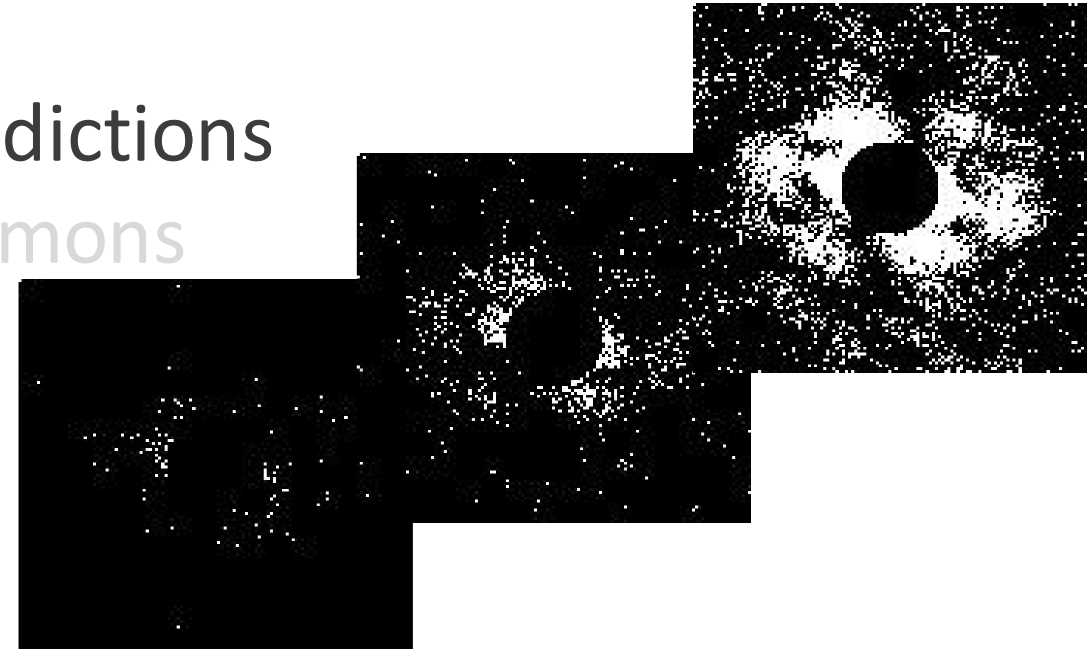


*NSGA-Net optimizes for ***minimal FLOPS***

Analytics for Neural Networks (A4NN)

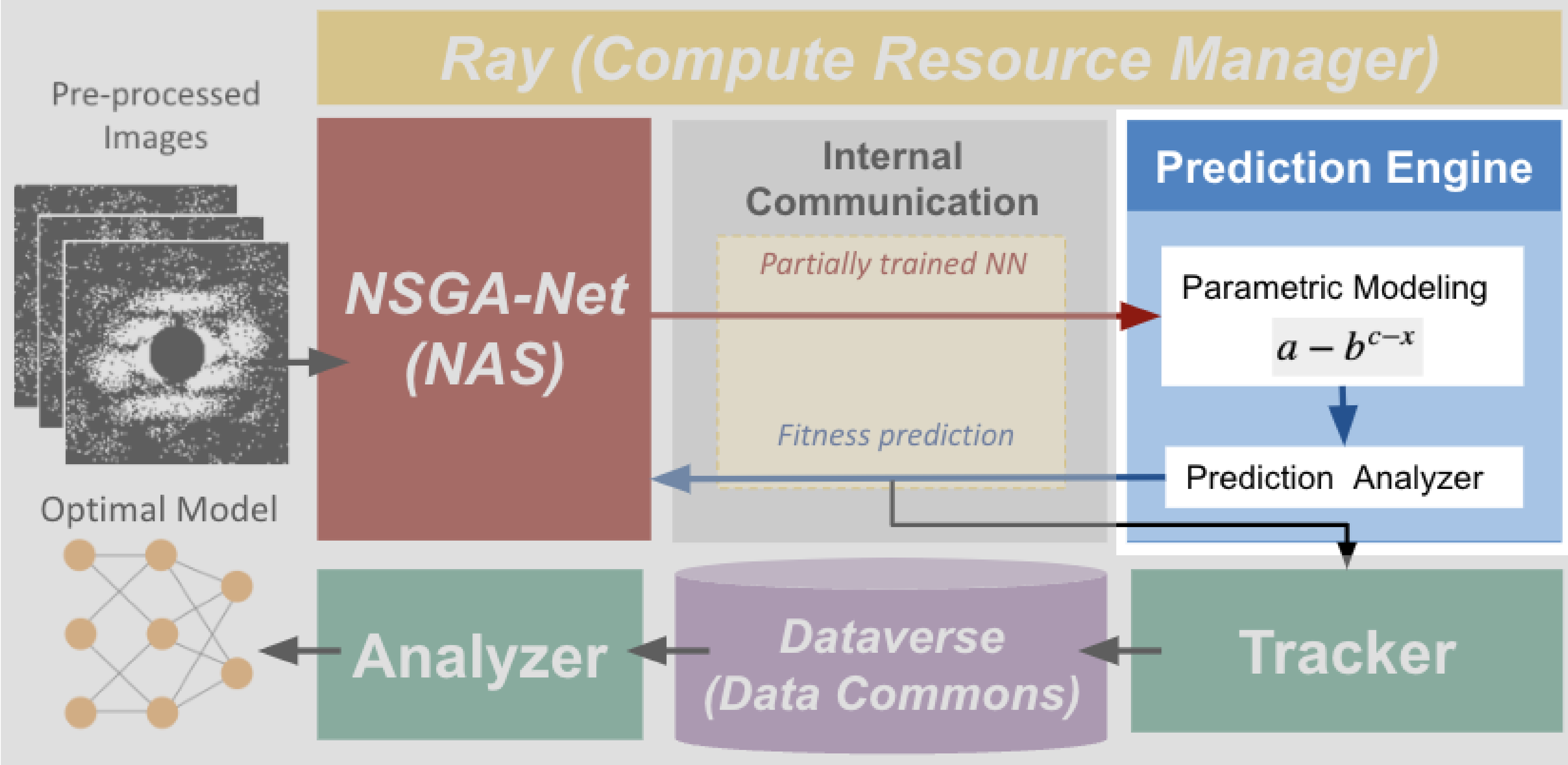
We propose to augment NAS with our A4NN workflow:

- Transform NAS implementations from monolithic software tools into a flexible, modular workflow
- Generate adaptable NN fitness predictions
- Build an open-access NN data commons
- Assess A4NN with a dataset of simulated XFEL Images



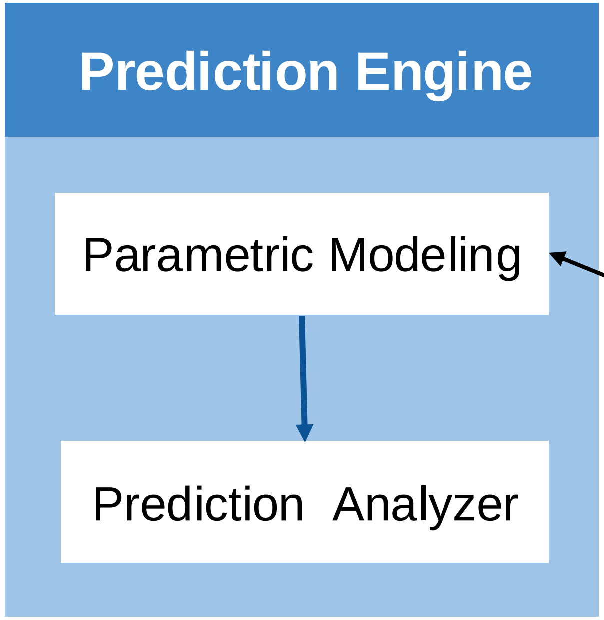
Predictive Engine for NNs (PENGUIN)

Predict NN performance to enable early stopping and expedite architecture optimization



Predictive Engine for NNs (PENGUIN)

- Generate fast, ***dynamic fitness predictions*** to inform the NAS
- Flexible learning curve modelling with ***parametric functions***
- ***Prediction validation*** over multiple epochs

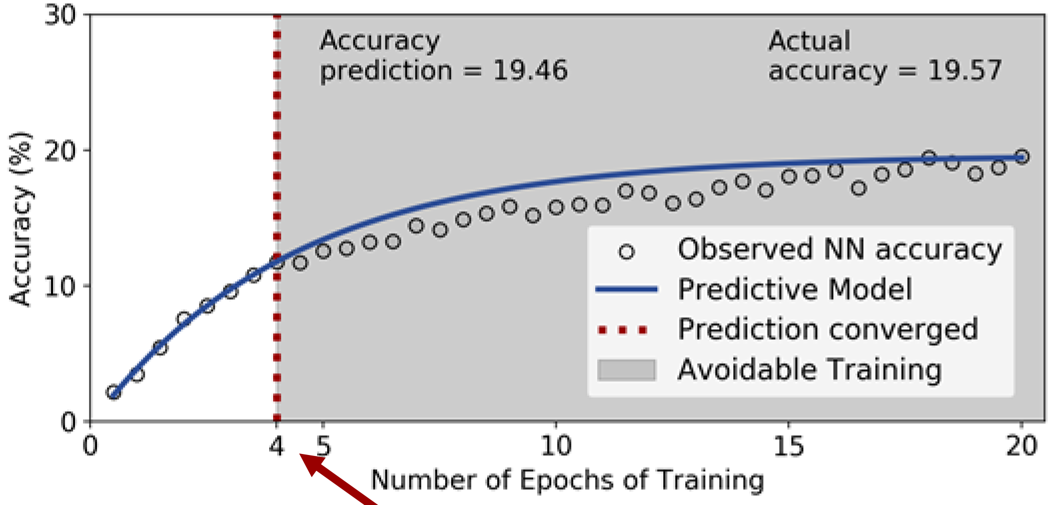


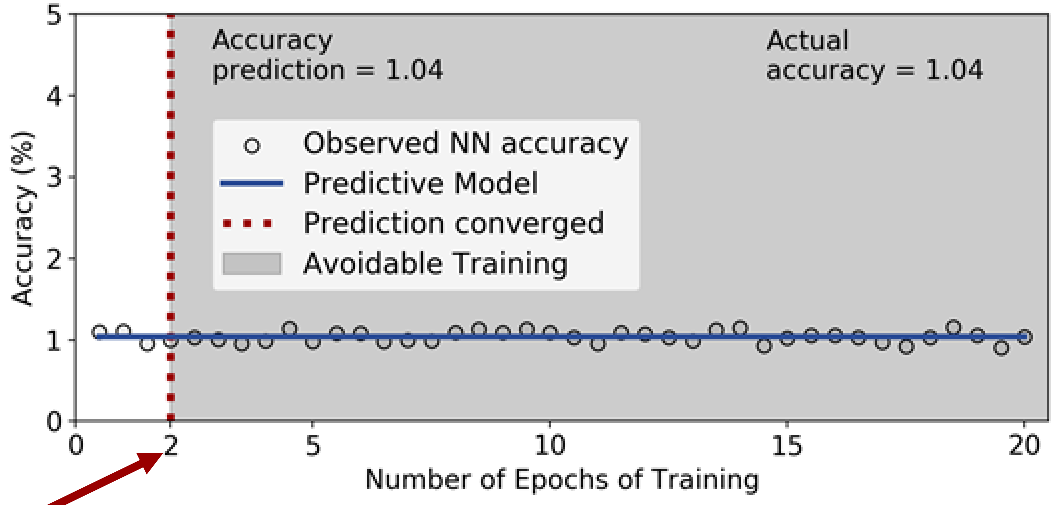
Users can plug in any parametric function to use for modeling– even a custom function!

Formula
an^{-b}
$an^{-b}+c$
$-a\log(n)+c$
$a\exp(-bn)+c$
$a\exp(-bn)$
$-an+b$
$\exp(a+b/n+c\log(n))$
$(ab+cn^d)/(b+n^d)$
$c-b\exp(-an^d)$
$c-\exp(-an^\alpha+b)$
$c-\exp((n-b)^\alpha)$
$c-(-an+b)^{-\alpha}$
$c-(a/\log(n))$
$c-(c-a)\exp(-bn)$

 Tom Viering and Marco Loog “The Shape of Learning Curves: a Review,” 2021 arxiv preprint.

Predictive Engine: An Example



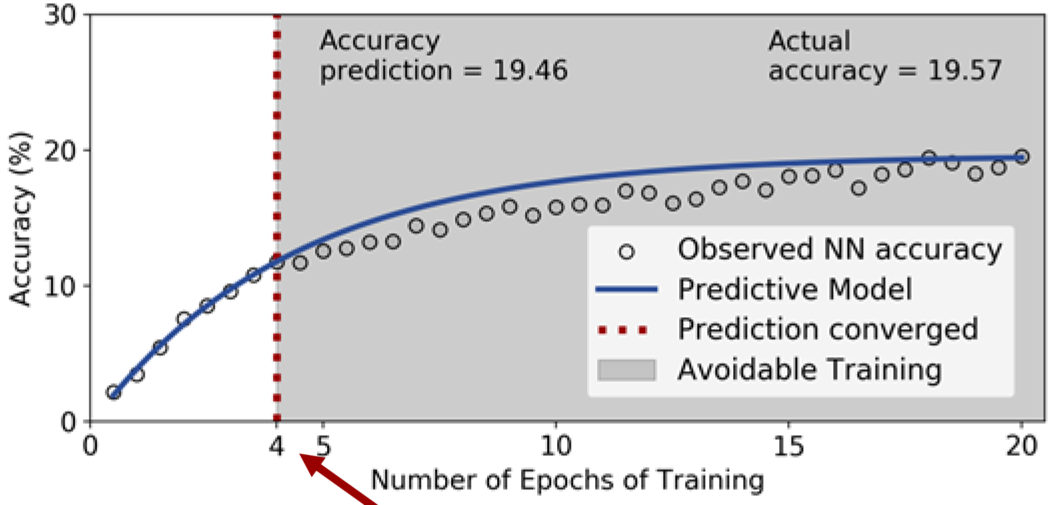


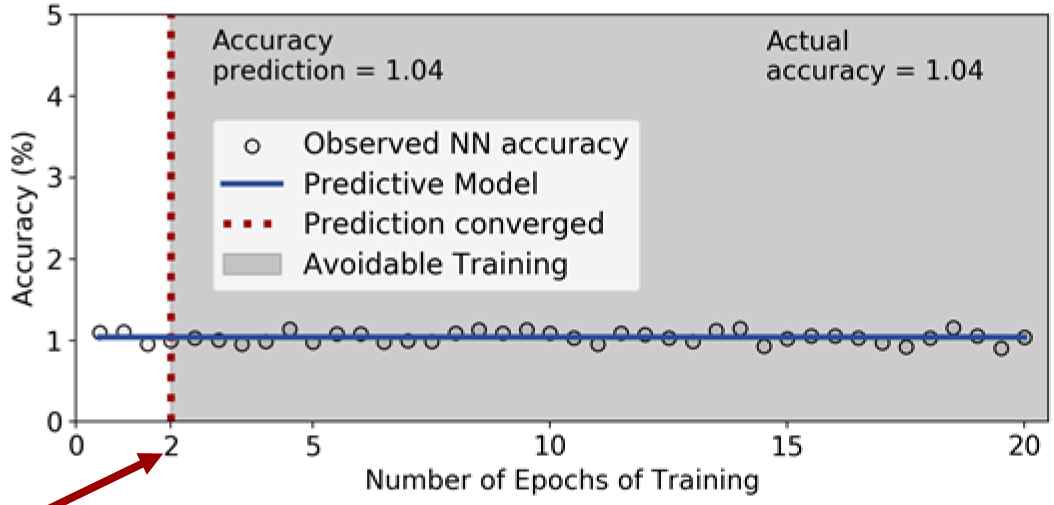
Prediction engine converges to a fitness prediction for epoch 20 (i.e., end of training)

Parametric function used:
$a - b^{c-x}$

Ariel Keller Rorabaugh, Silvina Caino-Lores, Travis Johnston, and Michela Taufer. Building High-throughput Neural Architecture Search Workflows via a Decoupled Fitness Prediction Engine *IEEE Trans. Parallel Distributed Syst. (TPDS), 33(11):2913–2926,* 2022.

Predictive Engine: An Example





Parametric function used:
$a - b^{c-x}$

Prediction engine converges to a fitness prediction for epoch 20 (i.e., end of training)

With this prediction, the workflow can:

- Terminate training of an NN
- Generate more NNs based on a top performing NN
- ***Save time and resources***

Ariel Keller Rorabaugh, Silvina Caino-Lores, Travis Johnston, and Michela Taufer. Building High-throughput Neural Architecture Search Workflows via a Decoupled Fitness Prediction Engine *IEEE Trans. Parallel Distributed Syst. (TPDS), 33(11):2913–2926,* 2022.

Analytics for Neural Networks (A4NN)

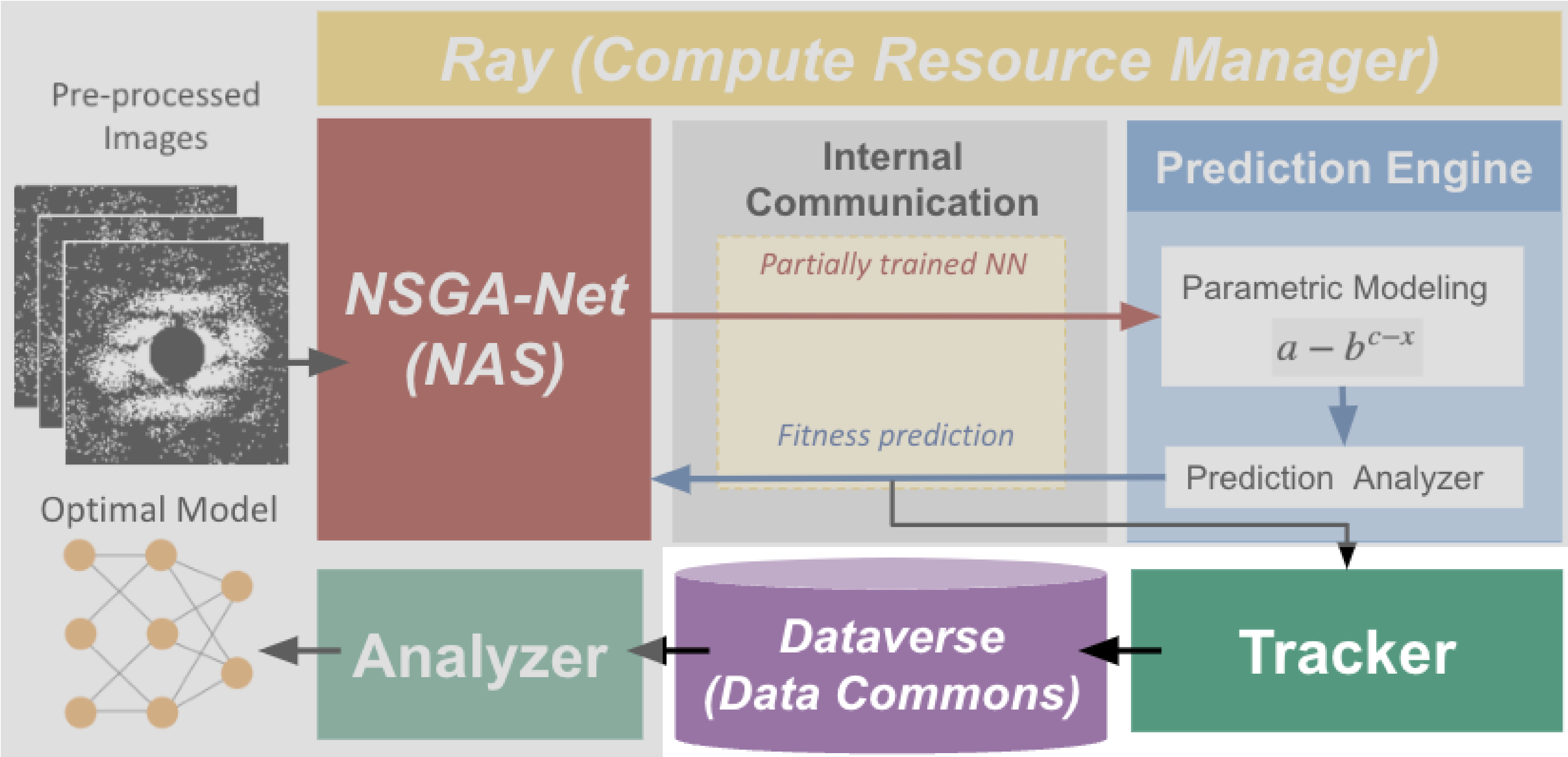
Analytics for Neural Networks

We propose to augment NAS with our A4NN workflow.

- Transform NAS implementations from monolithic software tools into a flexible, modular workflow
- Generate adaptable NN fitness predictions
- Build an open-access NN data commons
- Assess A4NN with a dataset of simulated XFEL Images

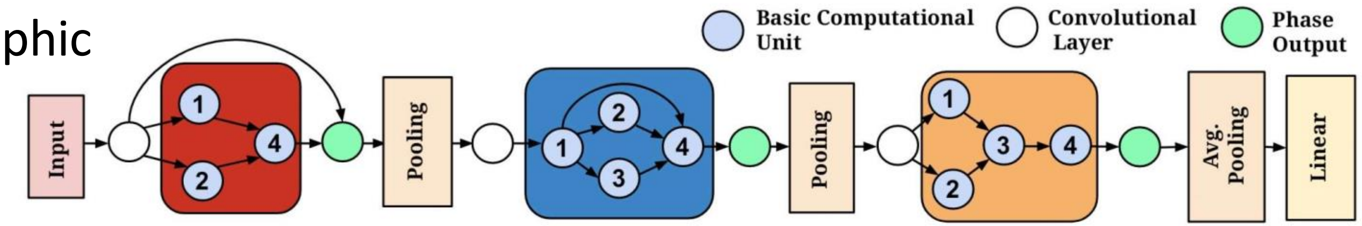
Tracker: Collecting Metadata

Analyze discarded NNs for similarities → understanding NN evolution and reproducibility

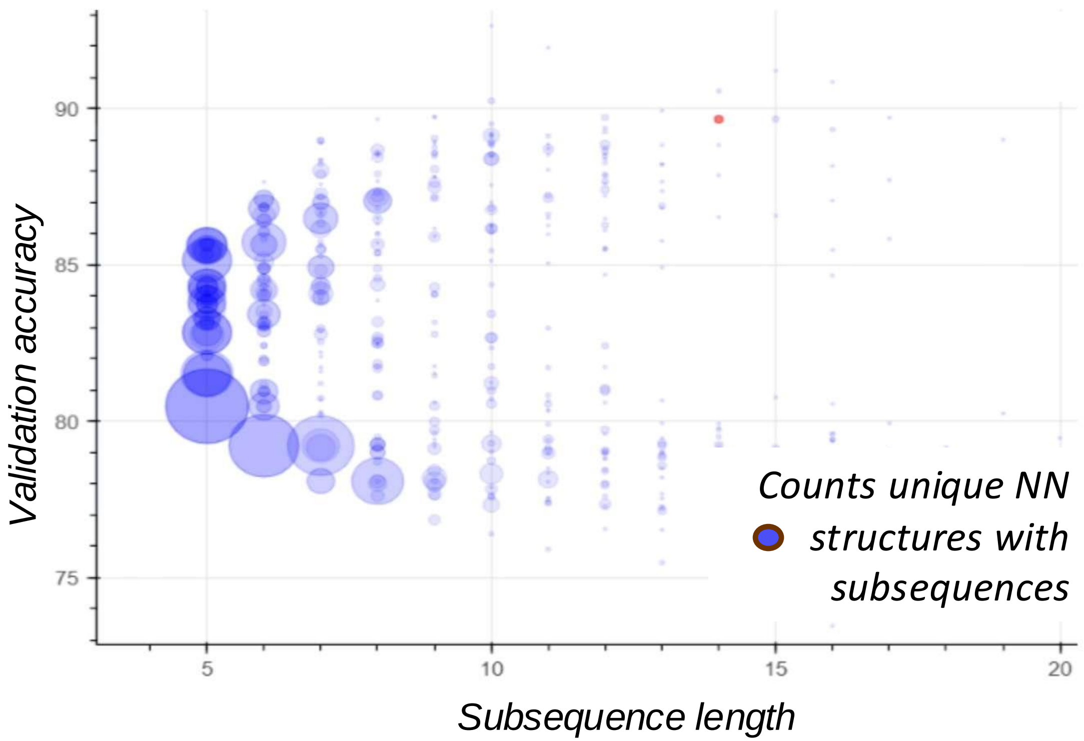


Visualizing Similarities

NN architecture generated with NSGA-Net as a graphic and binary array of connectivity phases → store matching subsequence and their length



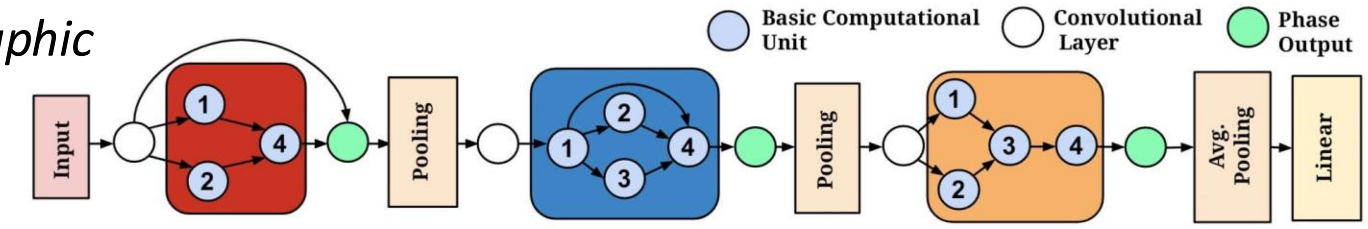
Scatter plot of matching network subsequences from NN commons



Seoyoung An, Georgia Channing, Catherine Schuman, and Michela Taufer. Visual Analytics Interactive Tool for Neural Network Archaeology. In Proceedings of the IEEE Cluster Conference (CLUSTER), 2023.

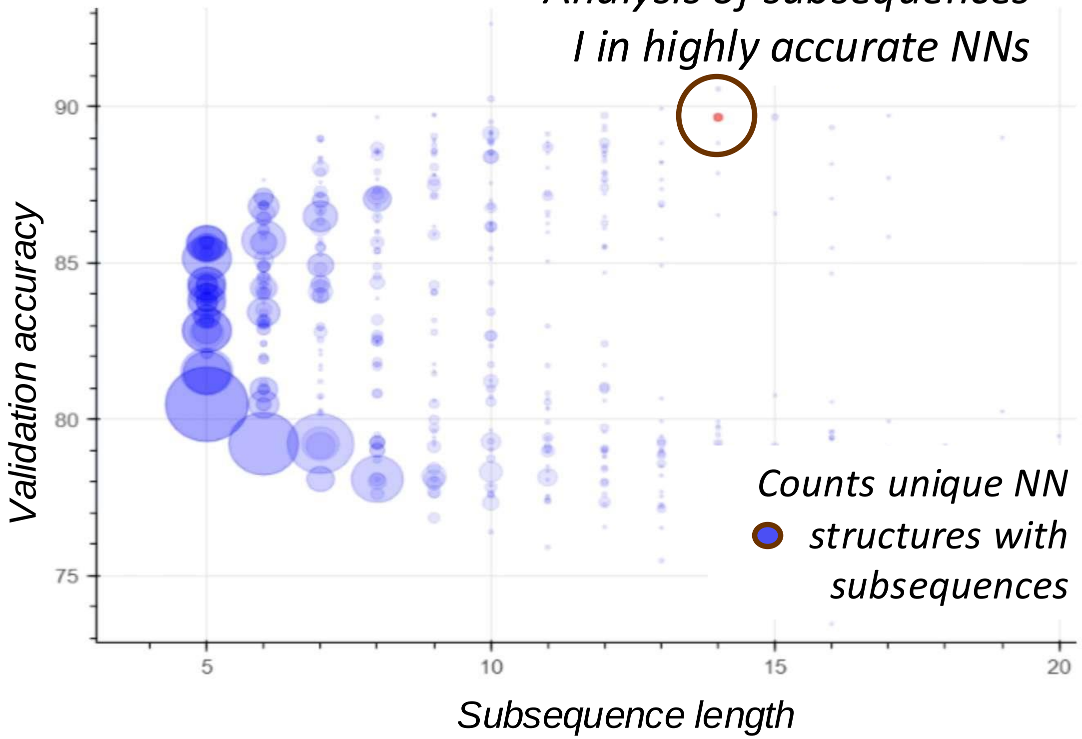
Visualizing Similarities

NN architecture generated with NSGA-Net as a graphic and binary array of connectivity phases → store matching subsequence and their length

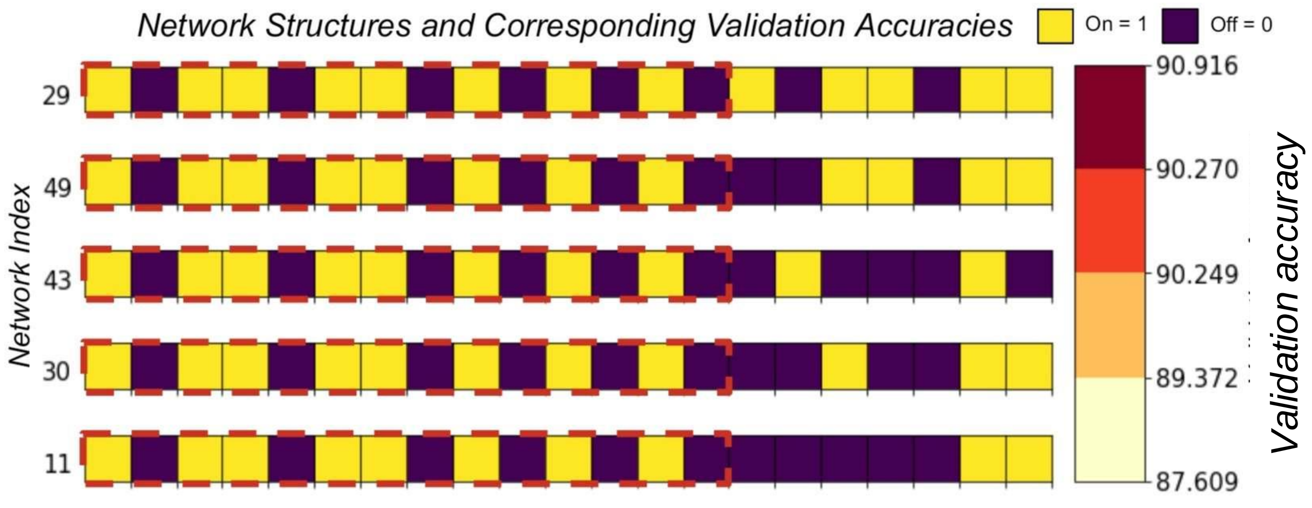


Scatter plot of matching network subsequences from NN commons

Analysis of subsequences I in highly accurate NNs



Counts unique NN structures with subsequences



The network structure of the selected network in our NN commons in ascending validation accuracy → dashed lines represent matching subsequence.

Seoyoung An, Georgia Channing, Catherine Schuman, and Michela Taufer. Visual Analytics Interactive Tool for Neural Network Archaeology. In Proceedings of the IEEE Cluster Conference (CLUSTER), 2023.

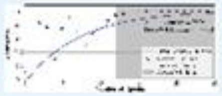
Data Commons: Accessing Metadata

HARVARD
DATAVERSE

Harvard Dataverse > Analytics for Neural Networks Dataverse >

Architecture Descriptions, Model Checkpoints, and Training Histories for A4NN Workflow on Protein Diffraction Data

Version 9.0



Channing, Georgia; Patel, Ria; Michela Taufer, 2023, "Architecture Descriptions, Model Checkpoints, and Training Histories for A4NN Workflow on Protein Diffraction Data", https://doi.org/10.7910/DVN/JS9FR6, Harvard Dataverse, V9, UNF:6:VC20hf9YdEgiPoEcOrhGjw== [fileUNF]

Cite Dataset ▾ Learn about Data Citation Standards.

Access Dataset ▾
Edit Dataset ▾
Link Dataset
Contact Owner | Share

Dataset Metrics

540 Downloads

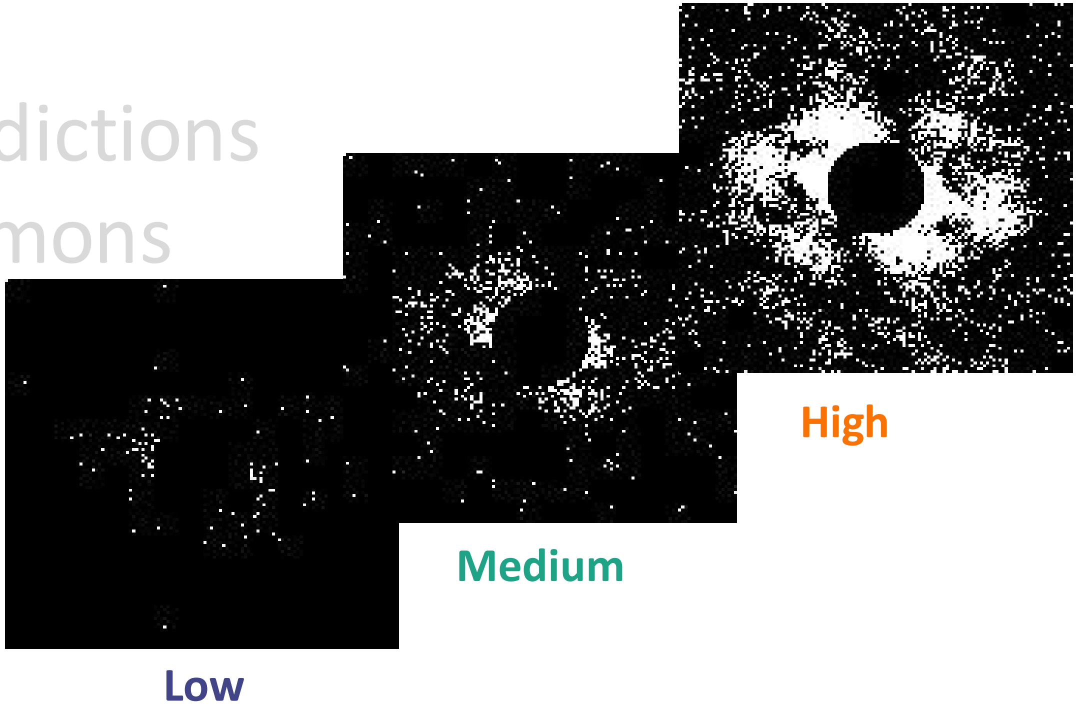
Description	This artifact contains scripts, input, and output datasets that reinforce the reproducibility of our results in our ICPP 2023 paper "Composable Workflow for Accelerating Neural Architecture Search Using In Situ Analytics for Protein Classification" (see README.txt). The input data comprises simulated protein diffraction patterns from X-ray Free Electron Laser (XFEL) experiments at low, medium, and high beam intensities. The datasets for each beam intensity contain 63,508 images for training and 15,876 images for testing (80/20 train-test split). The output dataset contains neural network (NN) models and metadata generated with the Analytics for Neural Networks (A4NN) workflow for several laser beam intensities on different GPU distributions. Each experiment contains 100 NN models that train for 25 epochs (max) each. There are approximately 72,900 model-related files in total.
Subject	Chemistry; Computer and Information Science; Physics
Keyword	Neural Architecture Search, neural networks, protein diffraction, predictive modeling, deep learning, early termination
License/Data Use Agreement	PUBLIC DOMAIN CC0 1.0

THE UNIVERSITY OF TENNESSEE KNOXVILLE

Analytics for Neural Networks (A4NN)

We propose to augment NAS with our A4NN workflow.

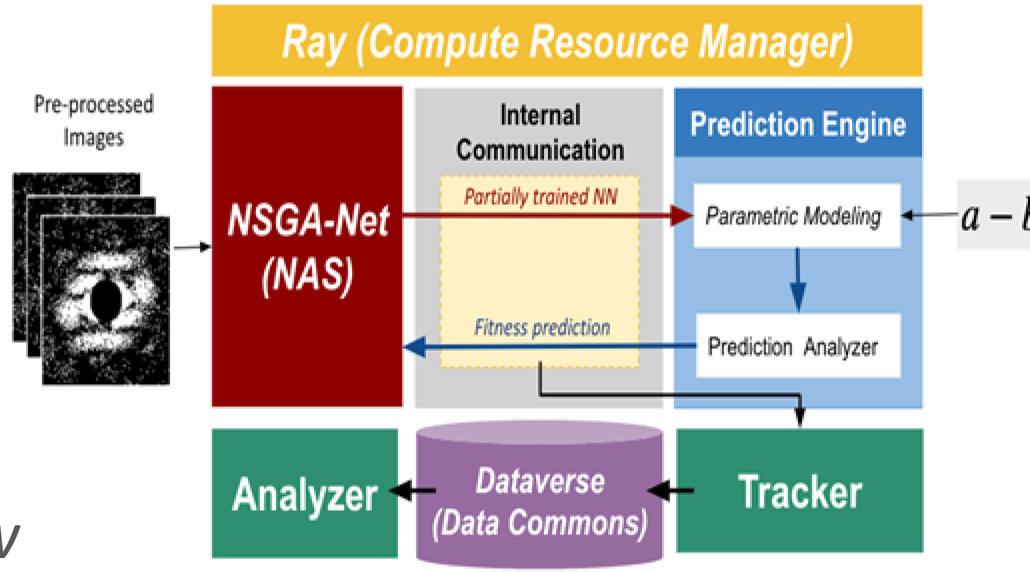
- Transform NAS implementations from monolithic software tools into a flexible, modular workflow
- Generate adaptable NN fitness predictions
- Build an open-access NN data commons
- Assess A4NN with a dataset of simulated XFEL Images



Performance Results: Metrics of Success

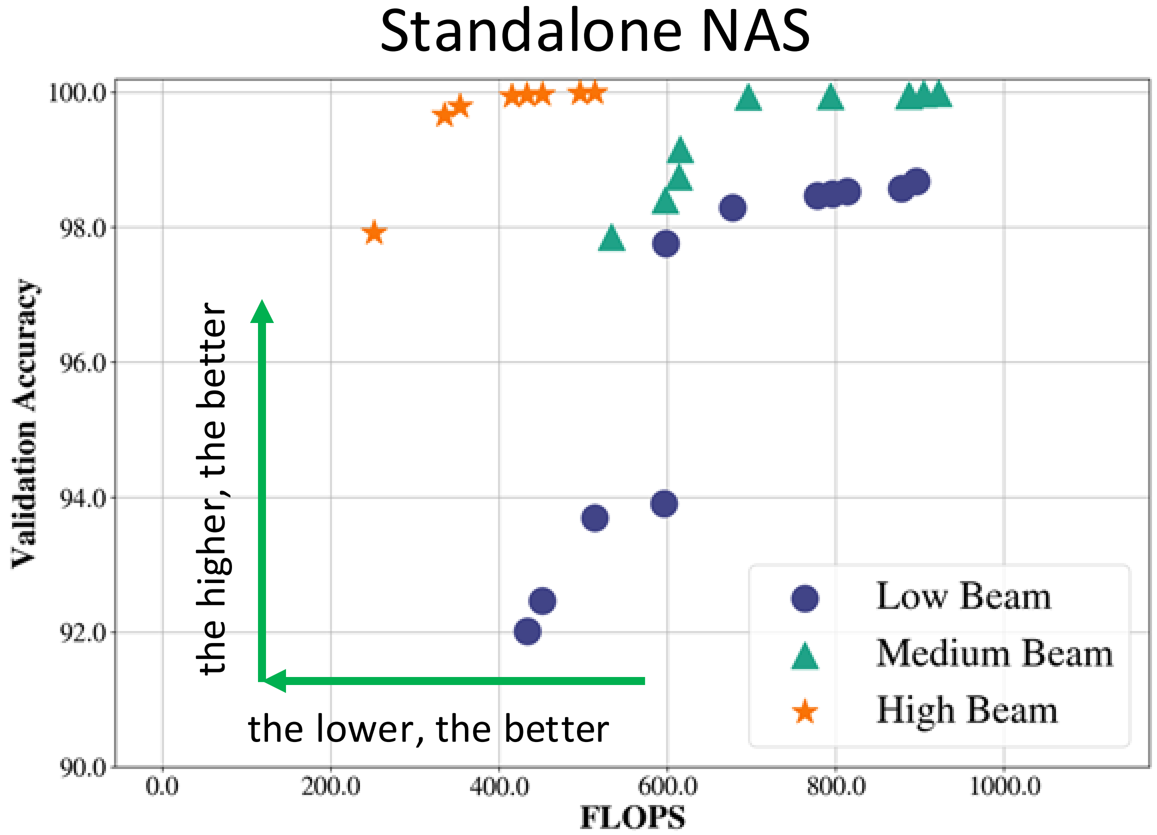
With A4NN, we strive for equal task performance with improved efficiency.
We evaluate our workflow by:

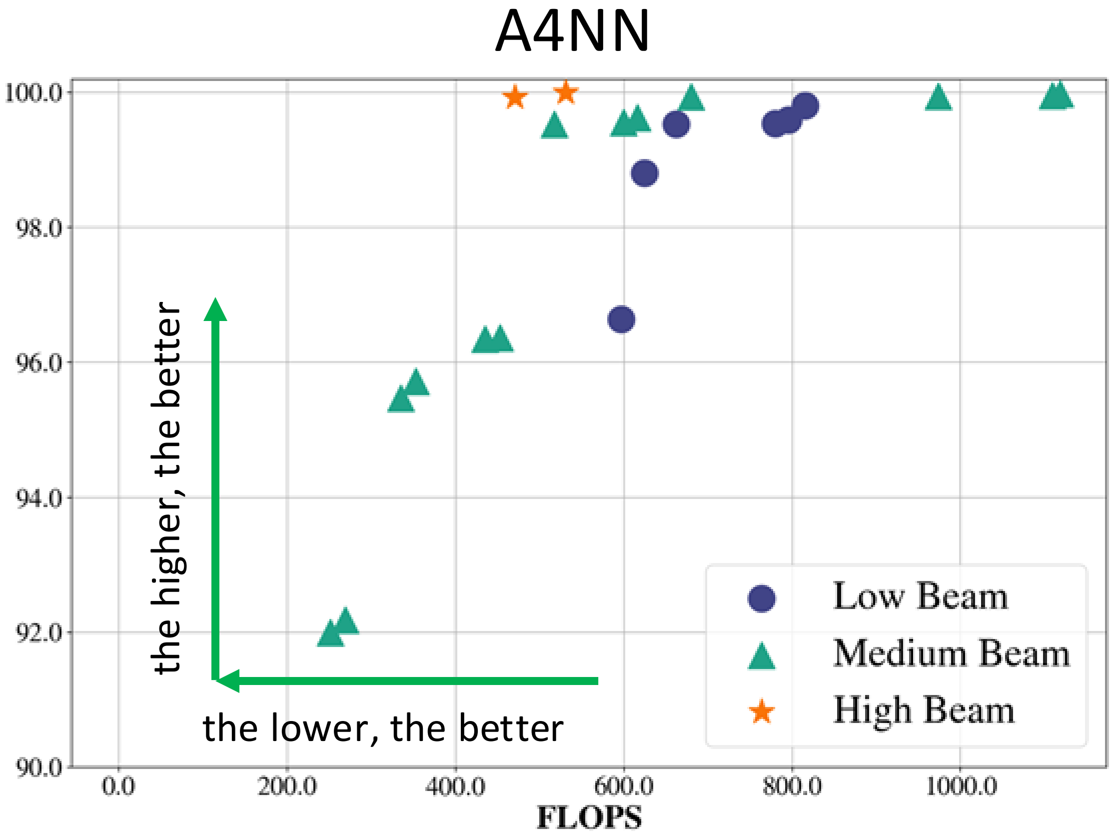
- ***Accuracy*** *(maximize)*
  - *Achieving results on par with SOTA*
- ***FLOPS*** *(minimize)*
  - *Proxy for energy consumption per model*
- ***Wall-time*** *(minimize)*
  - *Proxy for energy consumption for workflow*



Balanced conformation classes for each beam intensity
→ 80/20 train-test split of 63,508/15,876 images

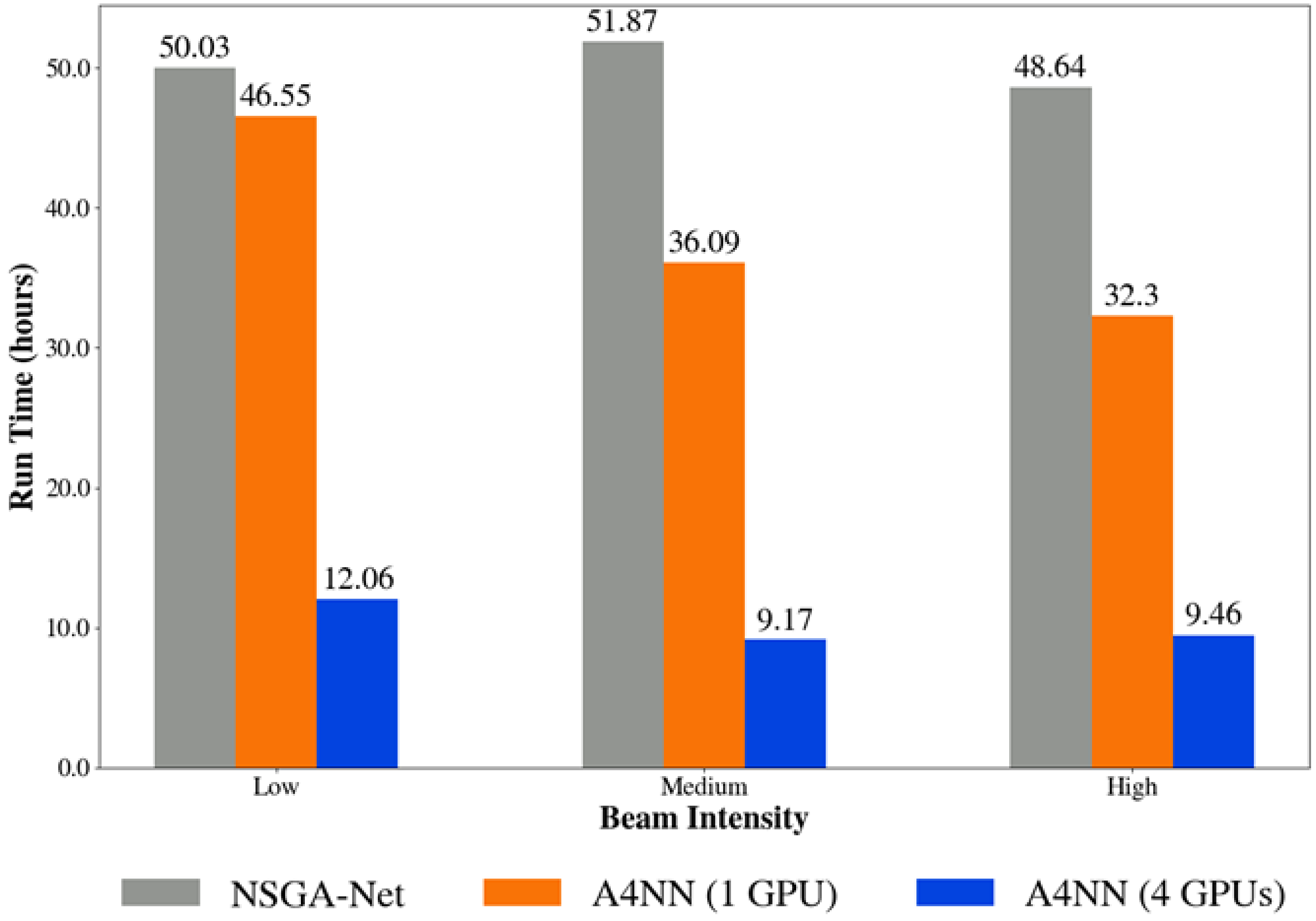
Performance Results: Accuracy and FLOPS





Comparable accuracy and FLOPS between standalone NAS and A4NN

Performance Results: Saved Wall-time



Hardware In-Use:

- DARWIN Cluster at U. Delaware
- 1 and 4 NVIDIA V100 GPUs

Configurations:

- NSGA-NET
- A4NN (1 GPU)
- A4NN (4 GPUs)

Data:

- low intensity
- medium intensity
- high intensity

5.5x faster wall-times when using A4NN compared to standalone NAS

A4NN Results Compared to State of the Art

We compare our results on *4 GPU* to a SOTA workflow for this dataset, called *XPSI*.

Beam	*Metric*	***XPSI***	***A4NN***
Low	Wall-Time	15.5 h	12.1 h
	Accuracy	92%	**97.8%**
Medium	Wall-Time	15.5 h	9.2 h
	Accuracy	99%	**99.9%**
High	Wall-Time	15.5 h	9.5h
	Accuracy	100%	100%

A4NN trains faster than XPSI and it's accuracies match or outperform XPSI

Paula Olaya, Silvina Caino-Lores, Vanessa Lama, Ria Patel, Ariel Rorabaugh, Osamu Miyashita, Florence Tama, and Michela Taufer. **Identifying Structural Properties of Proteins from X- ray Free Electron Laser Diffraction Patterns.** *In Proceedings of the 18th IEEE International Conference on e-Science (eScience), pages 1–10, Salt Lake City, Utah, USA, October 2022. IEEE Computer Society.*

Lessons Learned

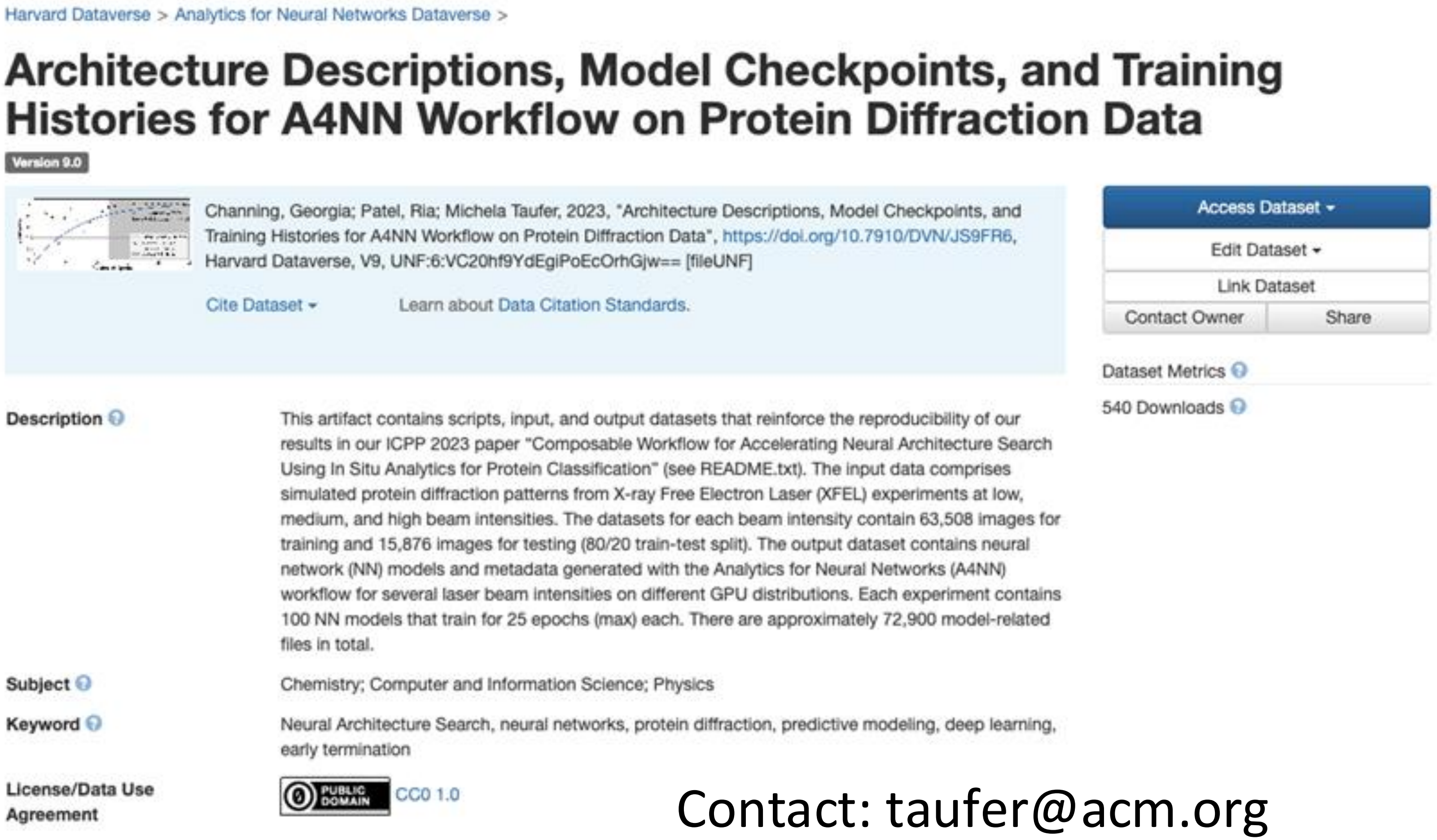
Strategies for successful optimization of NN training include:

- Decouple search and estimation strategies
- Minimize energy and training time
- Explain NN per performance by examining training NNs' history

With the A4NN workflow, we deliver:

- A composable, reusable deep-learning workflow for scientific datasets
- An efficient prediction methodology for any NAS
- 54 GB of metadata and model checkpoints for future study

Data Commons and Reproducibility



Harvard Dataverse > Analytics for Neural Networks Dataverse >

Architecture Descriptions, Model Checkpoints, and Training Histories for A4NN Workflow on Protein Diffraction Data

Version 9.0

Channing, Georgia; Patel, Ria; Michela Taufer, 2023, "Architecture Descriptions, Model Checkpoints, and Training Histories for A4NN Workflow on Protein Diffraction Data", https://doi.org/10.7910/DVN/JS9FR6, Harvard Dataverse, V9, UNF:6:VC20hf9YdEgiPoEcOrhGjw== [fileUNF]

Cite Dataset ▾ Learn about Data Citation Standards.

Access Dataset ▾

Edit Dataset ▾

Link Dataset

Contact Owner | Share

Dataset Metrics

540 Downloads

Description: This artifact contains scripts, input, and output datasets that reinforce the reproducibility of our results in our ICPP 2023 paper "Composable Workflow for Accelerating Neural Architecture Search Using In Situ Analytics for Protein Classification" (see README.txt). The input data comprises simulated protein diffraction patterns from X-ray Free Electron Laser (XFEL) experiments at low, medium, and high beam intensities. The datasets for each beam intensity contain 63,508 images for training and 15,876 images for testing (80/20 train-test split). The output dataset contains neural network (NN) models and metadata generated with the Analytics for Neural Networks (A4NN) workflow for several laser beam intensities on different GPU distributions. Each experiment contains 100 NN models that train for 25 epochs (max) each. There are approximately 72,900 model-related files in total.

Subject: Chemistry; Computer and Information Science; Physics

Keyword: Neural Architecture Search, neural networks, protein diffraction, predictive modeling, deep learning, early termination

License/Data Use Agreement: PUBLIC DOMAIN CC0 1.0

Contact: taufer@acm.org

THE UNIVERSITY OF TENNESSEE KNOXVILLE