Explaining Neural Network Predictions of Material Strength

(Machine Learning for Industry Forum 2021)

T. Nathan Mundhenk*
(mundhenk1@llnl.gov)
Ian A. Palmer* (MIT/USAF)
Brian Gallagher
Barry Chen
Gerald Friedland (UC Berkeley)
Yong Han

* Equal Contribution

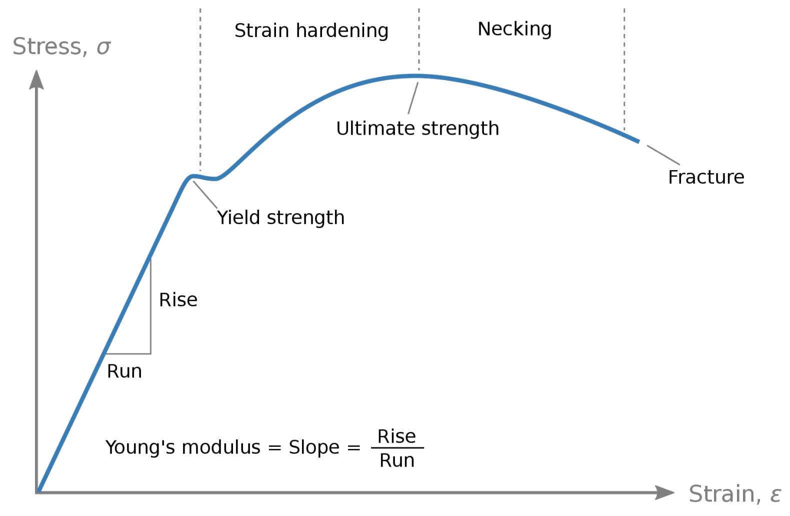
LLNL-PRES-821708
This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC

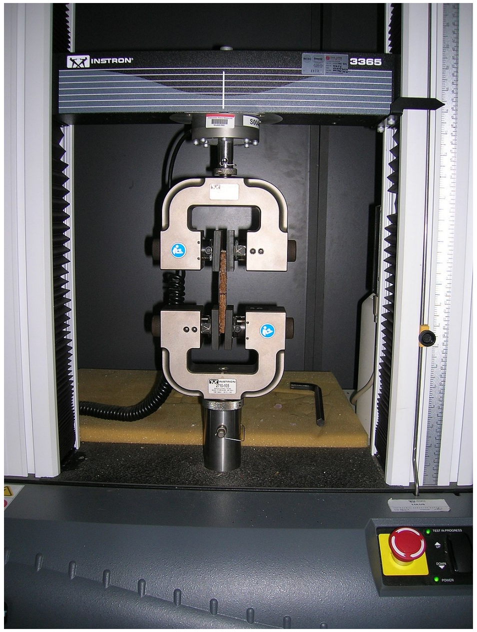
Lawrence Livermore National Laboratory

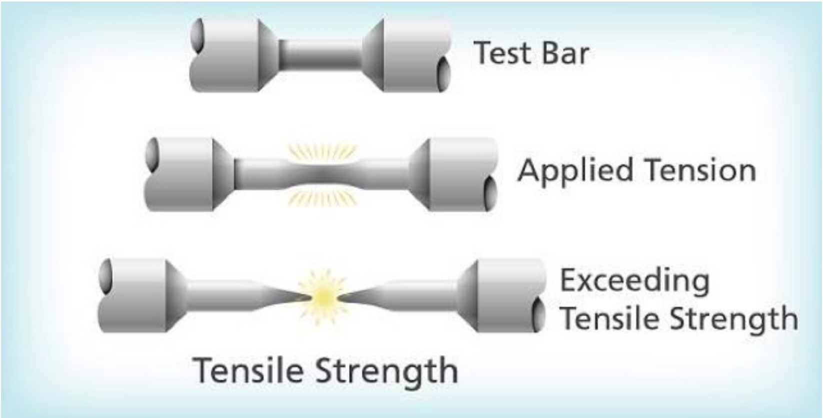
Goal: Increase Strength of TATB

- The material we are interested in is a powder called TATB.
- It must be combined with a **binder** into a solid pellet.
- Both the binder and TATB participate in forming the strength of the pellet.
- There are many ways to manufacture TATB as well as many choices of binder.
- What combination exhibits the most strength?

Tensile Strength Testing

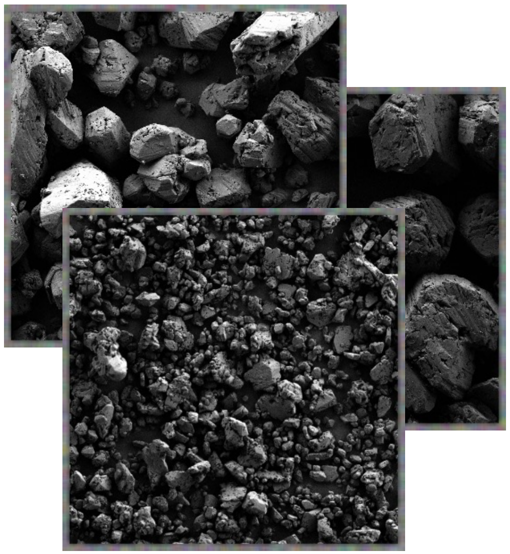


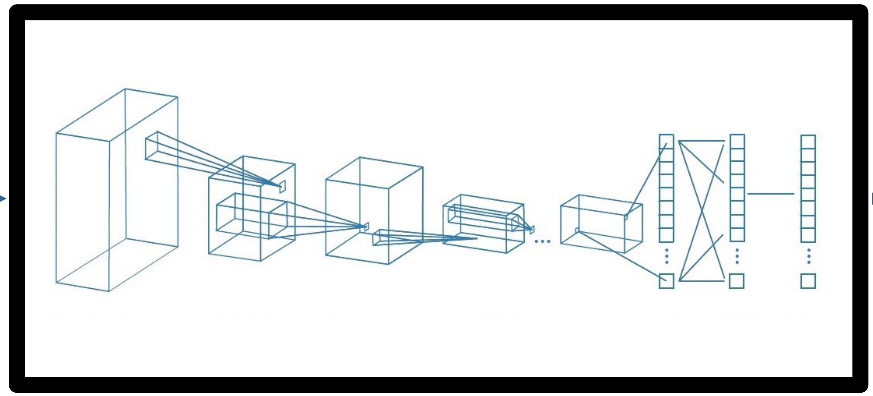


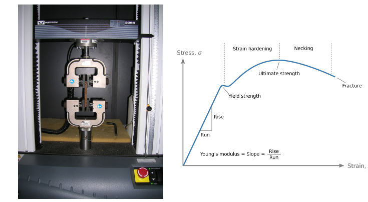


- Critical Peak Stress (aka Ultimate Strength): The point at which a material will not resume its original shape when stretched.
- This is an expensive test for this material.

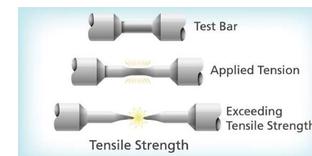
Initial Problem: Predict critical peak stress (CPS) of TATB by looking at SEM images







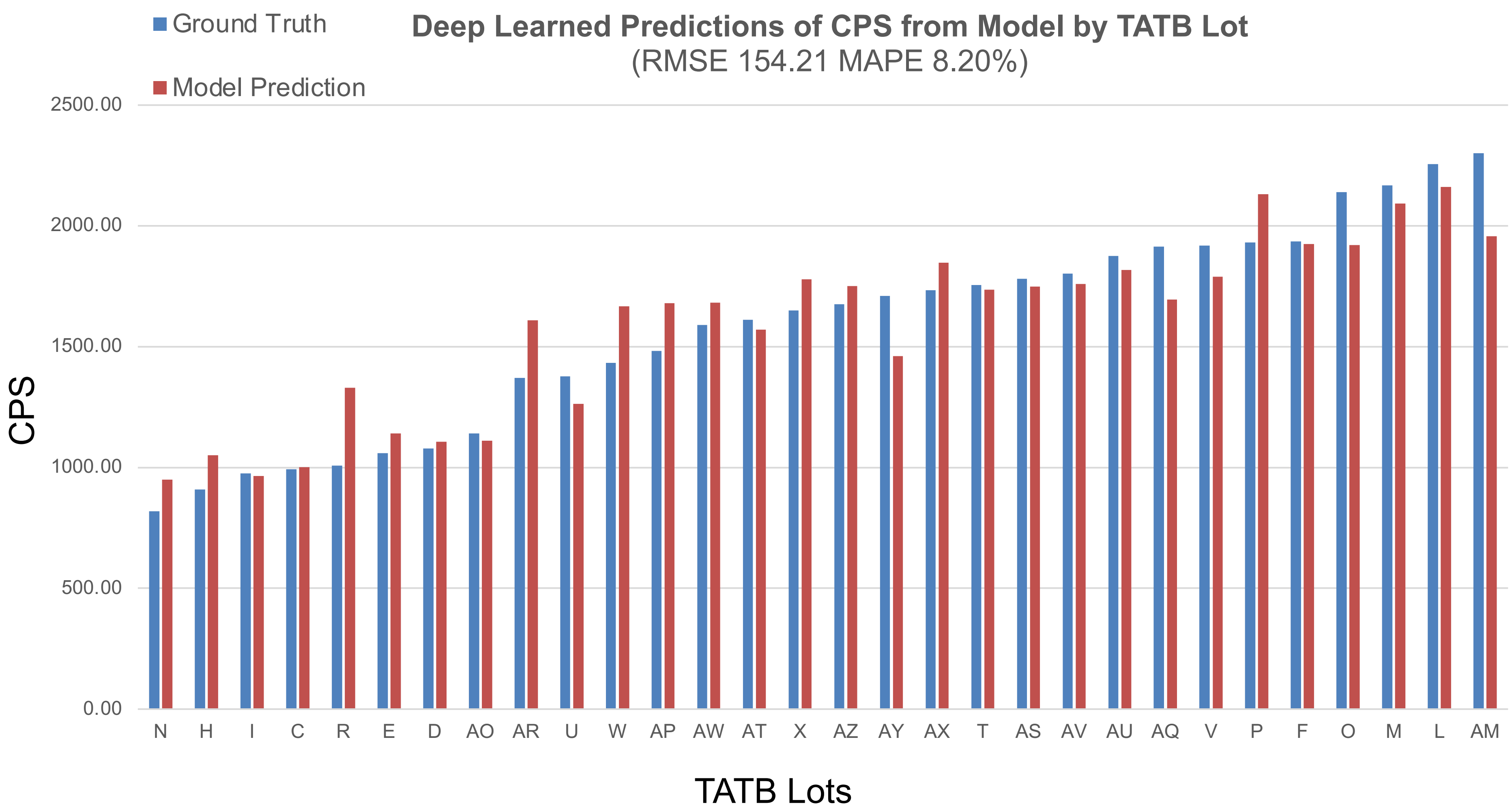
CPS



- By looking at a scanning electron microscope image of TATB crystals, can we predict what the stress-strain tested CPS (aka Ultimate Strength) will be when we have pressed it into a solid cylinder?

- From our paper: *Predicting compressive strength of consolidated molecular solids using computer vision and deep learning,* Materials and Design 2020
  – https://www.sciencedirect.com/science/article/pii/S0264127520300745

Results with our deep learning approach

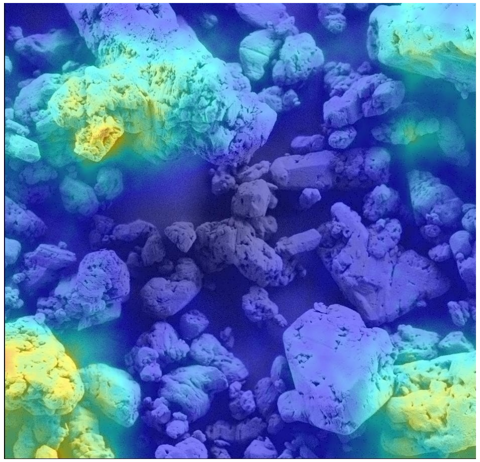


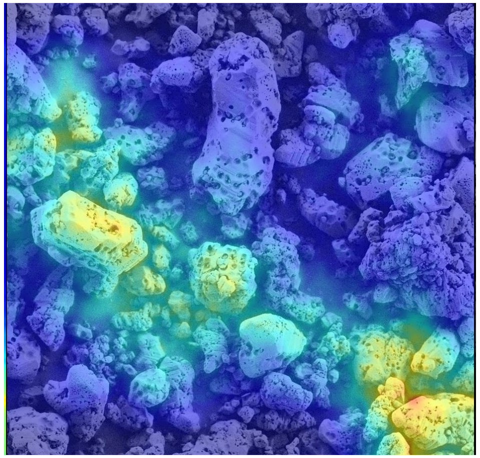
What is the network keying off of to make its predictions?

- Is the network making some sort of rational judgement about the material or is it possibly just lucky?
  - Can we expect it to continue to generalize and to what extent can we expect this method to work on other materials?
  - Do important visual features for the network correspond to useful physical properties?
  - Is the network using some feature we might not have thought of and can that give us new insight into material strength?

- Need an interpretable explainable AI (XAI) solution.
  - Most XAI methods are just saliency maps. Can we extract something *easier to interpret* than *where* did the network look?

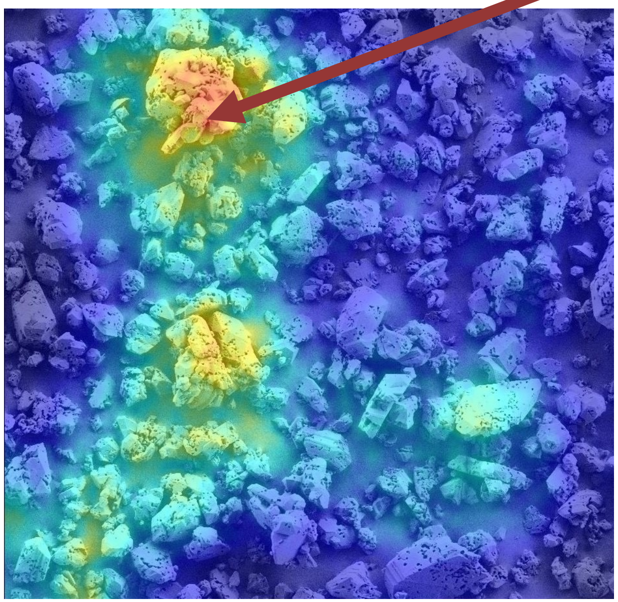
XAI Saliency Maps





Idea: Correlate describable textures to network decisions.

TATB SEM Image with Saliency Map Overlapped

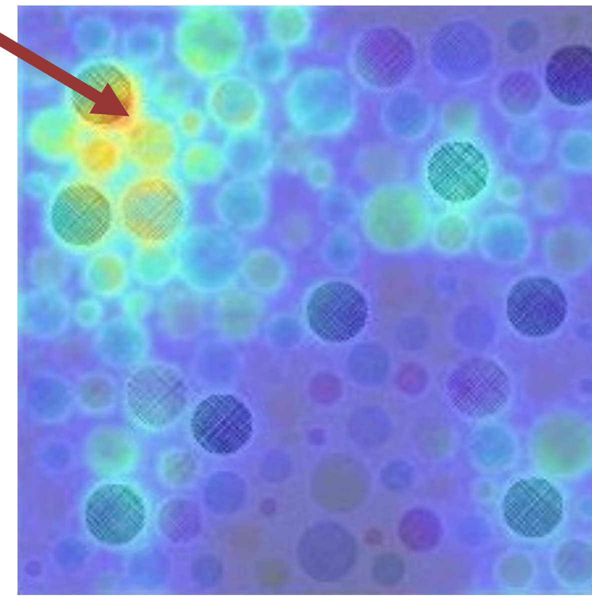


Extract Features at most salient locations from ***same*** model trained on SEM data.

Which texture features are the most like different SEM features?

Example: What kind of texture might correlate with a low CPS?

A Texture Image with Saliency Map Overlapped



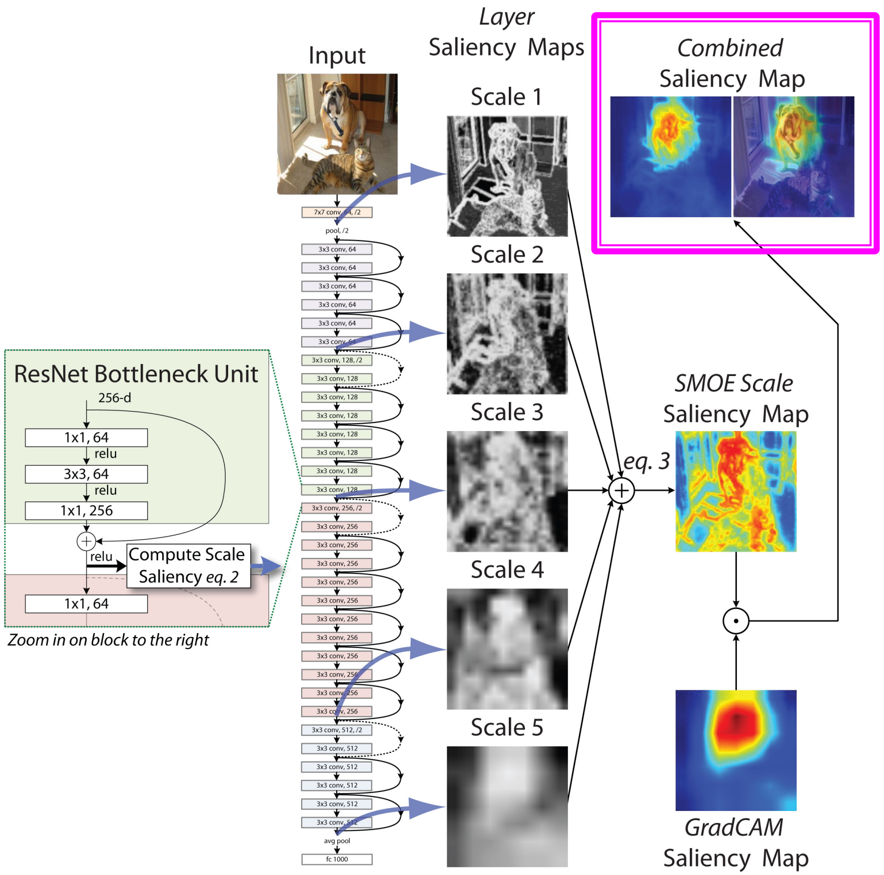
- We know *where* the network looks, but not *what* it finds interesting.
  - Features on SEM crystals are rather abstract to human observers.

- Can we correlate salient feature vectors from a SEM trained network between **texture images** and **SEM images** to tell us ***what*** the network is looking at?
  - This requires we use the same SEM trained network on both SEM and texture images to extract features.

Describable textures dataset

- Developed by VGG at Oxford, released in 2014
- 5640 images in 47 texture categories
- Example textures:
  - Bubbly
  - Honeycombed
  - Porous
  - Striped
- https://www.robots.ox.ac.uk/~vgg/data/dtd/

What is the image like to the network where it is most salient?



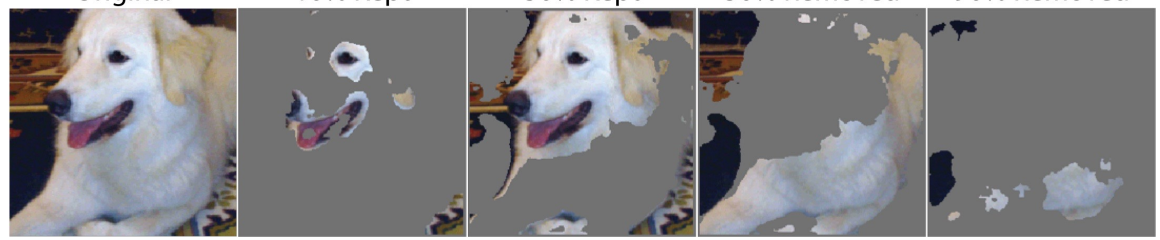
Formation of *SMOE Scale* map which will be element-wise multiplied by a *GradCAM* map.

- Use **FastCAM** to get most salient location.
  — https://github.com/LLNL/fastcam/
  — Efficient Saliency Maps for Explainable AI: https://arxiv.org/abs/1911.11293
  — Produces a saliency map of the parts of the image most important to the network’s decision.
  — Combines **SMOE Scale** a measures of layer activation variance activation with **GradCAM.**
  — Much faster than most other methods and more accurate.

Quantitative Results on FastCAM using ROAR/KAR

- If you mask out the most salient regions, performance should drop more for a better method when you train on the masked images.

- Conversely, if you mask out the least salient regions, performance should drop less when you retrain.



How faithful are the least salient locations?

How faithful are the most salient locations?

Combined Score Higher is Better

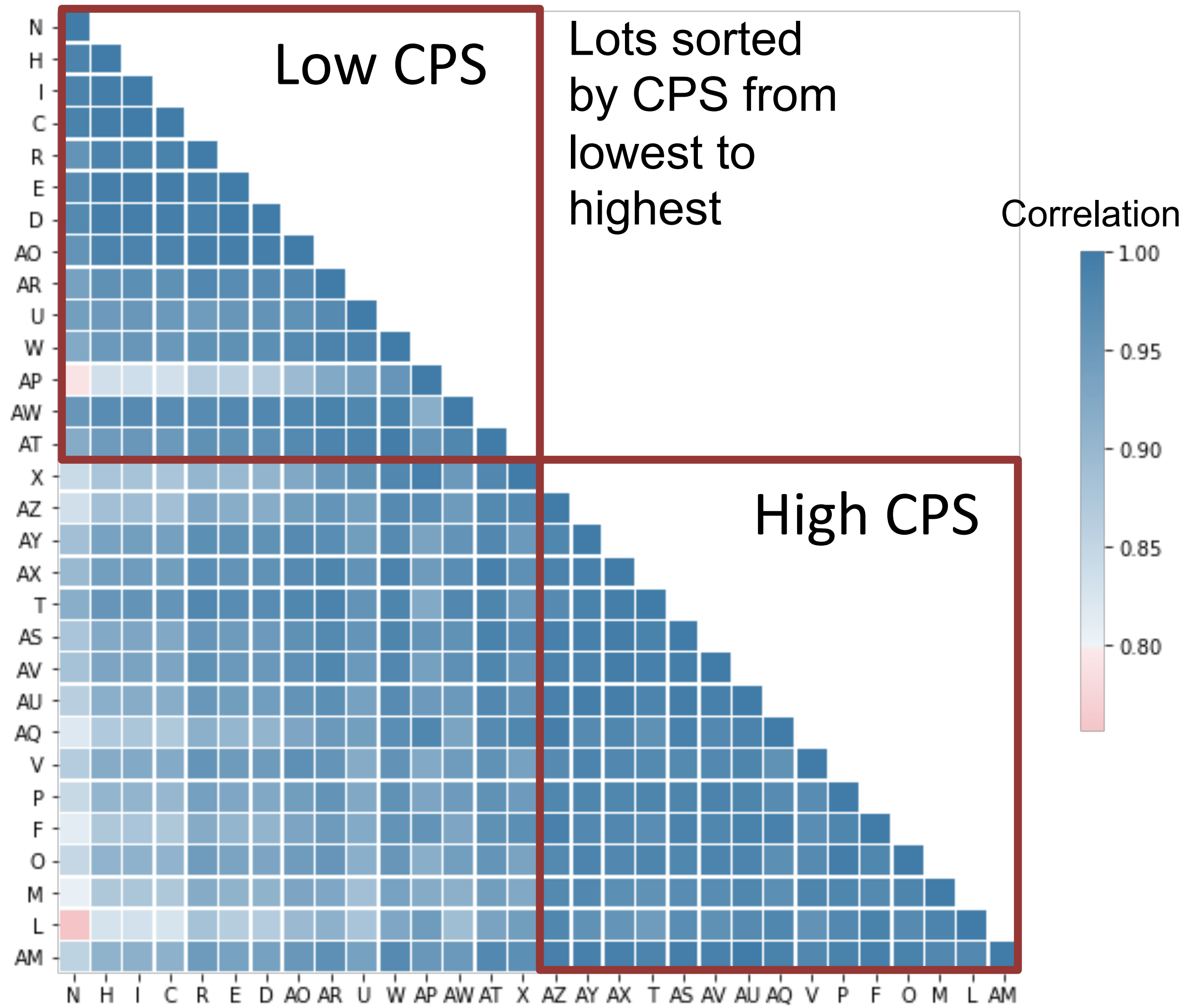
	Method	KAR	ROAR	COMBINED	Speed	Resolution
Existing Methods	Integrated Grad *Sundararajan 2017*	3.62	-3.58	0.03	Slow	Fine
	Gradient *Simonyan 2014*	3.57	-3.54	0.04	Medium	Fine
	Guided Backprop *Springenberg 2015*	3.60	-3.57	0.04	Medium	Fine
	Full Grad *Srinivas 2019*	3.66	-2.32	1.34	Medium	Fine
	Grad-CAM++ *Chattopadhyay 2018*	3.64	-2.27	1.37	Fast	Coarse
	XGrad-CAM *Fu 2020*	3.66	-2.27	1.38	Fast	Coarse
	Grad-CAM *Selvaraju 2017*	3.67	-2.27	1.40	Fast	Coarse
SMOE Scale Alone	SMOE Scale + Layer Weights [1,1,1,1,1]	3.62	-2.46	1.15	Fast	Fine
	SMOE Scale + Layer Weights [1,2,3,4,5]	3.62	-2.34	1.28	Fast	Fine
	SMOE Scale + Prior Layer Weights	3.61	-2.31	1.30	Fast	Fine
Combined Methods	Integrated Grad *-w-* SmoothGrad Sq. *Smilkov 2017*	3.56	-2.68	0.88	Slowest	Fine
	Guided Backprop *-w-* SmoothGrad Sq. *Smilkov 2017*	3.49	-2.33	1.16	Slow	Fine
	Gradient *-w-* SmoothGrad Sq. *Smilkov 2017*	3.52	-2.12	1.41	Slow	Fine
	SMOE Scale + Prior Wts. *-w-* Full Grad	3.68	-2.28	1.40	Medium	Fine
	SMOE Scale + Prior Wts. *-w-* Grad-CAM++	3.66	-2.22	1.44	Fast	Fine
	SMOE Scale + Prior Wts. *-w-* XGrad-CAM	3.68	-2.23	1.45	Fast	Fine
	SMOE Scale + Prior Wts. *-w-* Grad-CAM	3.67	-2.23	1.44	Fast	Fine
	Same as above Gamma CDF Normalizer	3.68	-2.23	1.45	Fast	Fine
	Same as above Layer Weights [1,1,1,1,1]	3.69	-2.24	1.45	Fast	Fine
	Same as above All Bottleneck Layers	3.68	-2.23	1.45	Fast	Fine

Method used in our Illustrations.

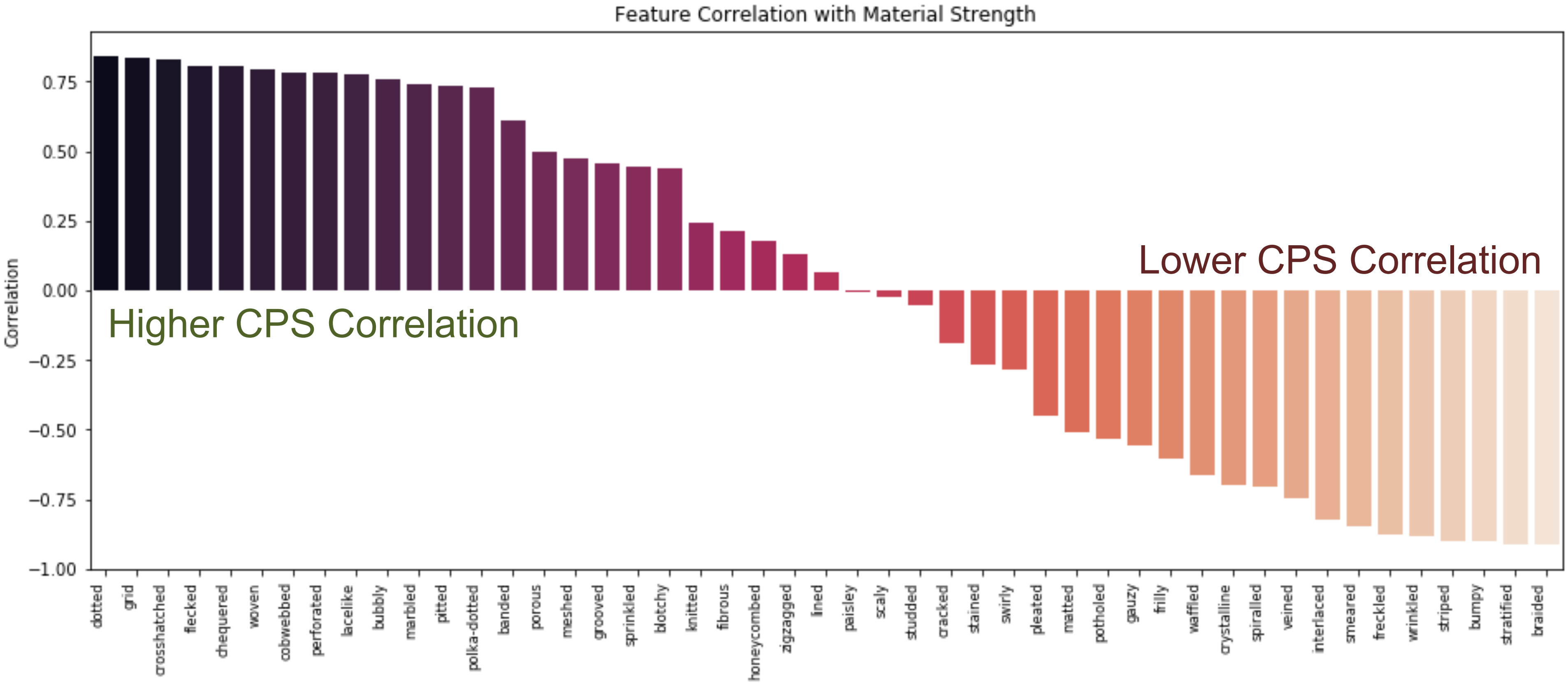
Our solution is approx. 1500 times faster!

Batches look more like other batches with similar CPS.

Without looking at texture yet: we can see that salient feature vectors from different lots look more like lots with similar CPS.

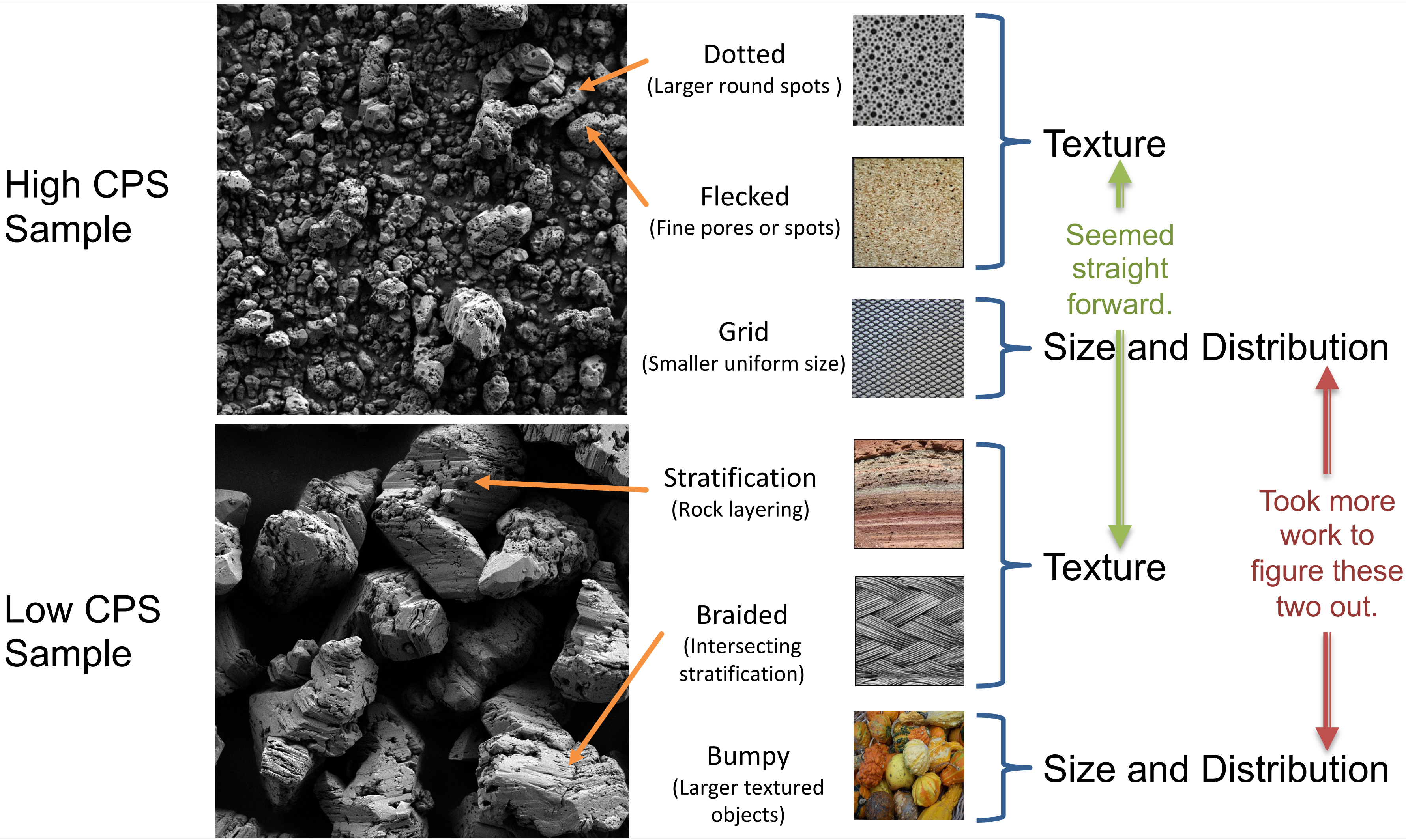


Correlation between textures and CPS



Texture Categories

What textures tell us the network is looking for



Conclusion

- *A priori* particle size (grid/bumpy) and porosity (dotted/Flecked) we suspected of playing a roll in CPS. However, stratification/braiding was something new we uncovered.
  - Note that we cannot eliminate confirmation bias as a factor.
- Did not see signs of some a priori suspected features playing a roll. These include facet, dispersity and surface area.
  - It's harder to exclude suspected visual features by this method.

Thank you

mundhenk1@llnl.gov

Lawrence Livermore National Laboratory