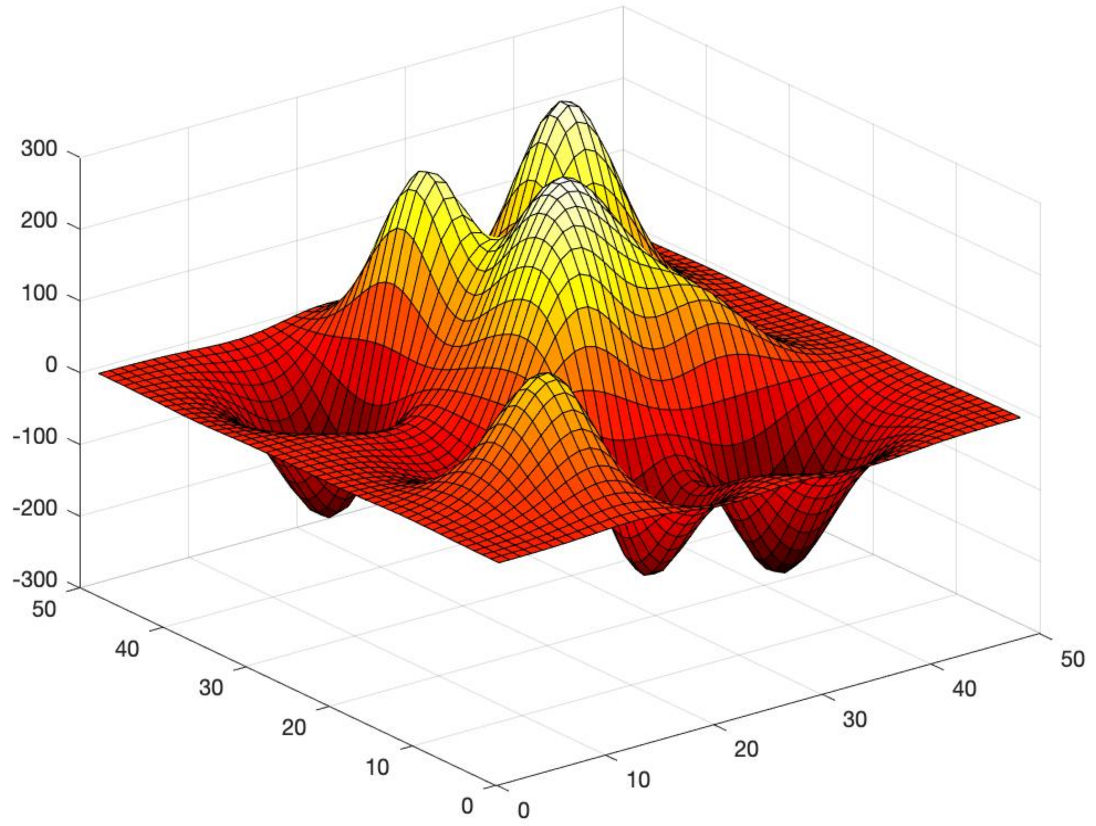
On the Effectiveness of Bayesian AutoML methods for Physics Emulators

Peetak P. Mitra, Niccolò Dal Santo, Majid Haghshenas, Shounak Mitra, Conor Daly, David P. Schmidt

NEURAL INFORMATION PROCESSING SYSTEMS
ML4Eng Paper

Motivation

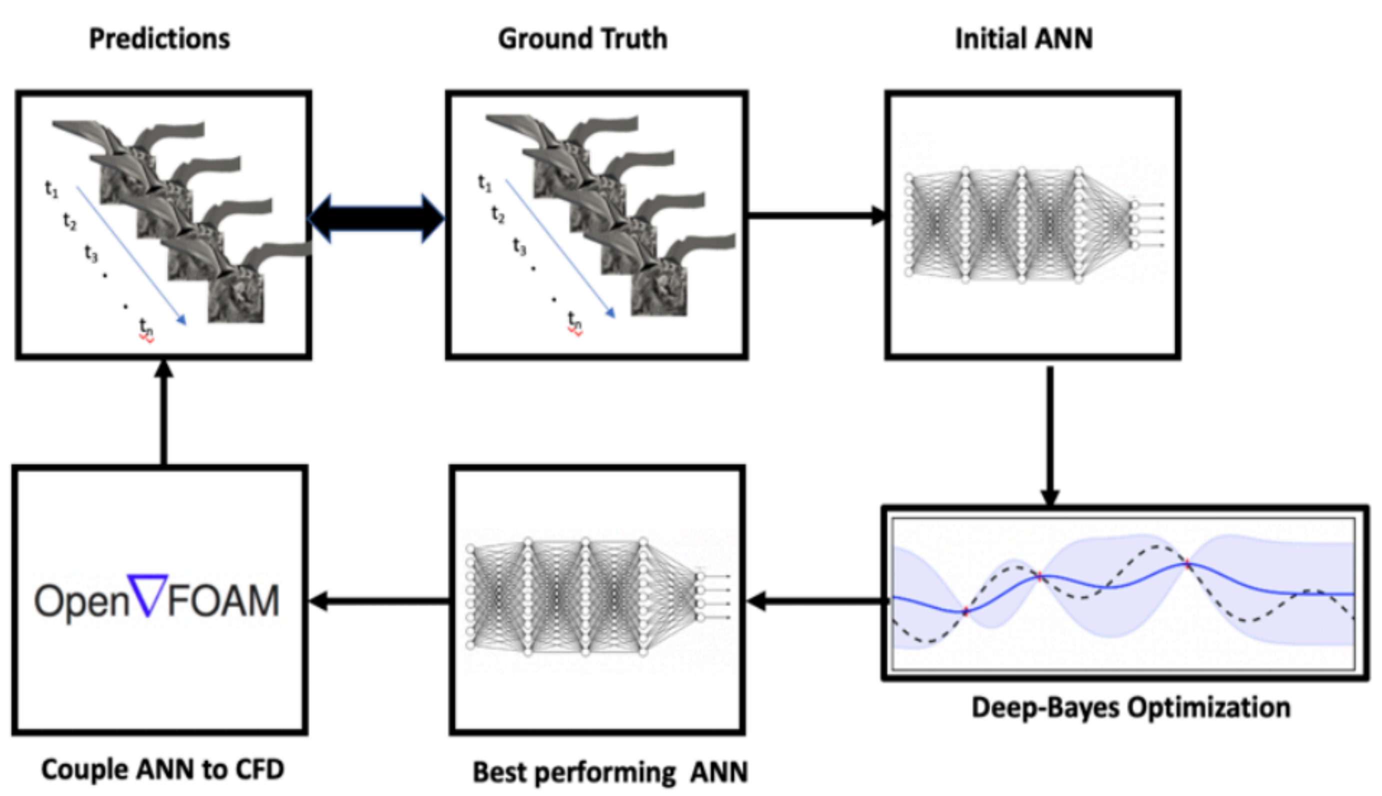
- Scientific data is often high-dimensional, complex, structured, sparse → Complicated loss manifold!
- Best practices in setting up network/hyperparameters translate poorly to scientific data
- Automatic ML methods are promising alternative to robustly train DNNs!



A translating, scaling Gaussian distribution-based loss manifold shows many peaks/troughs

Workflow

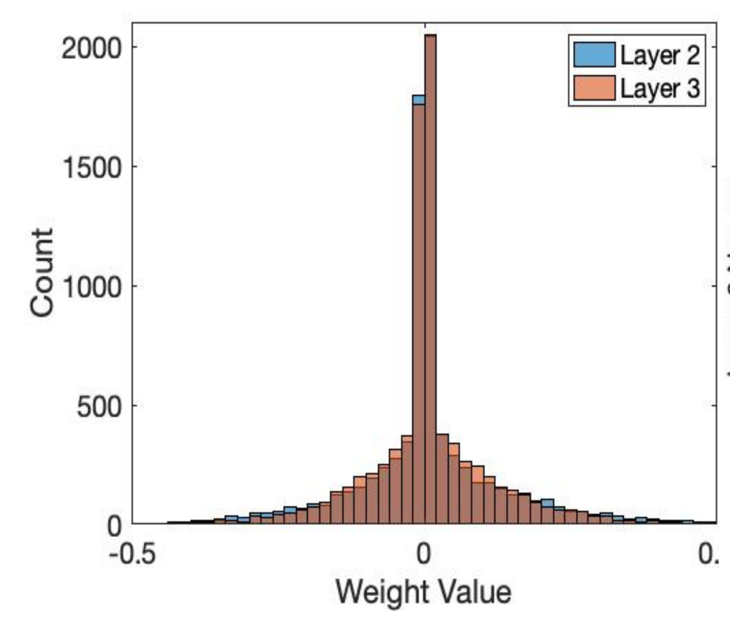
- Bayesian Optimization based AutoML explores parameter space to identify best performing settings, starting with an initial guess
- BayesOpt modeled as Gaussian Process (GP) with Expected Improvement (EI) acquisition function

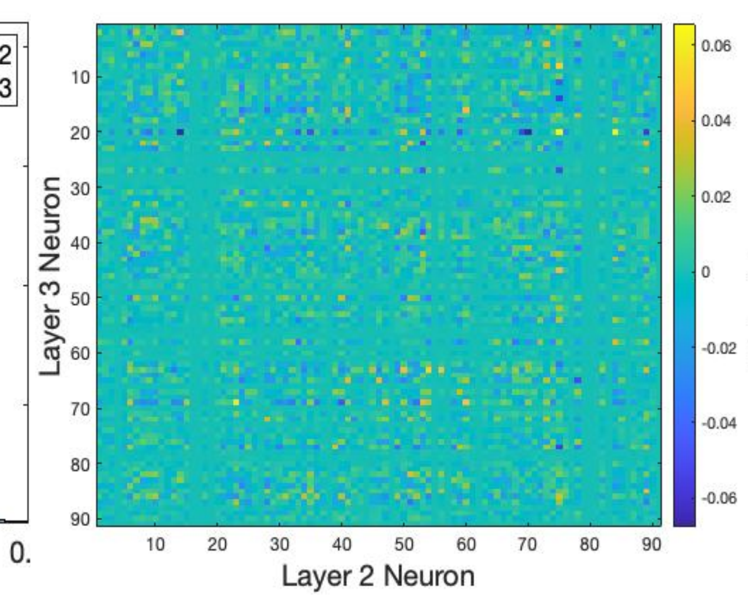


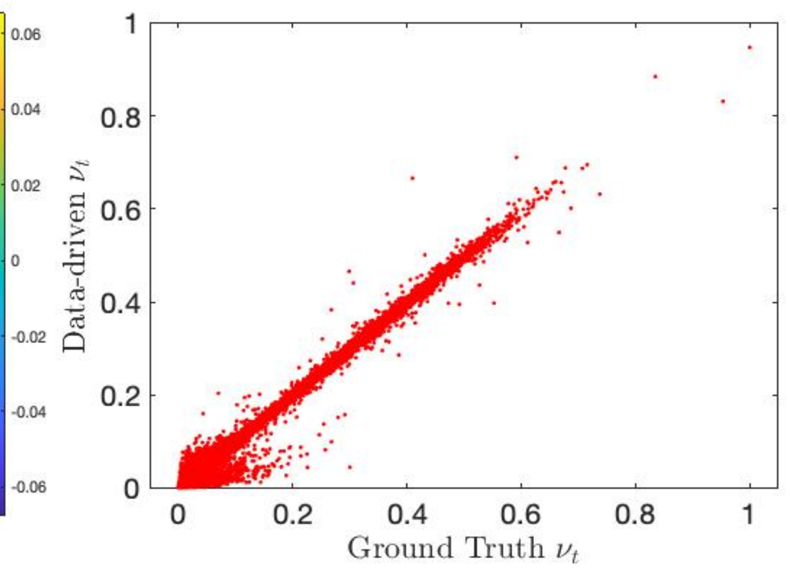
Workflow for AutoML for a physics emulation task. While the final goal of this work is to incorporate machine-learnt model to non-linear PDE solver OpenFOAM, this study will limit to understanding the learning process

Results

Optimizer	Initialization	Batch Size	LR	W	D	$\bar{\varepsilon}$	N
ADAM	Glorot	1426	9.56e-04	91	10	1e-05	84904
ADAM	He	12719	3.77e-04	50	7	2e-04	18207
SGDM	Glorot	6814	0.0098	91	9	2e-03	75812
SGDM	He	1161	0.0098	89	6	1e-03	48778
RMSProp	Glorot	291	1e-04	78	4	8e-04	25432
RMSProp	He	11630	1.86e-05	55	9	1e-03	28004





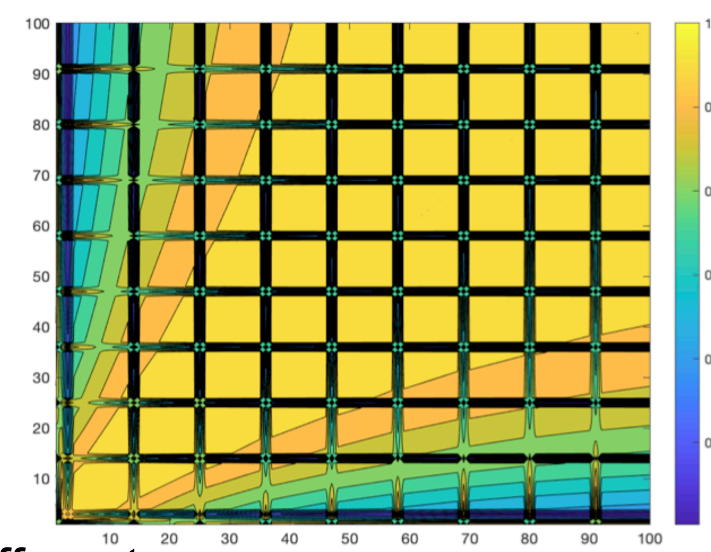


Weight Space Similarity

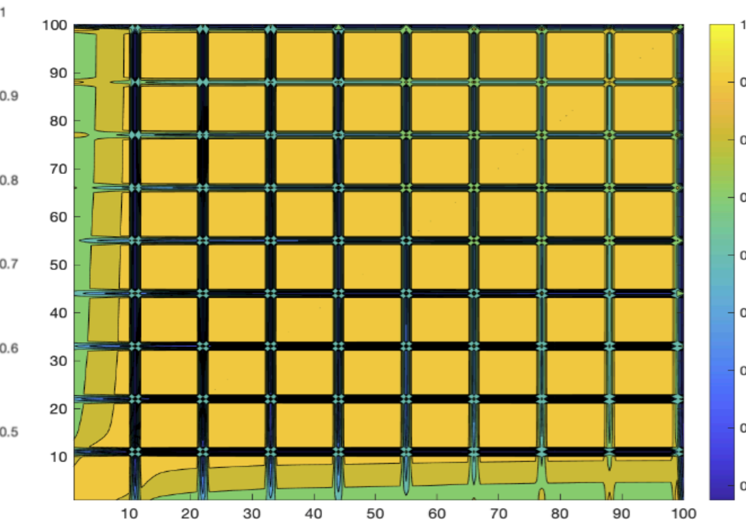
Network checkpointed after every epoch

$$\cos(\theta_1, \theta_2) = \frac{\theta_2 \theta_1^T}{\|\theta_2\| \|\theta_1\|}$$

Self-similar training evolution
Different optimizers have vastly different trajectories



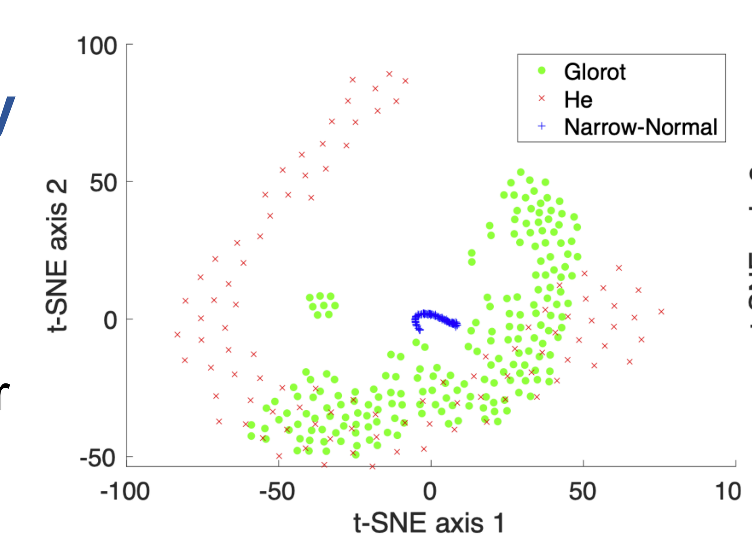
ADAM-Glorot

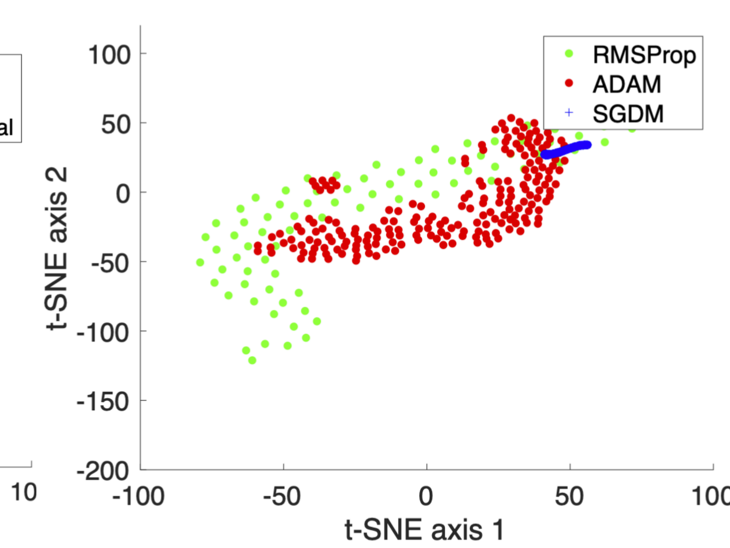


SGDM-Glorot

Function Space Similarity

- t-SNE based dimensionality reduction
- SGDM is stuck in local optima
- ADAM/RMSProp explore larger swaths
- Glorot/He have overlapping exploration



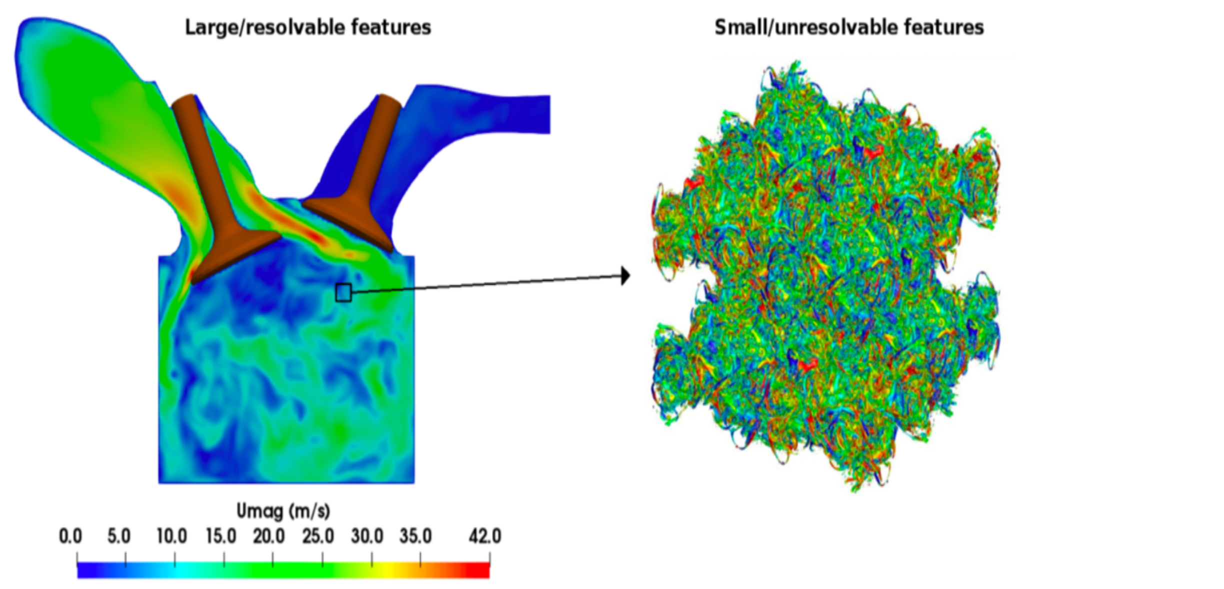


TL;DR

- AutoML methods are successfully in robustly identifying best performing settings for a complicated physics problem
- Network training evolution is heavily dependent on choice of optimizers, adaptive LR optimizers outperform
- Layer-by-Layer learnt weight comparisons reveals complicated nature of network learning

Physics

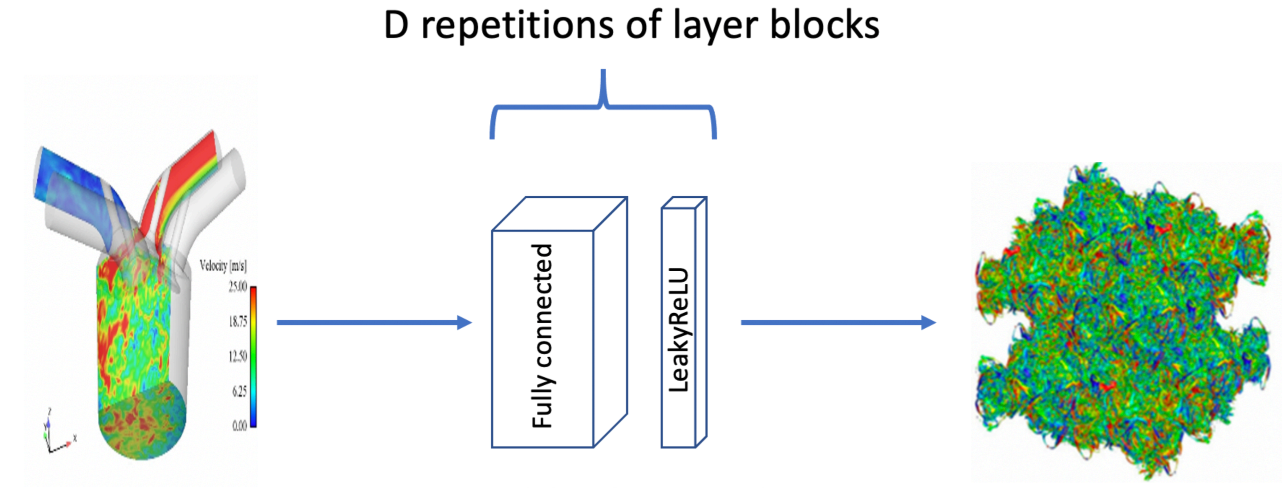
- Accurate turbulence closure is critically important for LES models
- Most methods involve some form of data-fitting, so why not a purely data-driven method!
- Eddy viscosity modeled as a function of large scale filtered variables $\nu_t = f(Re_c, S, \Omega, \nabla K, Y)$



Multiscale nature of a typical Internal Combustion Engine (ICE) simulation. Adopted from Dias Riberio et al.

Network Architecture Search

- NAS, Including optimize for weight initialization and solver



Hyper-parameter	Min. Range	Max. Range	Interpolation
Initial LR	1e-6	1e-2	Logarithmic
LR Drop Factor	10	1000	Integer
Batch Size	100	16000	Logarithmic
Network Depth	2	10	Integer
Network Width	10	100	Integer

References

Dias Ribeiro, Mateus, Alex Mendonça Bimbato, Maurício Araújo Zanardi, José Antônio Perrella Balestieri, and David P. Schmidt. "Large-eddy simulation of the flow in a direct injection spark ignition engine using an open-source framework." *International Journal of Engine Research* (2020): 1468087420903622.Marge Simpson (2010).

Acknowledgements