Group Meeting Report

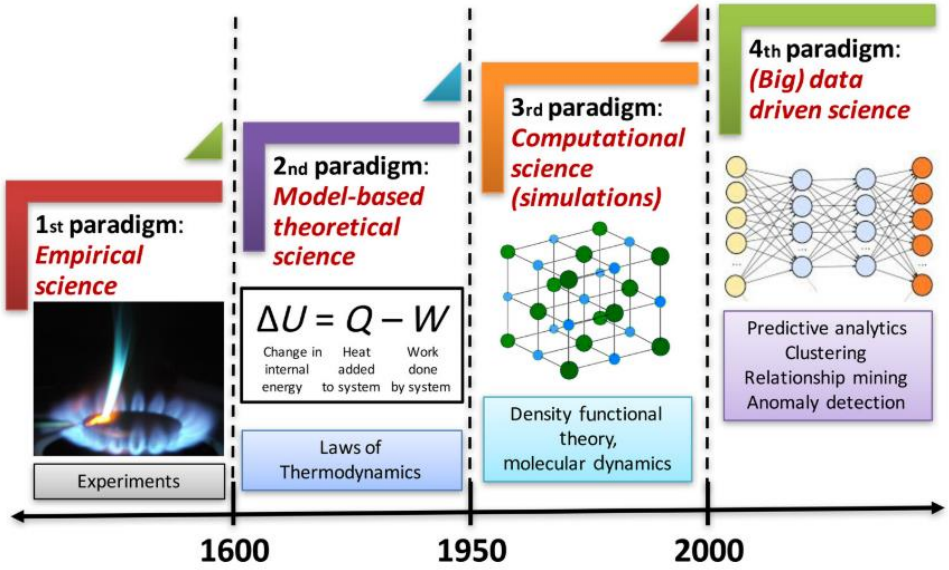
Efficient Material Design aided by Machine Learning Techniques

PRESENTED BY **Liu Shubo**

shuboliu.5guav.com.cn

Paradigm evolution of Material Science

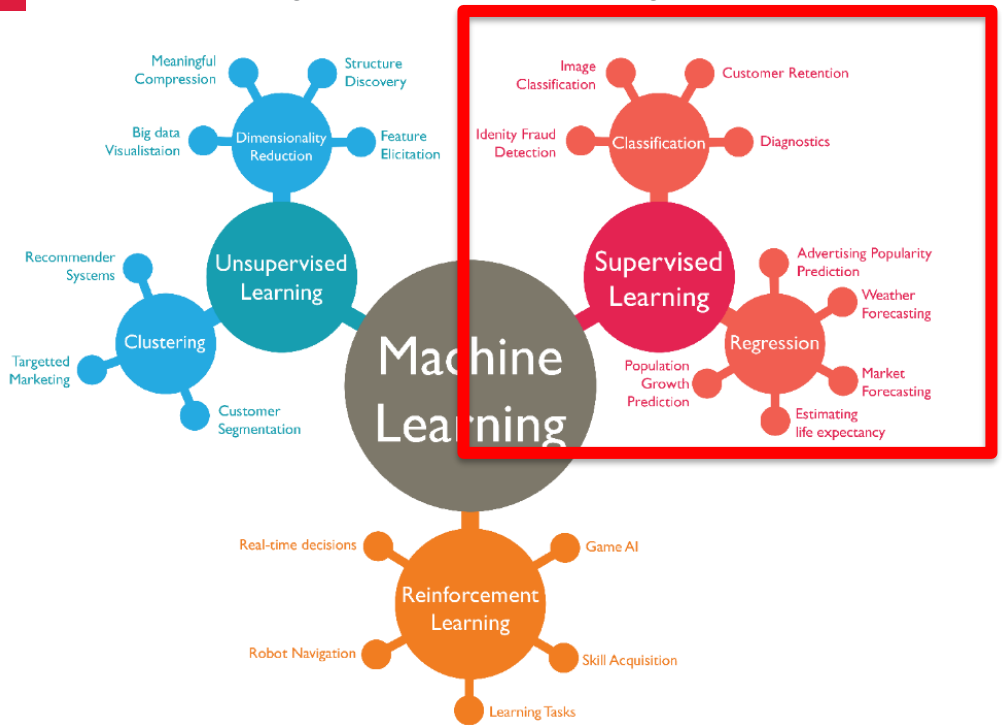
Material research evolutes into big data driven paradigm



- Traditional material science study based on empirical and experience.
- After the invention of calculus in 17th century, quantitively description of natural phenomena became possible.
- Invention of complex computer enables large scale simulation based on DFT and MD method.
- Data science developed in the last two decades enables material researcher to find the rules behind both experimental and simulation data.

ML / DL help material design

Supervised Learning is the key for material design

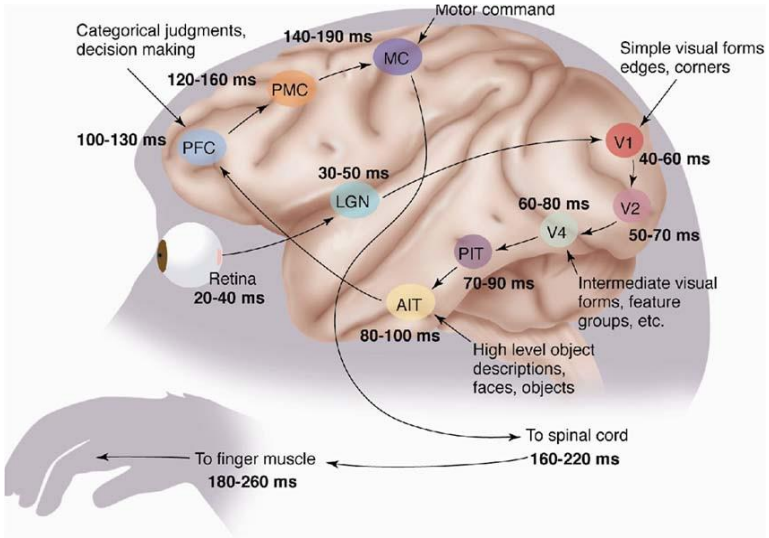


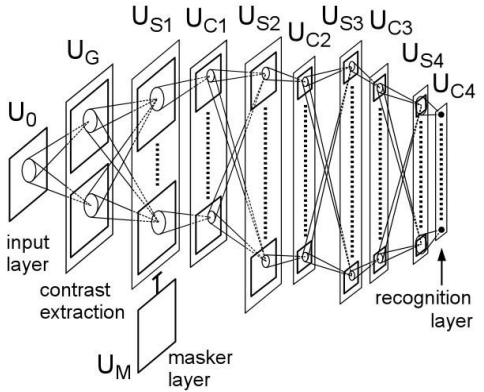
- Supervised learning is the most useful tool in machine learning added material design
- Unsupervised learning is used to clustering task while reinforcement learning is used for game theory
- Deep learning method as a subsequence of machine learning have raised researcher's attentional due to its unimageable ability in multi-factor prediction

Deep Learning: Bio-mimic Structure

Multiple neural network model was proposed by Fukushima

Fukushima

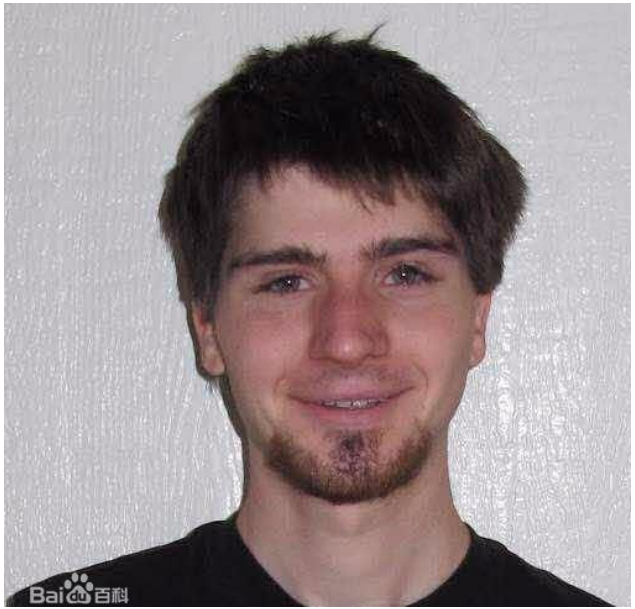




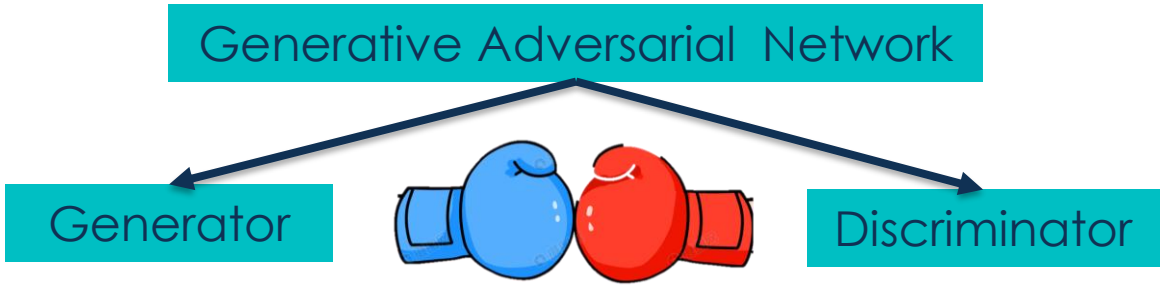
Deep learning network is a typical bio-mimic structure

GAN: A Masterpiece of Two-player Game

Generative adversarial network model was proposed by Ian Goodfellow

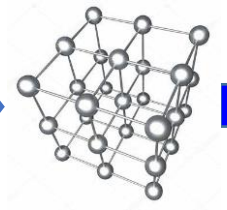


Ian Goodfellow

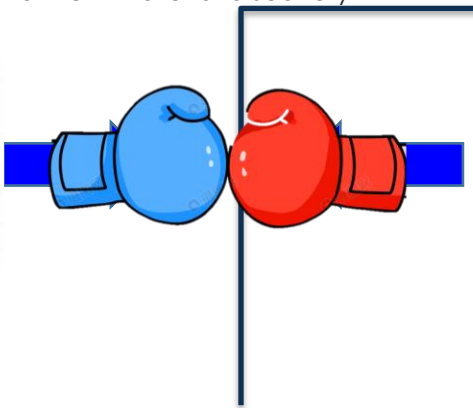


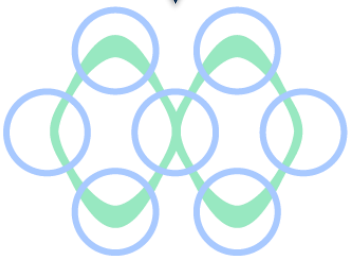
Crystal-GAN: Guide New Material Discovery

Generative Adversarial Networks is a potential way for new material discovery

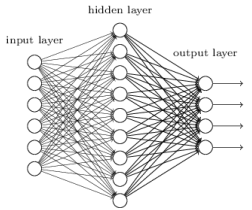


Generated Structure

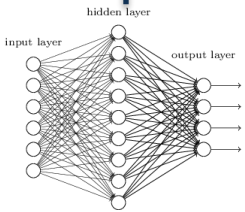




Crystal Structure Lib



Generator



Discriminator

Update Generator

Loss of Generator

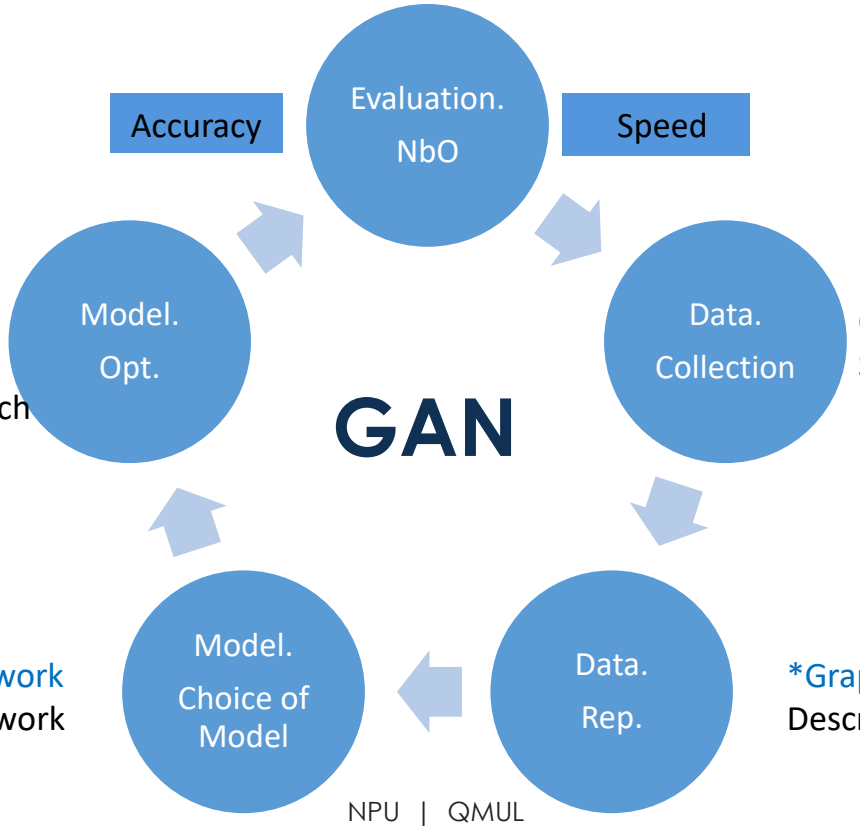
Loss

Loss of Discriminator

Update Discriminator

Crystal-GAN: Data Pipeline

ML/DL added material design method is the hotpot in recent years



GPU Acceleration
Grid Search
Random Search
*Network Architecture Search

Open Database: ICSD, OQMD
Self-built database (DFT)

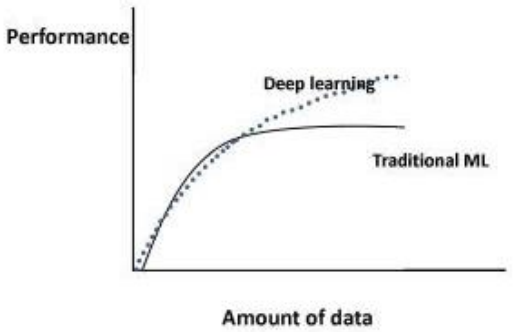
*Graphical Neural Network
Descriptor: MaML

CNN Network
ANN Network

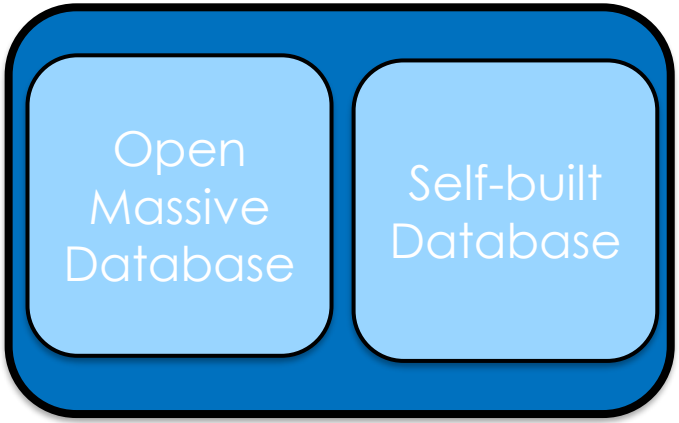
Data Collection

There are three main resources for database building

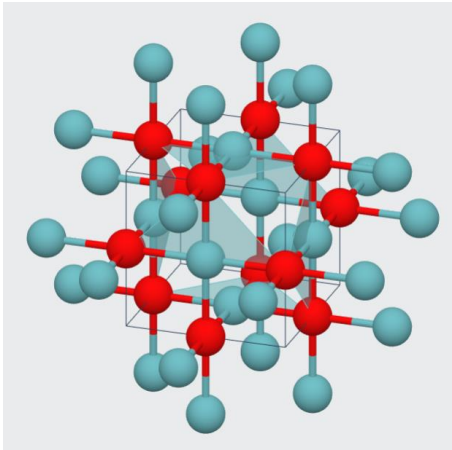
- With the increase of training data amount, performance of trained model improves (Figure below)
- Data Sources: Open Massive Database, Self-built Database
- Open Massive Database includes: ICSD (Material Project), OQMD



Learning Curve



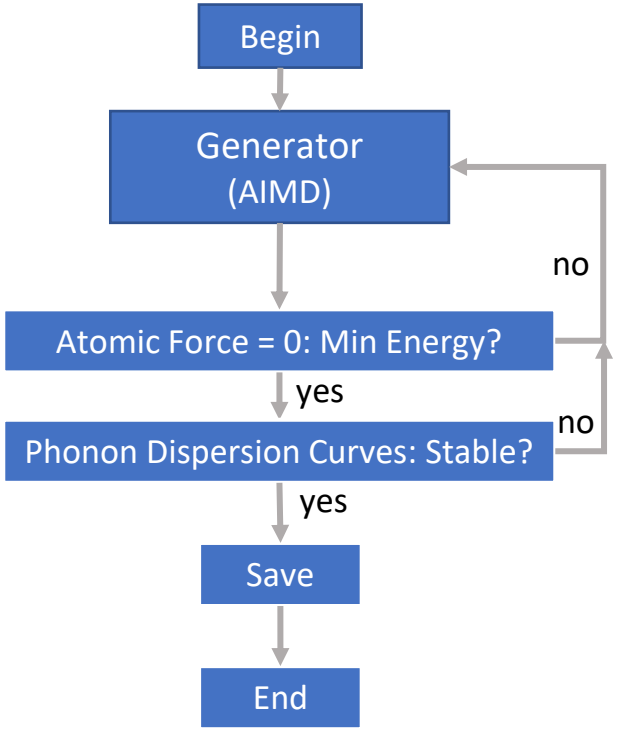
Database Used



Structure of NbO

Data Collection: Self-built Database

There are at least 2 method for self-built database generation:



There are at least 2 method for self-built database generation:

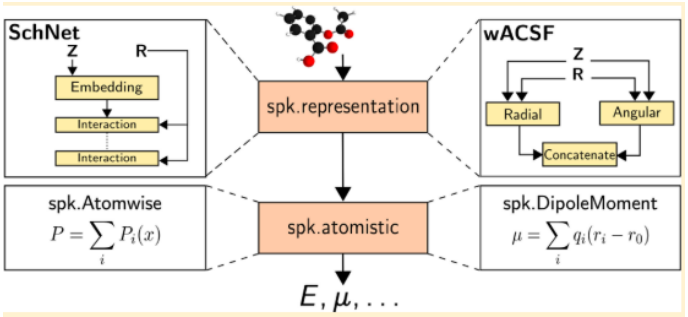
Method 1. Generate structure and use phonon criteria to judge its validation.

~~Method 2. Generated structure forward by phonon spectrum knowledge.~~

Data Representation

There are at least 2 method for data representation method

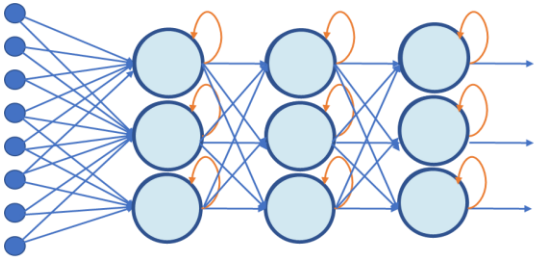
- Data Representation method must cover 4 properties: reference invariance, degeneracy, smoothness and transformability
- Data Representation method might include the following method: Graphical Neural Network (GNN) and Describer
- GNN method is commonly used in molecular dynamic. As a newly developed method, its typical representative includes SchNet, PhyNet



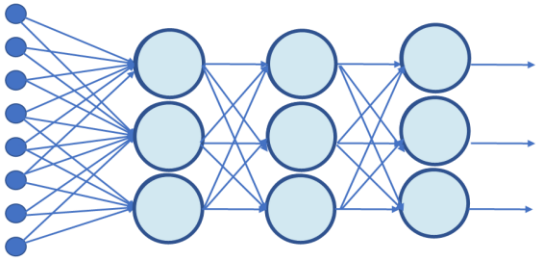
Model Choice

Choice of model is the key to the robustness and accuracy of the final prediction

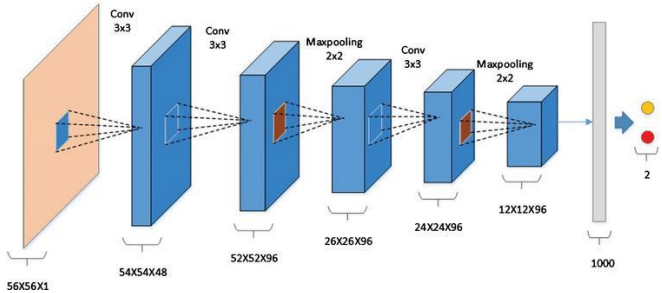
- Choice of model is the key to the robustness and accuracy of the final prediction. It have to follow the nature of data.
- Commonly, hieratical data employs CNN model while sequential data for RNN model and ANN for the rest.



Recurrent Neural Network



Artificial Neural Network



Convolutional Neural Network

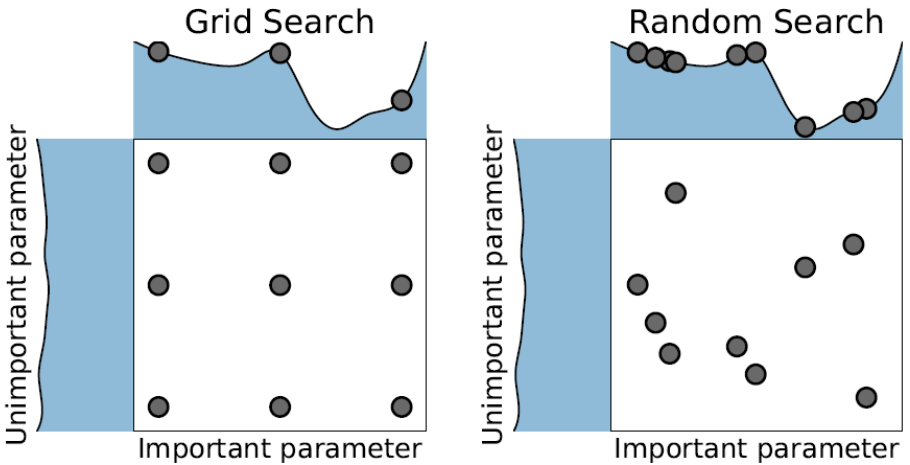
Model Optimization

Model optimization method directly affect the speed of the final prediction

- Optimization of hyperparameters: connectivity between neuron, number of layers, number of neuron in each layers, type of activate function, learning rate, batch size
- Choice of Loss Function

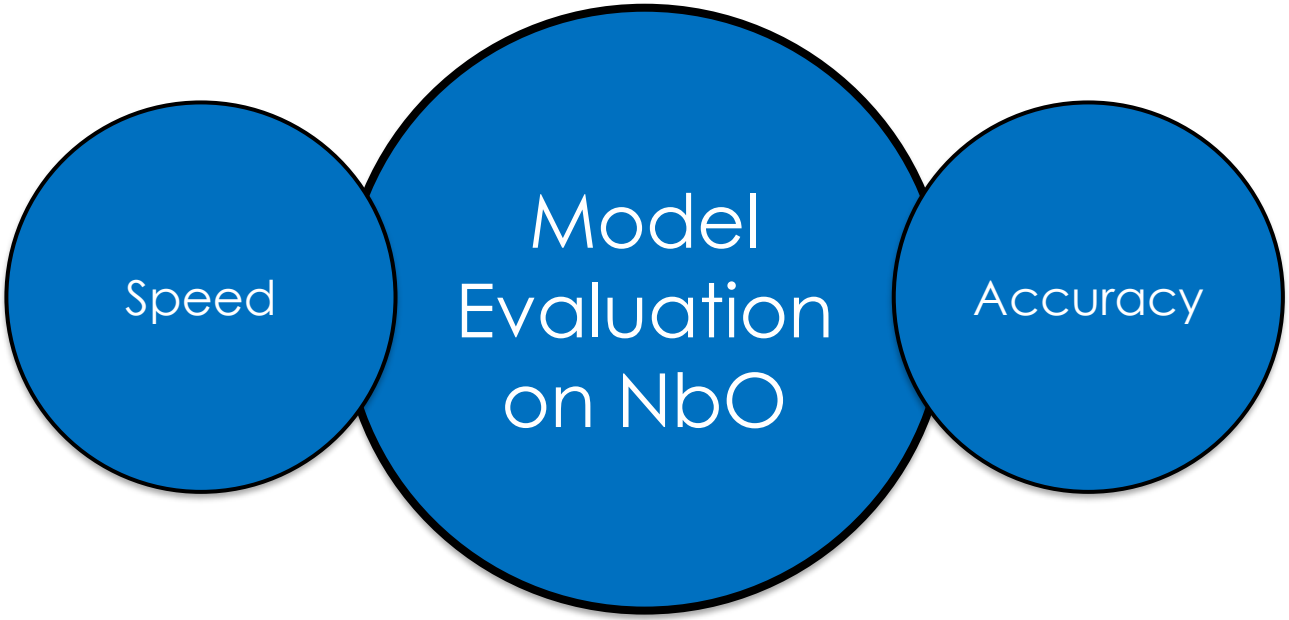
Methods:

- Grid Search algorithm,
- Random Search algorithm
- Network Architecture Search (NAS) algorithm (if possible)



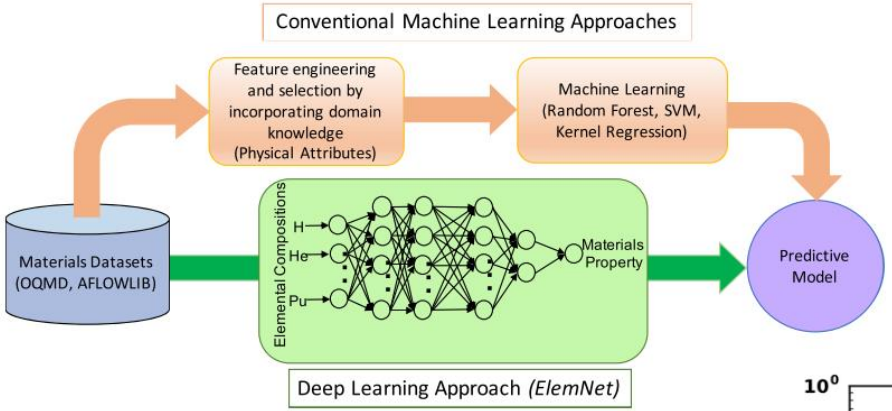
Evaluation: Take NbO system as an example

NbO binary system will be an example to testify the performance of proposed model



Related work: ElemNet

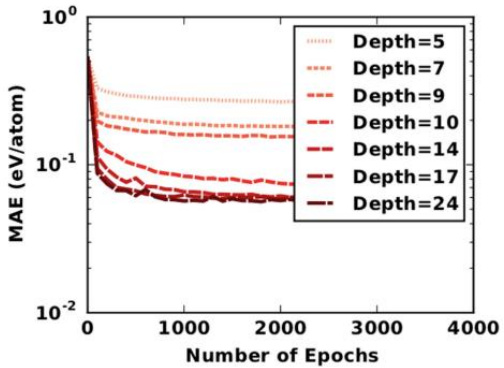
ElemNet that is a deep learning-based material prediction model proposed by Jha, Northwest University



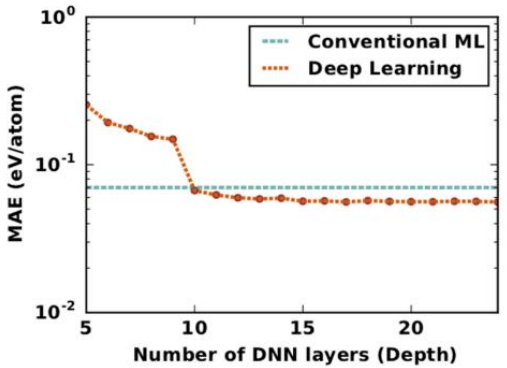
Establish the relationship between chemical composition and macroscopical properties

With the deepen of network, performance will be better, but not always;

More epochs, more accurate



(a) Test Error



(b) Impact of depth

MAE: Mean Absolute Error

D. Jha et al., "ElemNet: Deep Learning the Chemistry of Materials From Only Elemental Composition," Sci. Rep., vol. 8, no. 1, pp. 1–14, 2018, doi: 10.1038/s41598-018-35934-y.

Reference (1)

Efficient Material Design aided by Machine Learning Techniques

[1]A. Agrawal and A. Choudhary, "Deep materials informatics: Applications of deep learning in materials science," MRS Commun., vol. 9, no. 3, pp. 779–792, 2019, doi: 10.1557/mrc.2019.73.
[2]L. Yan, C. Alfredo, G. Mark, and L. Zeming, "Deep Learning," NYU CENTER FOR DATA SCIENCE, 2020. https://atcold.github.io/pytorch-Deep-Learning/.
[3]K. Ryczko, D. A. Strubbe, and I. Tamblyn, "Deep learning and density-functional theory," Phys. Rev. A, vol. 100, no. 2, 2019, doi: 10.1103/PhysRevA.100.022512.
[4]W. Ye, C. Chen, Z. Wang, I. H. Chu, and S. P. Ong, "Deep neural networks for accurate predictions of garnet stability," Nat. Commun., pp. 1–18, 2018.
[5]J. E. Saal, S. Kirklin, M. Aykol, B. Meredig, and C. Wolverton, "Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD)," JOM, vol. 65, no. 11, pp. 1501–1509, 2013, doi: 10.1007/s11837-013-0755-4.
[6]P. Karamanis, "The Importance of the DFT method on the computation of the second hyperpolarizability of semiconductor clusters of increasing size: A critical analysis on prolate aluminum phosphide clusters," John Wiley & Sons, Ltd, 2012.
[7]D. Jha et al., "ElemNet: Deep Learning the Chemistry of Materials From Only Elemental Composition," Sci. Rep., vol. 8, no. 1, pp. 1–14, 2018, doi: 10.1038/s41598-018-35934-y.
[8]N. Artrith and A. M. Kolpak, "Understanding the Composition and Activity of Electrocatalytic Nanoalloys in Aqueous Solvents: A Combination of DFT and Accurate Neural Network Potentials," Nano Lett, vol. 14, pp. 2670–2676, 2014, doi: 10.1021/nl5005674.
[9]L. Huang and C. Ling, "Practicing deep learning in materials science: An evaluation for predicting the formation energies," J. Appl. Phys., vol. 128, no. 12, 2020, doi: 10.1063/5.0012411.
[10]T. Xie and J. C. Grossman, "Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties," Phys. Rev. Lett., vol. 120, no. 14, 2018, doi: 10.1103/PhysRevLett.120.145301.
[11]W. Ye, C. Chen, Z. Wang, I. H. Chu, and S. P. Ong, "Deep neural networks for accurate predictions of crystal stability," Nat. Commun., vol. 9, no. 1, pp. 1–6, 2018, doi: 10.1038/s41467-018-06322-x.

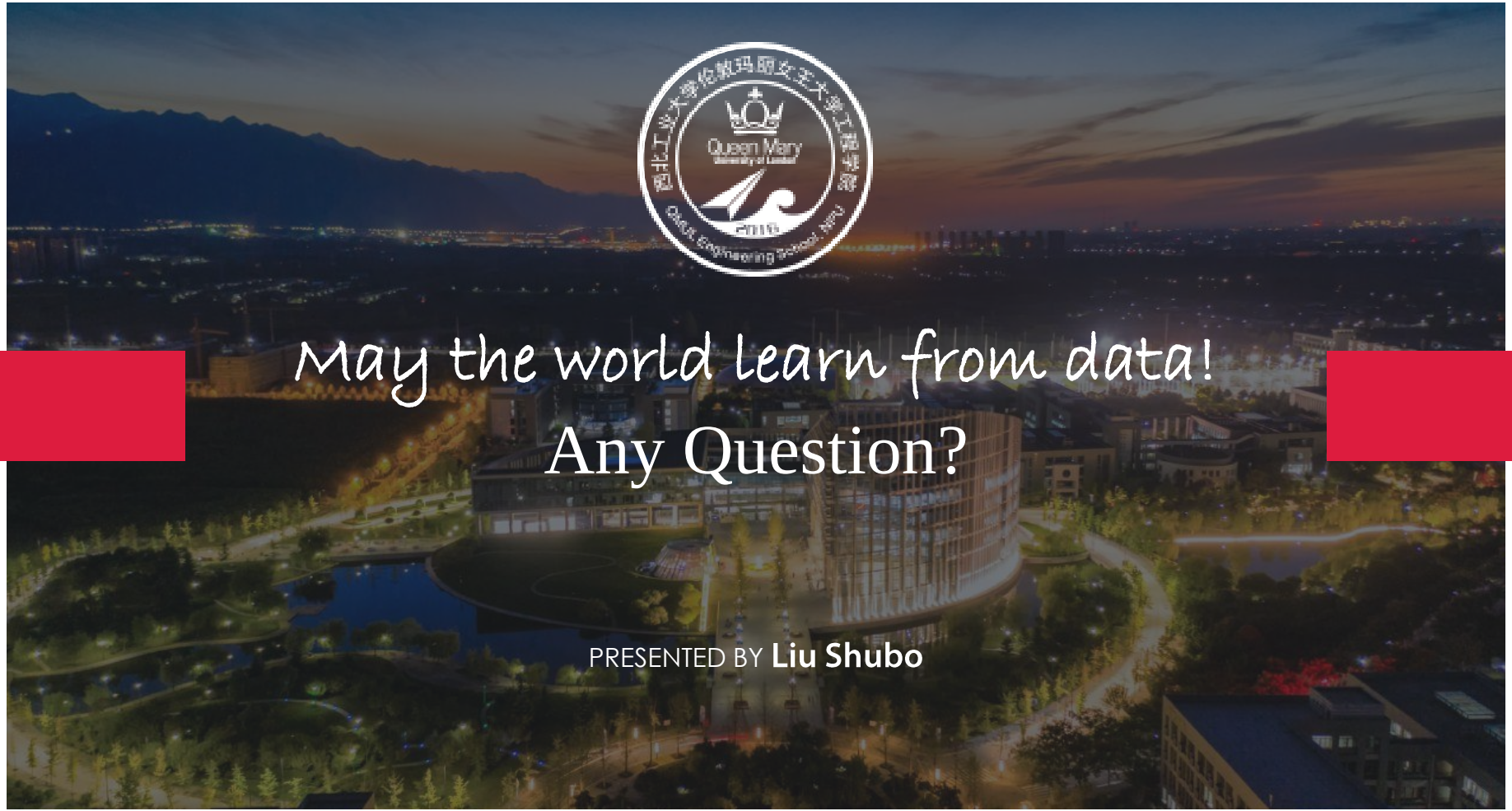
Reference (2)

Efficient Material Design aided by Machine Learning Techniques

[12]"Latest Trends on Computer Vision Market – InData Labs Blog." https://indatalabs.com/blog/trends-computer-vision-software-market?cli_action=1555888112.716 (accessed Dec. 25, 2020).
[13]X. Zhu, C. Vondrick, C. C. Fowlkes, and D. Ramanan, "Do We Need More Training Data?," Int. J. Comput. Vis., vol. 119, no. 1, pp. 76–92, 2016, doi: 10.1007/s11263-015-0812-2.
[14]J. Hestness et al., "Deep learning scaling is predictable, empirically," arXiv, pp. 1–19, 2017.
[15]S. Kirklin et al., "The Open Quantum Materials Database (OQMD): Assessing the accuracy of DFT formation energies," npj Comput. Mater., vol. 1, no. November, 2015, doi: 10.1038/npjcompumats.2015.10.
[16]M. Lu, "CNN vs.RNN vs.ANN - Analysis of Three Deep Leaning Networks," Zhihu.com, 2020. https://zhuanlan.zhihu.com/p/107993566 (accessed Dec. 24, 2020).
[17]Hoogle, "Introduction to Black Box Optimization," Zhihu.com, 2019. https://zhuanlan.zhihu.com/p/66312442 (accessed Jan. 12, 2021).
[18]B. Da, "Deep Learning No.10: Optimization of Hyperparameters," Zhihu.com, 2018. https://zhuanlan.zhihu.com/p/69025104 (accessed Jan. 12, 2021).
[19]K. Georg, V. Doris, and M. Martijn, "Vienna Ab initio Simulation Package." VASP Software GmbH, Vienna, 2017, [Online]. Available: https://www.vasp.at/.
[20]Y. Lanren, "ResNet Analysis," CSDN blog, Jan. 14, 2018. https://blog.csdn.net/lanran2/article/details/79057994 (accessed Jan. 08, 2021).

Acknowledgement

I would like to express a great thank to Dr. Zhenwei Li from NPU, Dr. Lei Su from QMUL for their great contribution and Mr. Hongsheng Zhang, Mr. Yufei Bai and Ms. Shuang Feng from NPU for their huge support



May the world learn from data!

Any Question?

PRESENTED BY **Liu Shubo**