Introduction to Machine Learning, Materials Imaging, and Segmentation

Dr. Tiberiu Stan – tiberiu.stan@northwestern.edu

Northwestern University

MatSci 395 – ML Lecture 1 – March 1, 2021

Who Am I? How did I get here?

Dr. Tiberiu Stan

- Research Associate with Prof. Peter Voorhees
- PhD in Materials Science and BS in Physics from UCSB

Research Interests

- Metallurgy, phase transformations, solidification, x-ray tomography, artificial intelligence, big data processing

MatCNN Software

- Segmentation is one of the most challenging parts of materials image analysis
- Development of MatCNN to segment images
- I have taught researchers from around the world how to use MatCNN

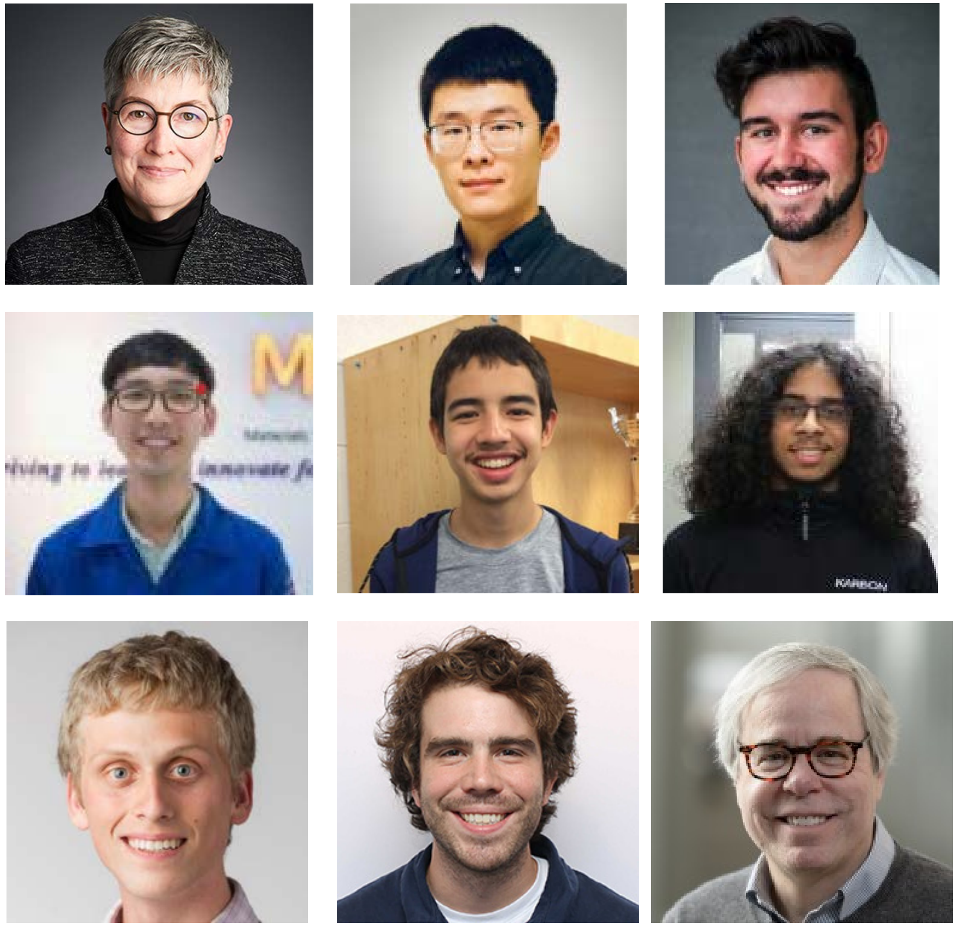
Acknowledgements

Thank you

- Prof. Elizabeth Holm and Dr. Bo Lei (CMU) – initial collaboration to develop MatCNN
- Joshua Pritz (NWU) and Jiwon Yeom (KAIST) – Improvements to Python code
- Nathan Pruyne, Jim James, Marcus Schwarting (Argonne) – Adaptation to Google Colab
- Prof. Jonathan Emery (NWU) – Course development
- Prof. Peter Voorhees (NWU) – General guidance

Funding by the Center for Hierarchical Materials Design (CHIMAD)

CHIMaD

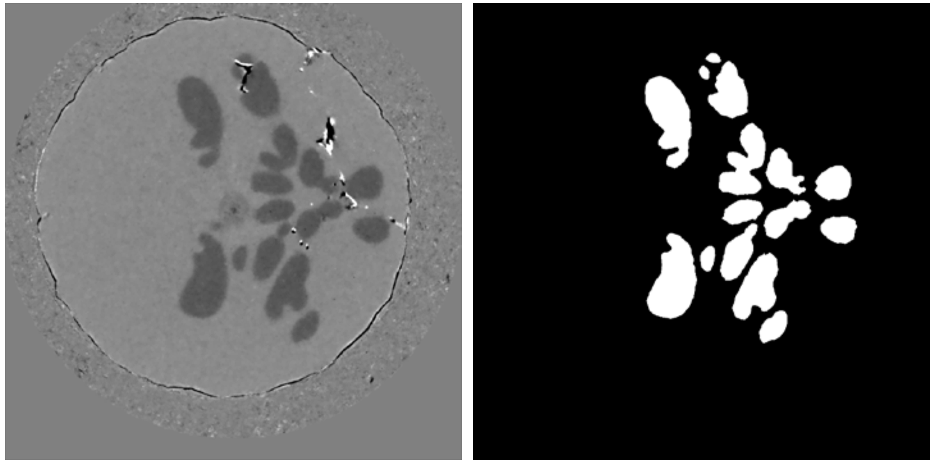


Scope

By the end of this module you should be able to

- Understand the fundamental construction and functionality of Convolutional Neural Networks
- Segment materials science images both by hand, and using the machine learning code MatCNN
- Extract useful materials information from images

Image Segmentation



Tentative Schedule

Monday	Wednesday	Friday
L1 – Introduction to machine learning, imaging, and segmentation	L2 – **In-class activity: segmentation using GIMP**	L3 – Convolutional Neural Network basics
L4 – Training Convolutional Neural Networks and MatCNN introduction	L5 – **In-class activity: segmentation using MatCNN**	L6 – Segmentation parameters, metrics, and other questions

Please download and install GIMP before next class

Growing Importance of Data-Driven Approaches

Materials Genome Initiative (MGI)

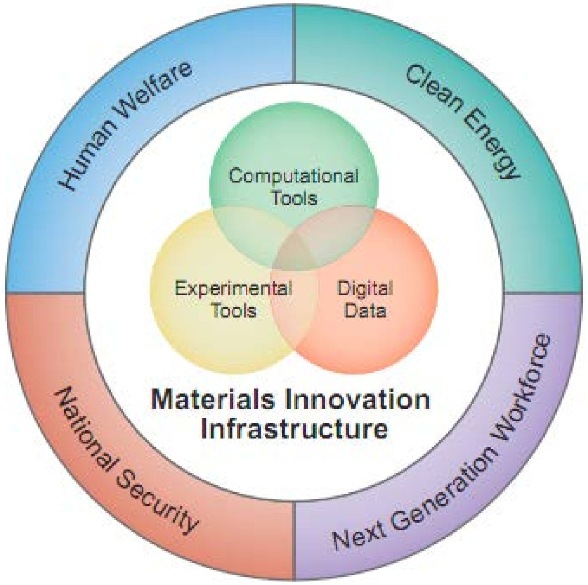


Figure 3: Initiative overview

- **Challenge:** It takes years to get technologies into the marketplace
- **Goal:** Help businesses discover, develop, and deploy new materials twice as fast by reducing the time-to-market by 50%
- **Approach:** Artificial intelligence has enhanced and accelerated our ability to process data

Artificial Intelligence Overview

Artificial Intelligence (AI) - The science of making things smart

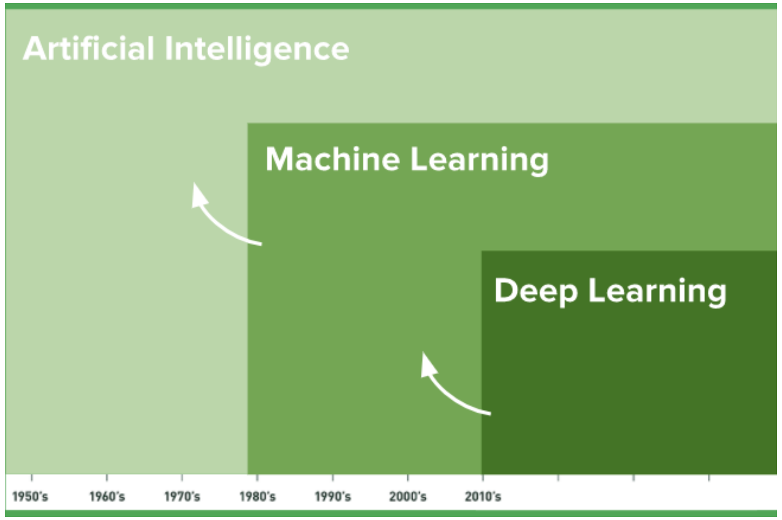
- Object and speech recognition
- Natural Language Processing (NLP)

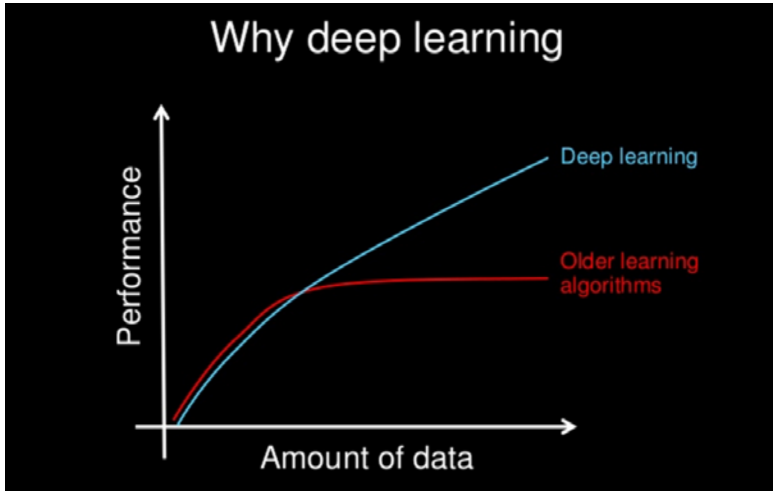
Machine Learning (ML) - Achieving AI though systems that can "learn from experience" to find patterns in data

- Systems recognize patterns by example, rather than programing it with specific rules
- ML algorithms learn complex functions (or patterns) from the data and make predictions on it

Deep Learning - Layered-code structures that mimic the brain

- Convolutional Neural Networks (CNNs) – **MatCNN**
- Generative Adversarial Networks (GANs)

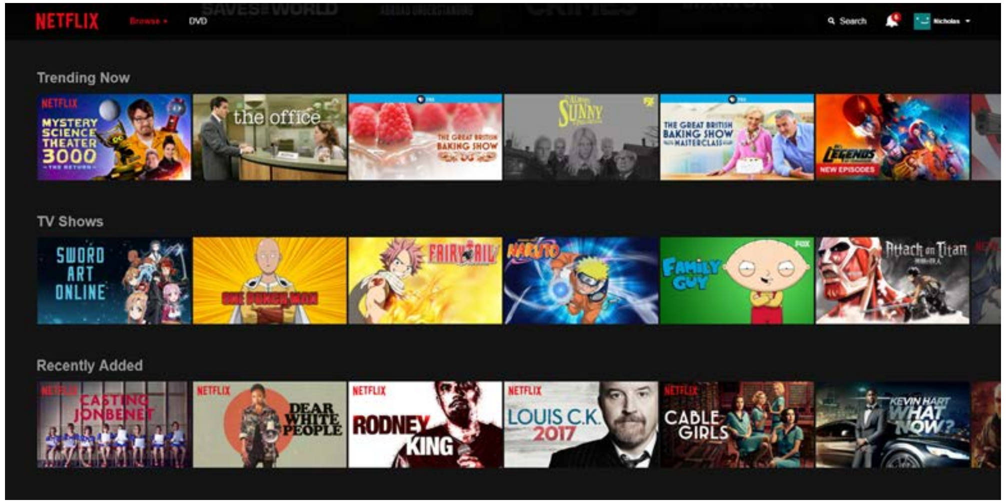


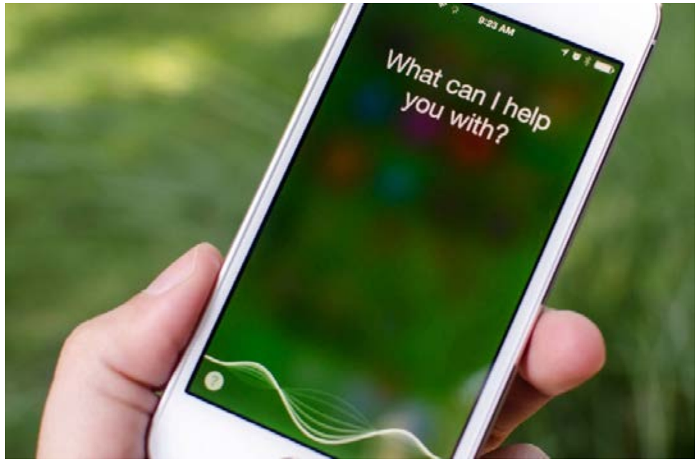


Machine Learning in the Customer Experience

Companies use ML to improve customer experience

- Google: Photos, Smart Search, Gmail, etc.
- Amazon: recommended purchases
- Netflix: recommended shows and movies
- Apple: News app
- Facebook: photo tagging
- Robotics: object recognition
- Many more examples…



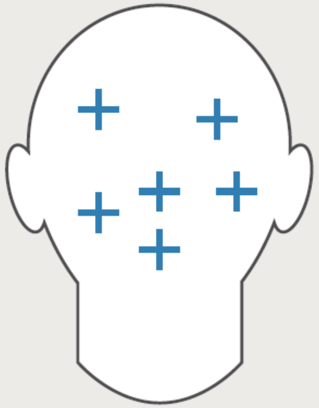


When to Use Machine Learning

Machine learning is best used when

- Problems are complex and relationships between variables are not obvious
- Problems contain a large amount of data and/or variables
- There are no existing formulas or equations

Hand-written rules and equations are too complex—as in face recognition and speech recognition.



The rules of a task are constantly changing—as in fraud detection from transaction records.

The nature of the data keeps changing, and the program needs to adapt—as in automated trading, energy demand forecasting, and predicting shopping trends.

Types of Machine Learning

Machine Learning

Unsupervised Learning

Clustering

- Group and interpret data based on input data
- Describe structure in data without known values

Supervised Learning

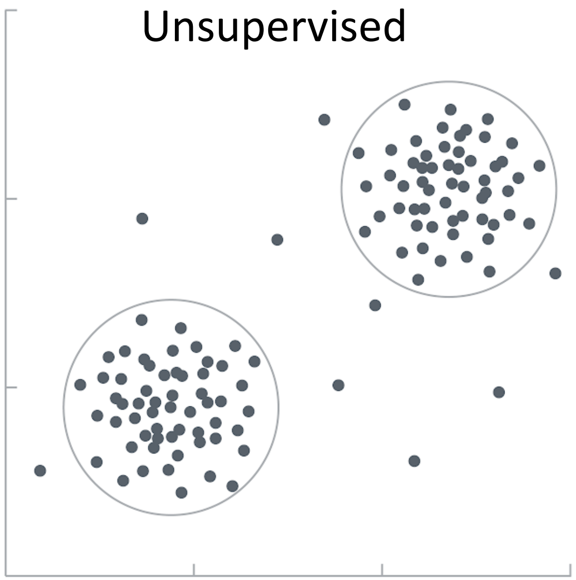
Classification

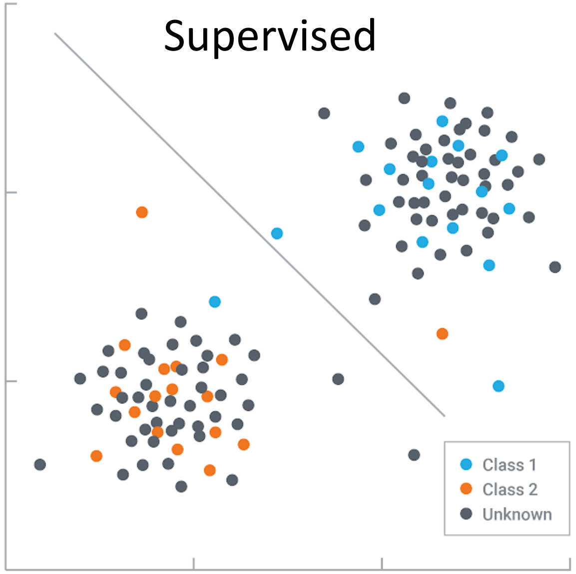
Regression

- Develop predictive models based on examples of both input and output data
- Predict a value based on some known variables

Reinforcement Learning

- Learn optimal behaviors given a state of the world and a desired reward





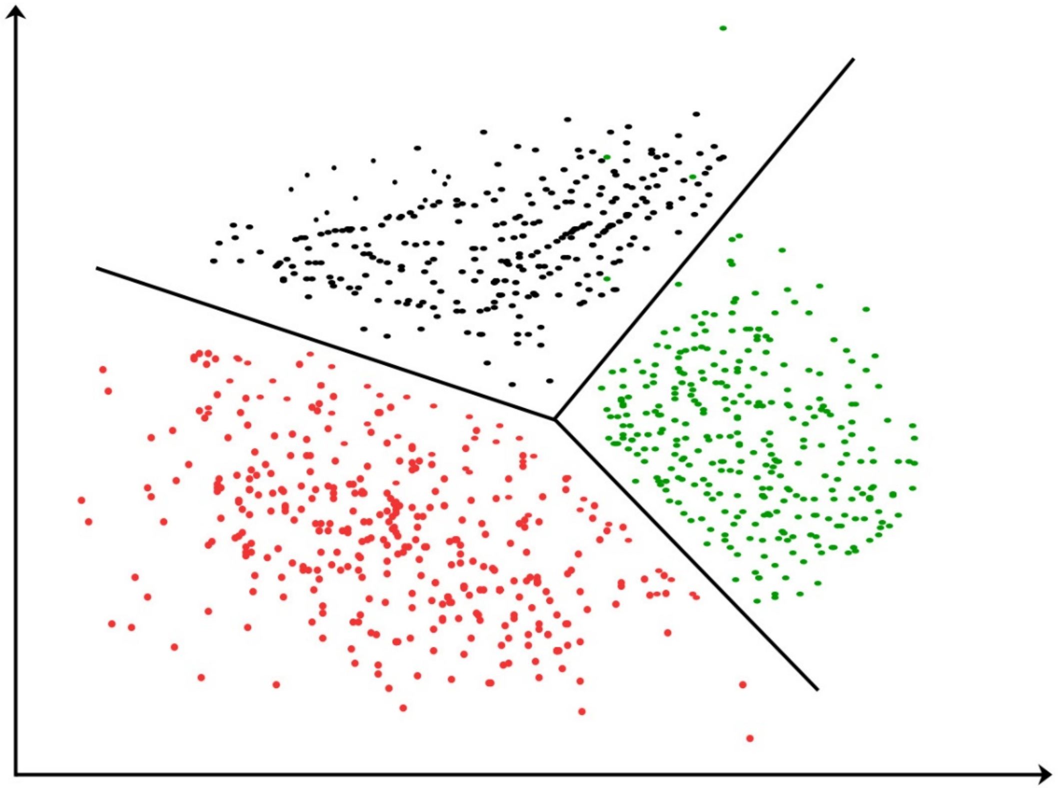
Unsupervised Learning

Goals

- Find patterns or intrinsic structure in the data
- Make inferences from the data sets consisting of input data without labeled responses

Clustering

- The most common technique
- Used for exploratory data analysis to find hidden patterns in groupings of data
- Examples: K-means clustering, spectral clustering, Principal Component Analysis (PCA), t-SNE
- Uses: gene sequence analysis, market research, object recognition



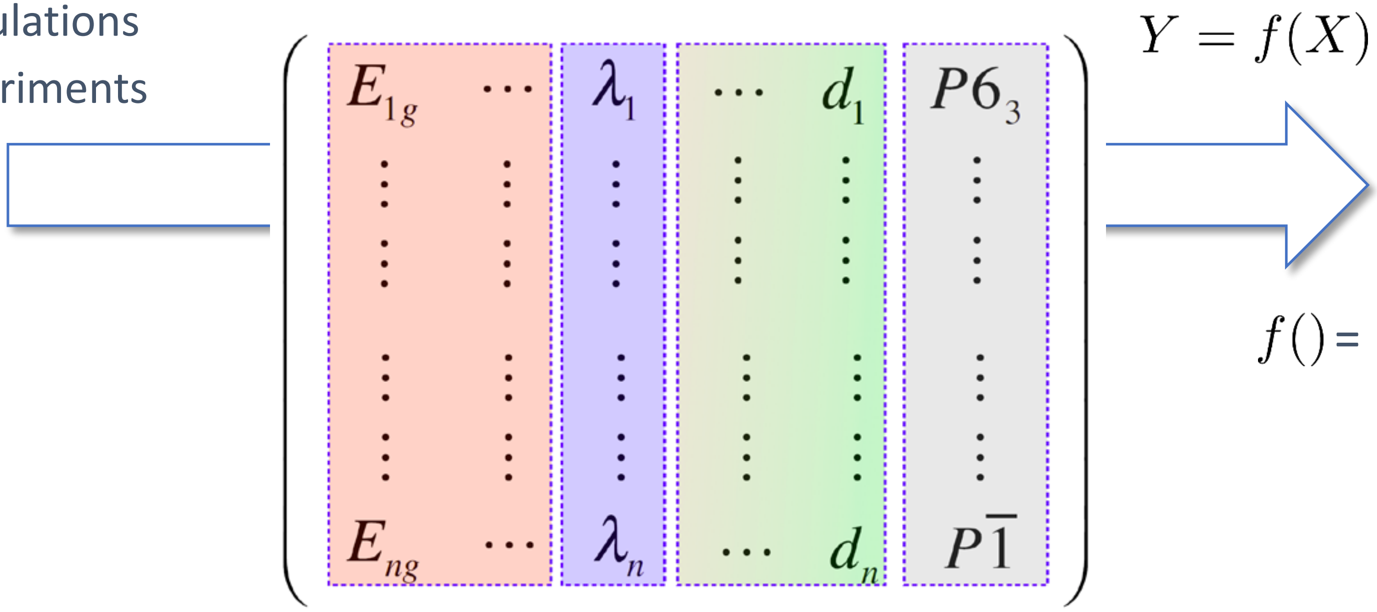
Supervised Learning

Goal: Build a model that makes predictions based on evidence

Algorithms are "trained" on a set of input data (X = E_{ng}, λ_n, d_n, etc.) and known output responses to the data (Y) to generate a model (Y = *f*(X)) that gives reasonable predictions when applied to new data

Calculations

Experiments

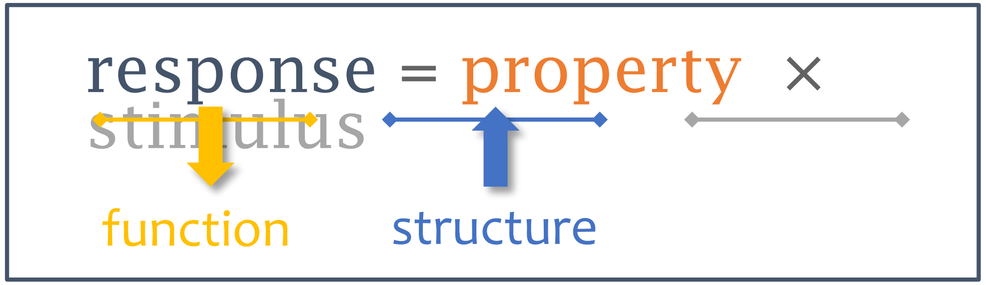


$f()$ = Learning/classification method based on probability theory

Why is Supervised Learning Attractive to MSE?

What is a material property?

- Relationship between two measurable quantities (density = mass / volume)
- Formally - the response of a material to a specific change in a given set of conditions that relates independent and dependent quantities in a particular process
- There exists an underlying structure to properties in constitutive relationship, which are mathematical, thus computers should be able to learn them given enough data and appropriate features



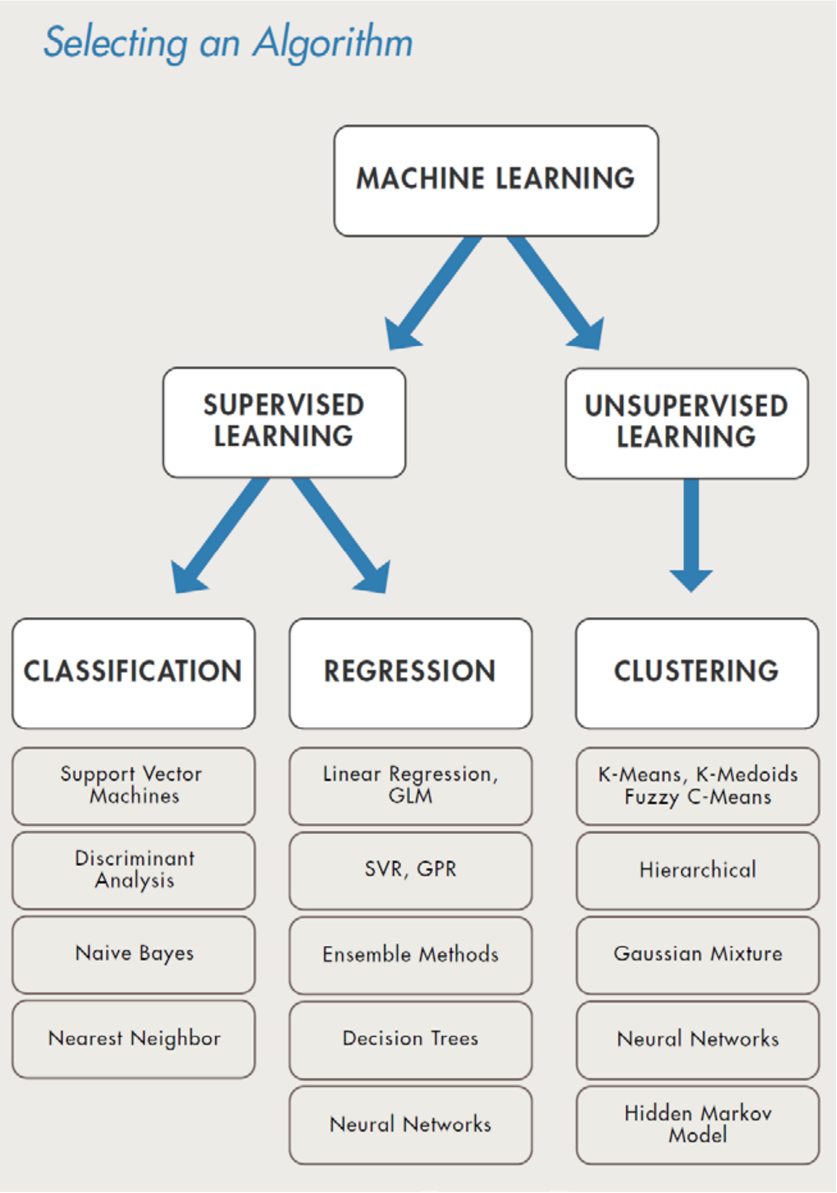
Which Algorithm to Use?

Choosing an algorithm can be overwhelming

- Model should be simple but not simpler (Einstein)
- Diverse algorithms with different approaches
- Not obvious a priori which is best
- Often requires trial-and-error

Algorithm selection will often depend on

- Size of the dataset you are working with
- Insights you want from the data
- Use of those insights

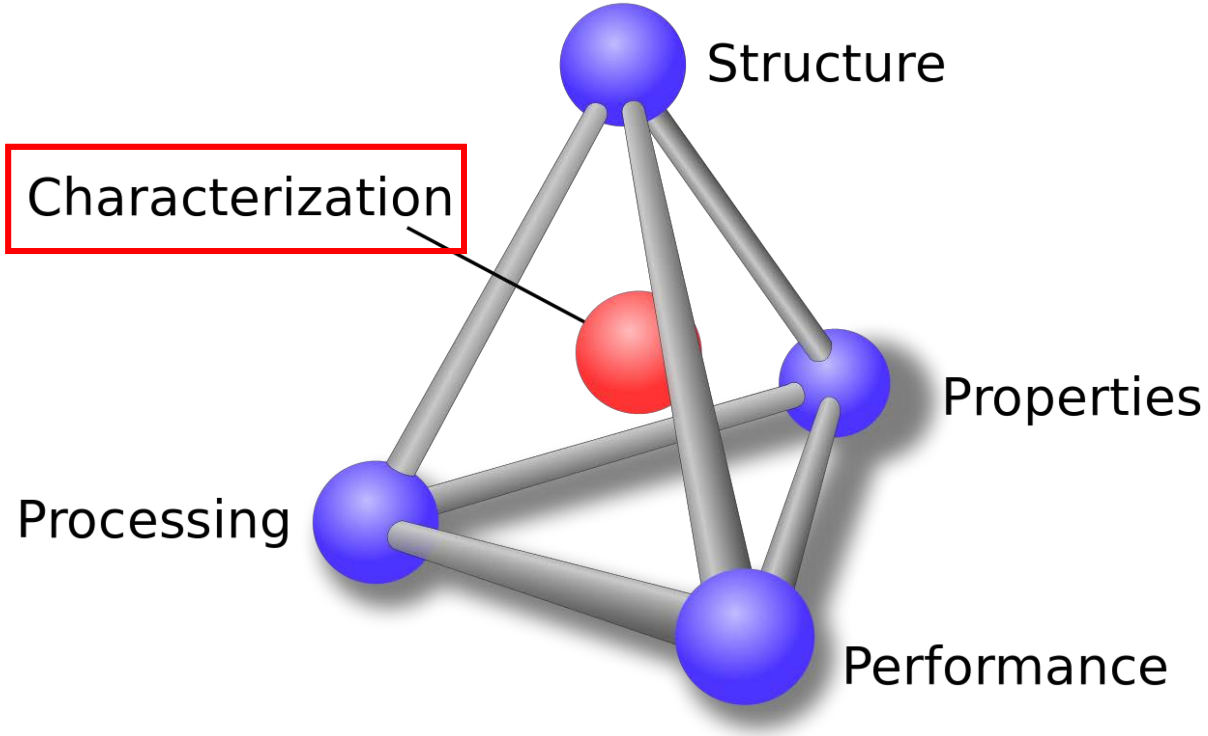


Imaging in Materials Science

Materials imaging is a form of characterization

By taking images we learn about

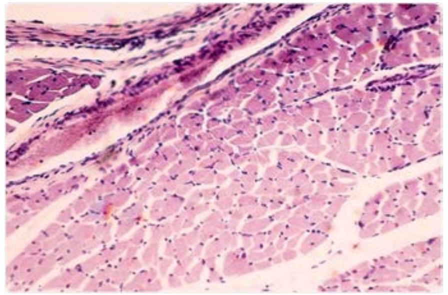
- The structure of materials
- How processing influenced that structure
- How the structure impacts properties



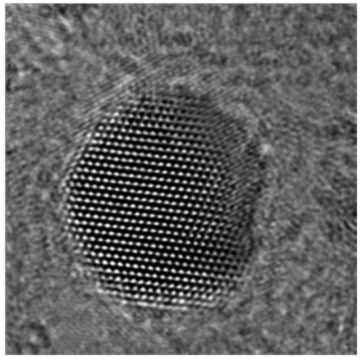
What Is Imaging?

Imaging is

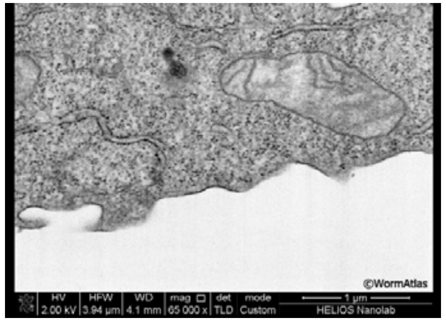
- A visual representation or reproduction of an object's form
- A measurement of how a "source" interacted with an object of interest
- Some common sources in materials imaging are:



Visible Light
(Optical Microscopy)



Electrons
(Electron Microscopy)



Ions
(Focused Ion Beam
Microscopy)

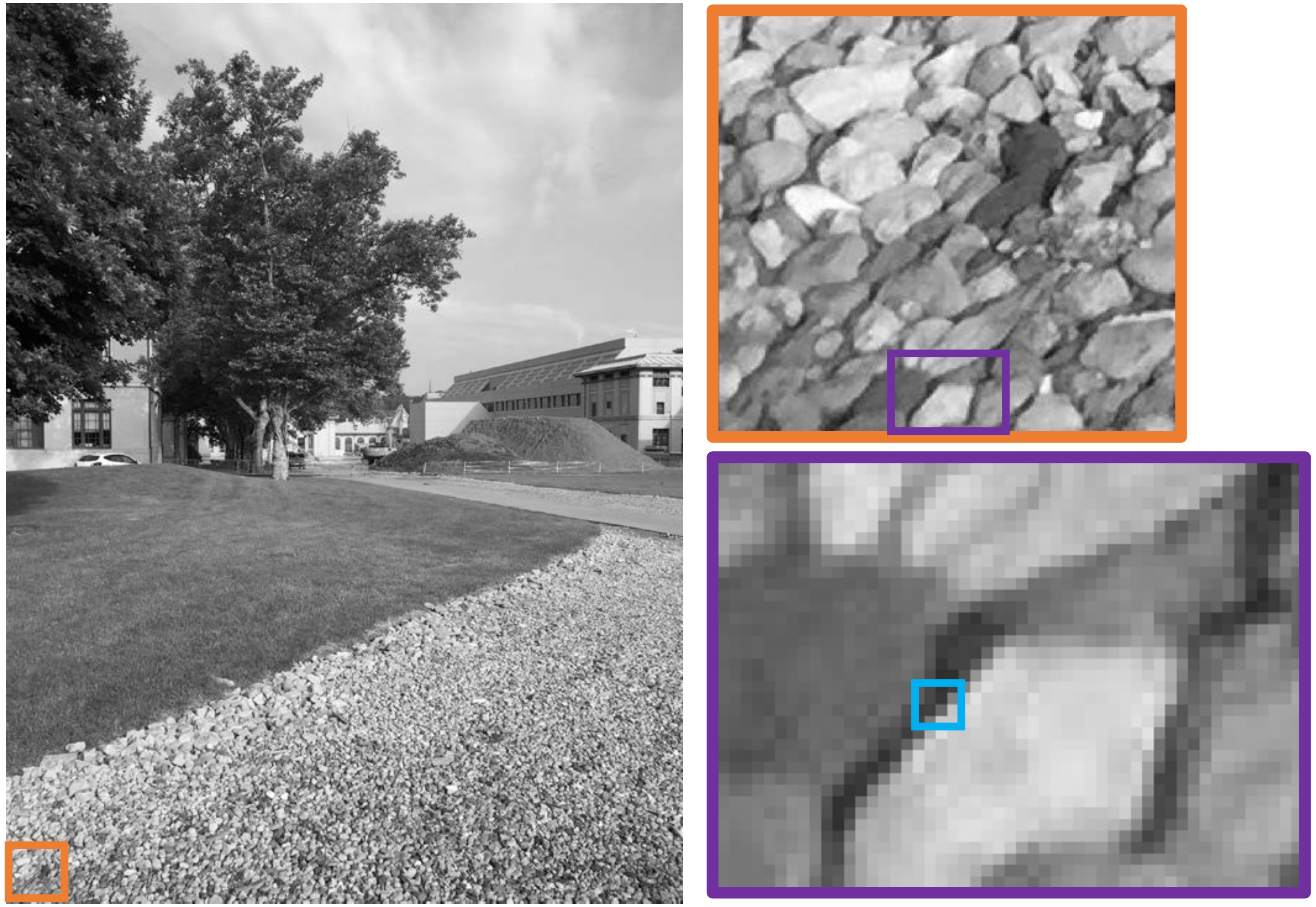
X-Rays
(X-Ray Computed Tomography)

https://www.news-medical.net/life-sciences/Limitations-of-Optical-Microscopy.aspx
https://www.uni-due.de/cenide/ican/tem_index_en
https://www.wormatlas.org/EMmethods/FIBSEM.htm
https://en.wikipedia.org/wiki/X-ray_microtomography

What Is Imaging?

What happens when you take a photo?

- The interaction of light with the objects of interest is stored as intensity values within a grid of "pixels". Intensity values are usually normalized to [0 - 255]



3x3 Pixel Sub-Image

46	47	79
51	63	128
55	128	201

Segmentation

Image segmentation is the act of partitioning an image into multiple useful sections

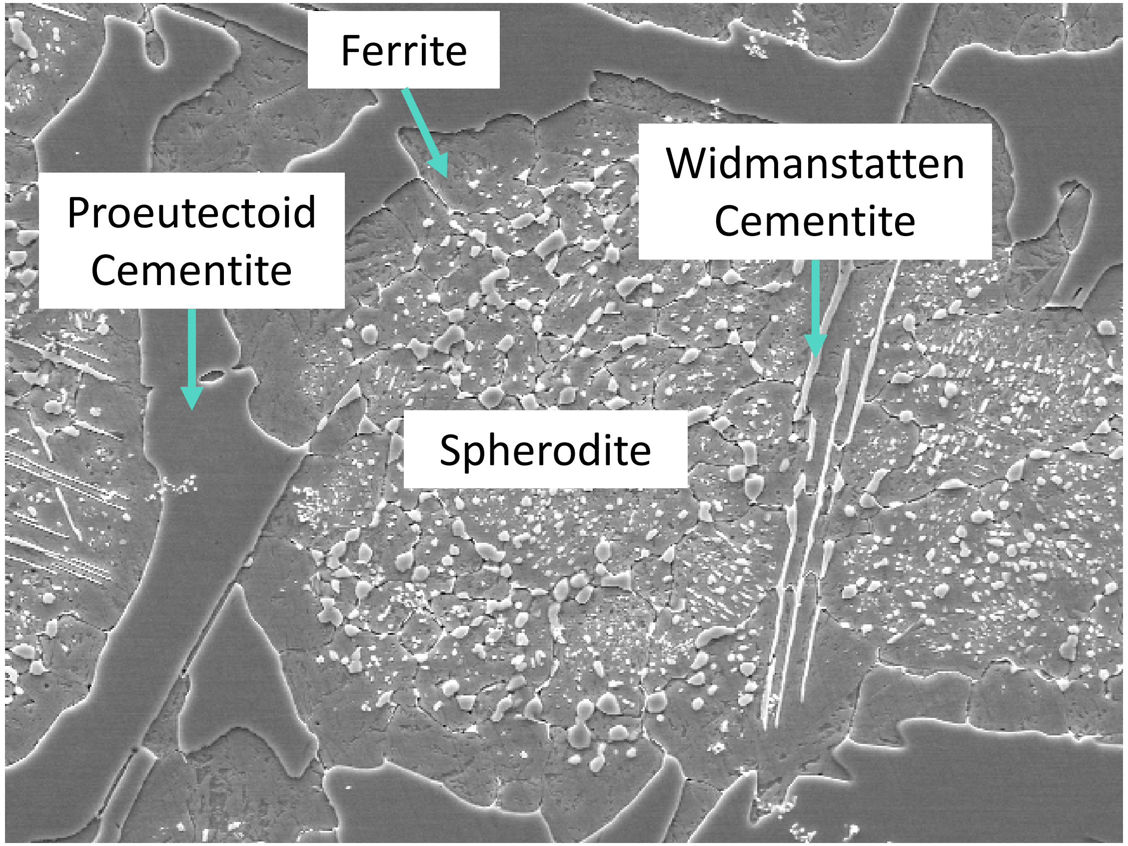
- Semantic segmentation: using machine learning to assign a class to each pixel in an image
- Many self-driving cars use neural networks to segment images

Image | Semantic Segmentation

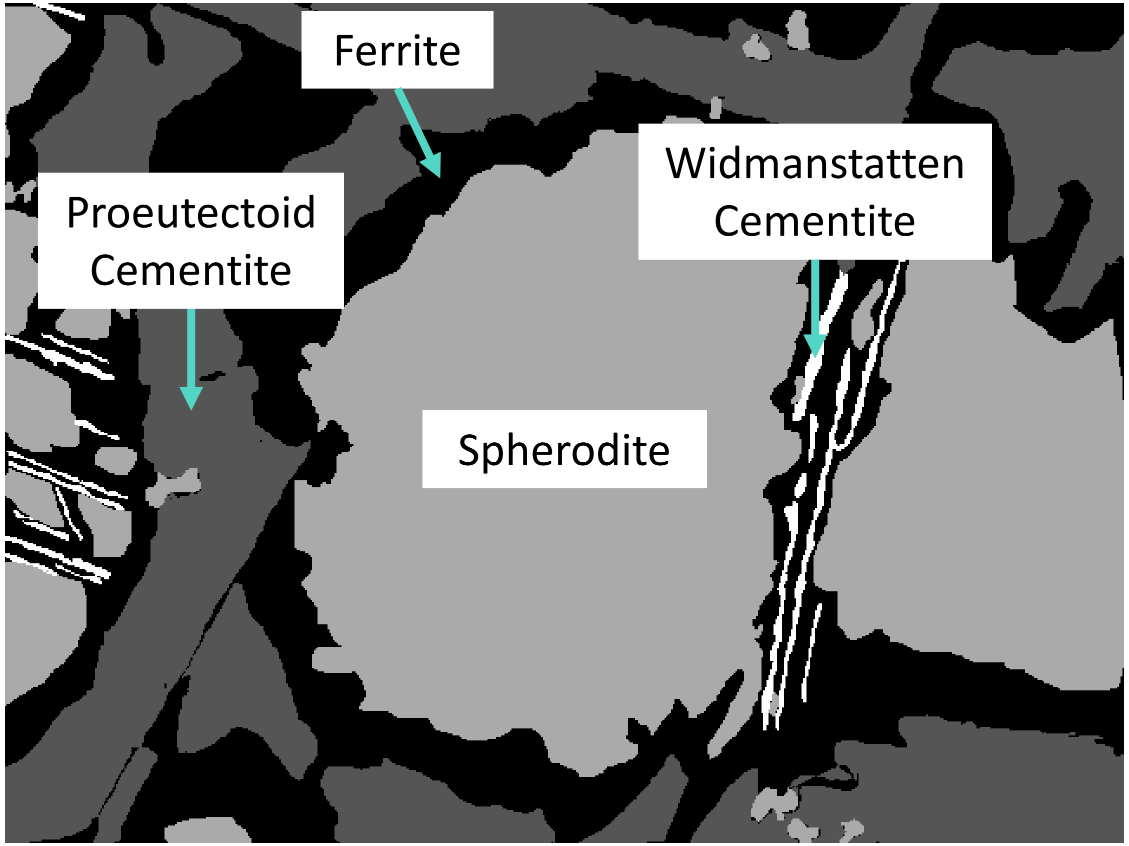
Segmentation

- Segmentation is used in materials science to count particle sizes, determine object shapes, determine area fractions of different phases, etc.
- Scanning electron microscopy image from an ultra-high carbon steel with 4 phases

Image



Manual Segmentation

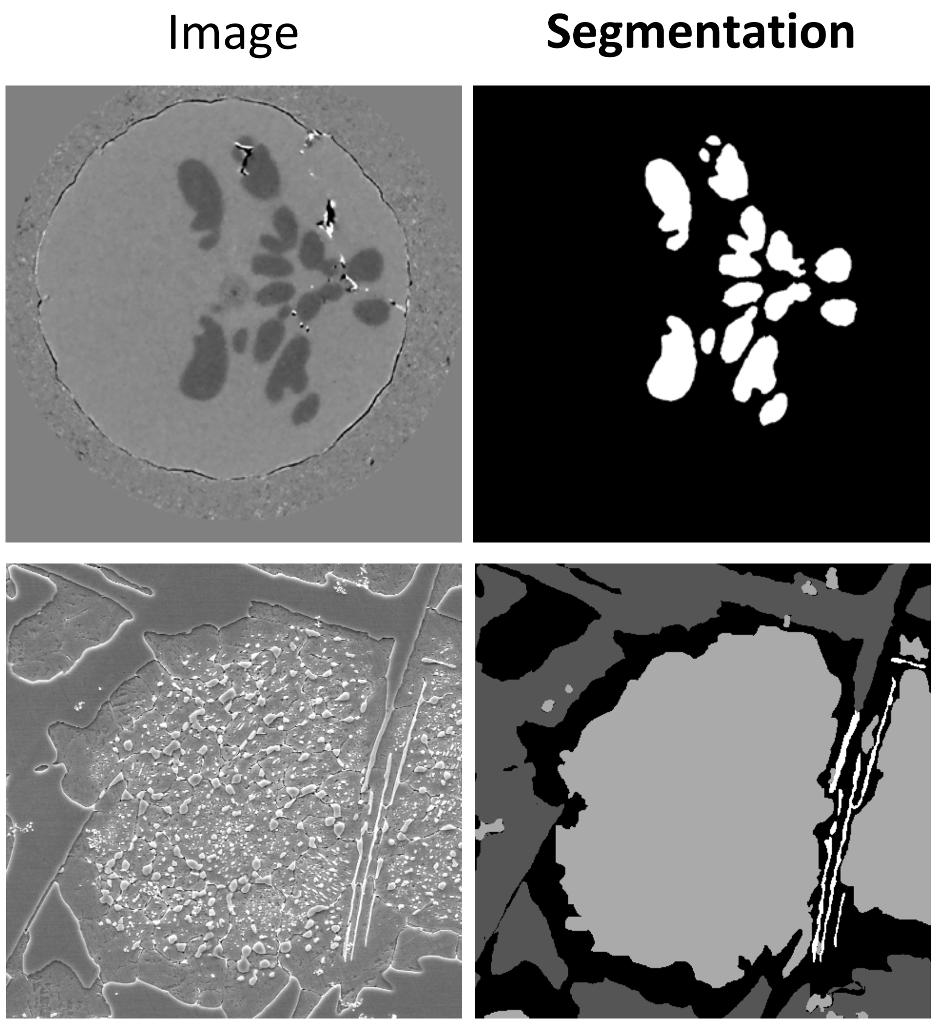


B. De Cost et al. *Microscopy and Microanalysis* 2018

Summary

Lecture Summary

- Artificial intelligence is a useful tool for processing large datasets and accelerating the rate of materials discovery and development
- An image measures the interaction of a source with an object of interest
- Segmentation is the act of partitioning an image into multiple useful parts



Next lecture

- Introduction to solidfication and x-ray tomography
- In-class activity: manual segmentation using GIMP
- Download and install GIMP before class