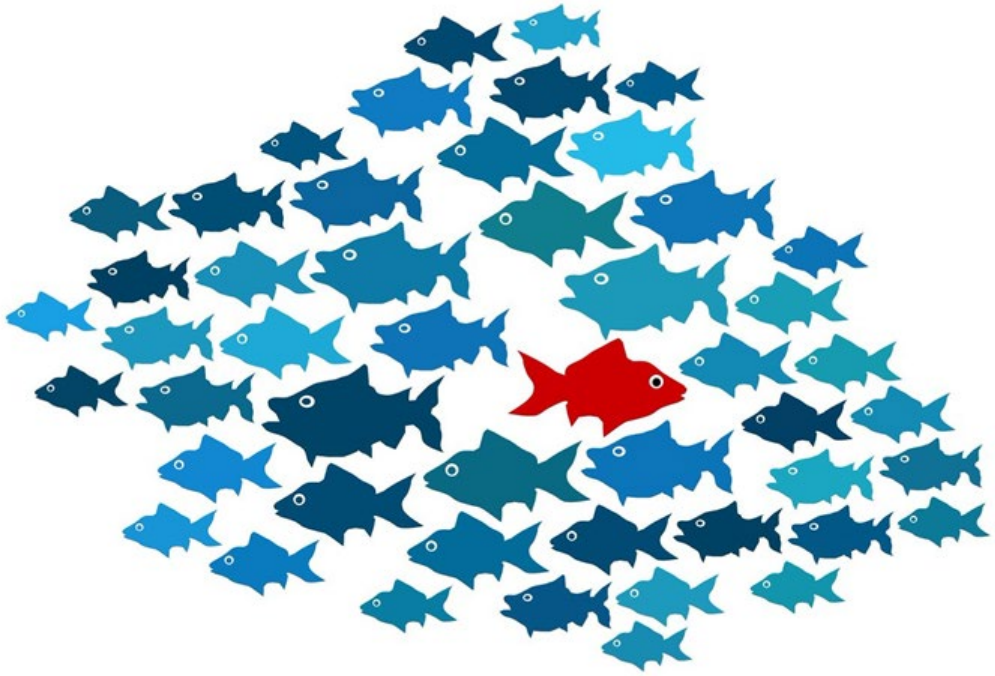
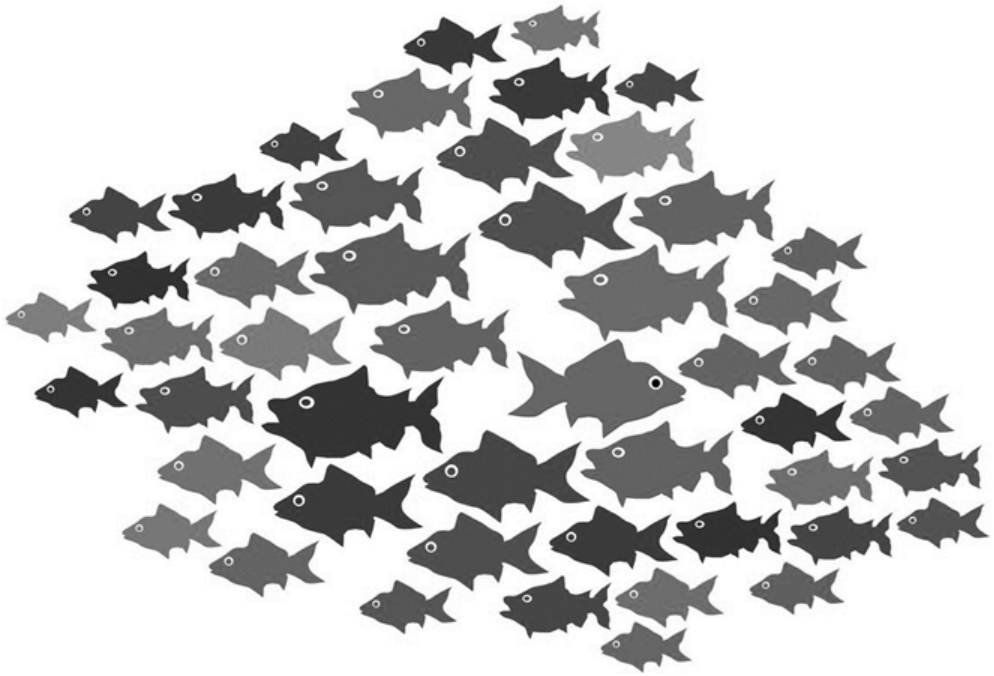
August 30, 2022| AKI FUJIMURA, CEO, D2S, Inc.

Going Deep: Without the Right Data, Deep Learning Stops as a Prototype

Humans Can Find an Anomaly Easily

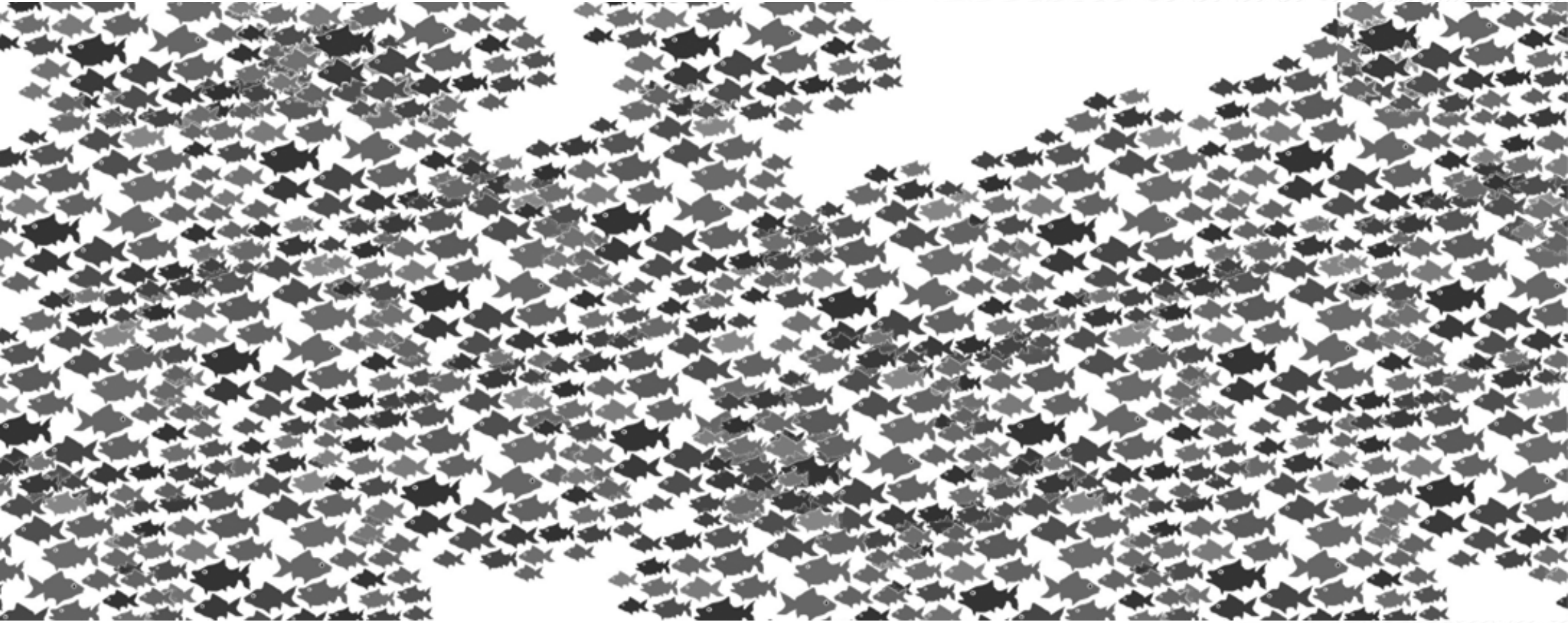


How About Now?

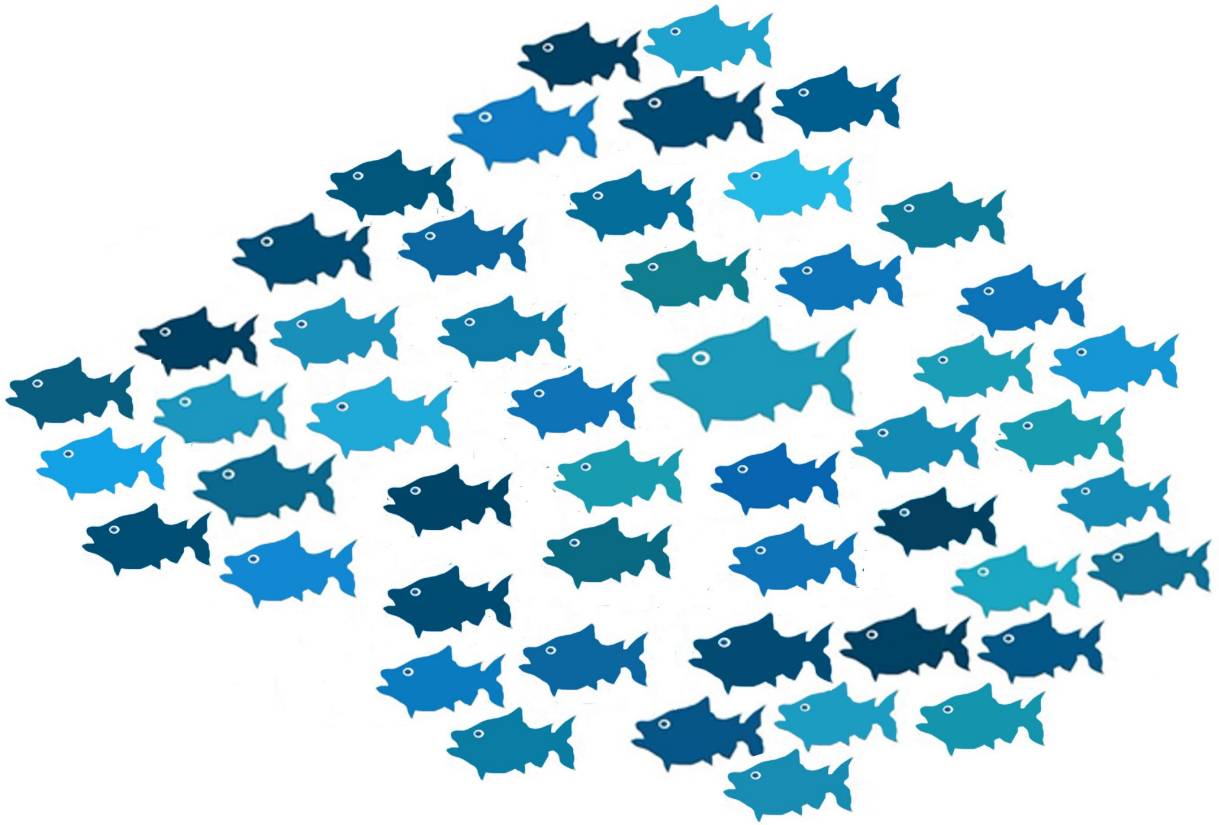


And Now?

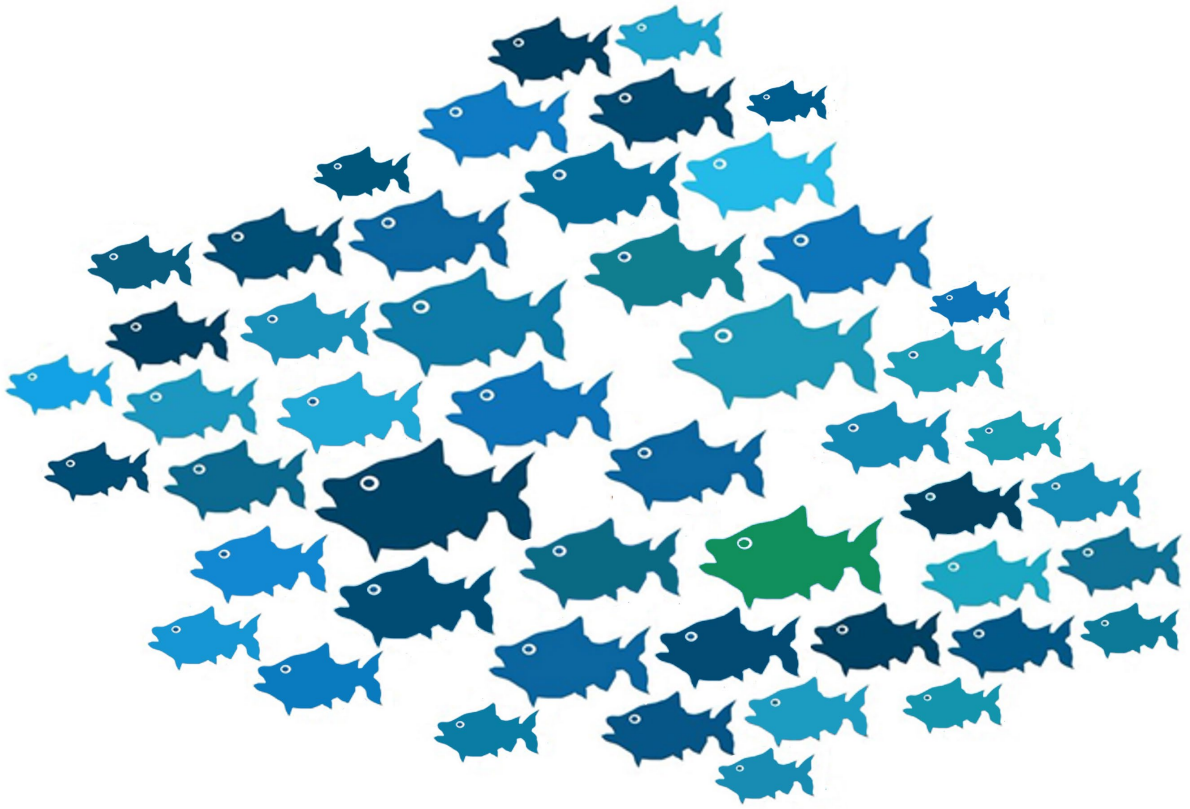
Machines are better at Tedious and Error-Prone Processes



It’s Amazing How A Human Brain Does This

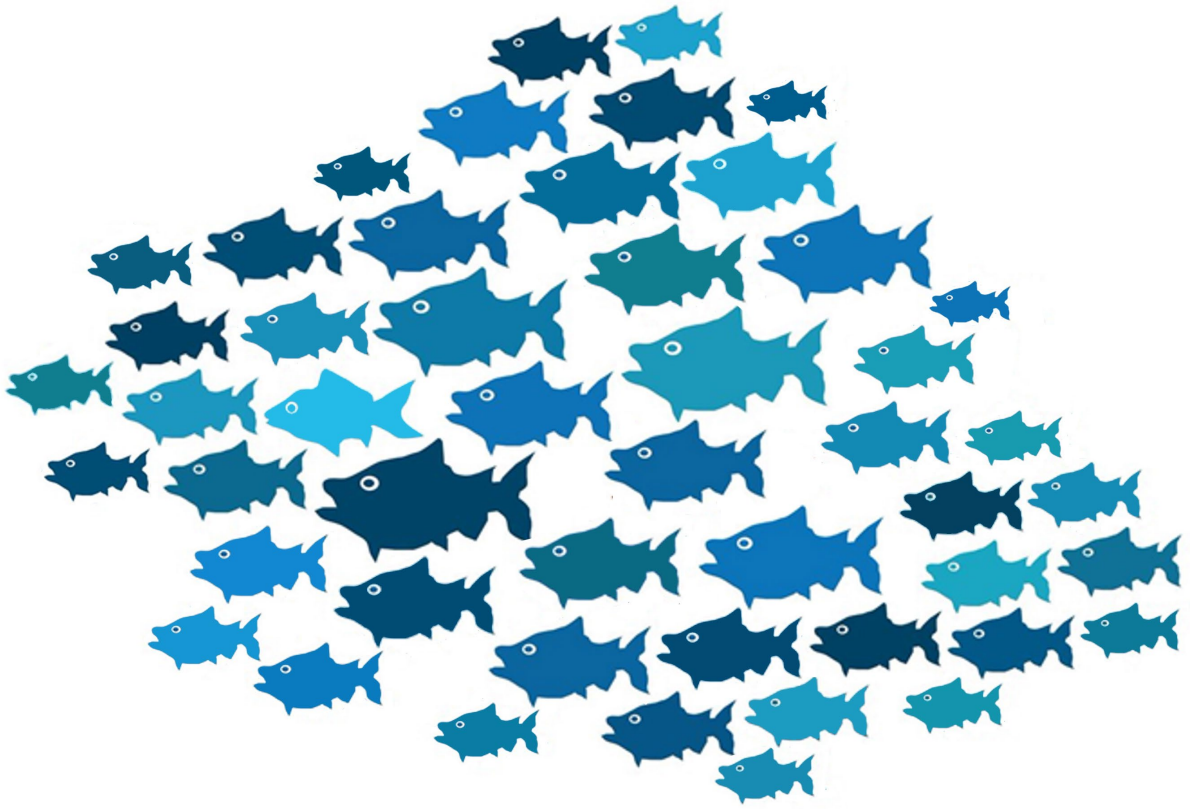


It’s Amazing How A Human Brain Does This



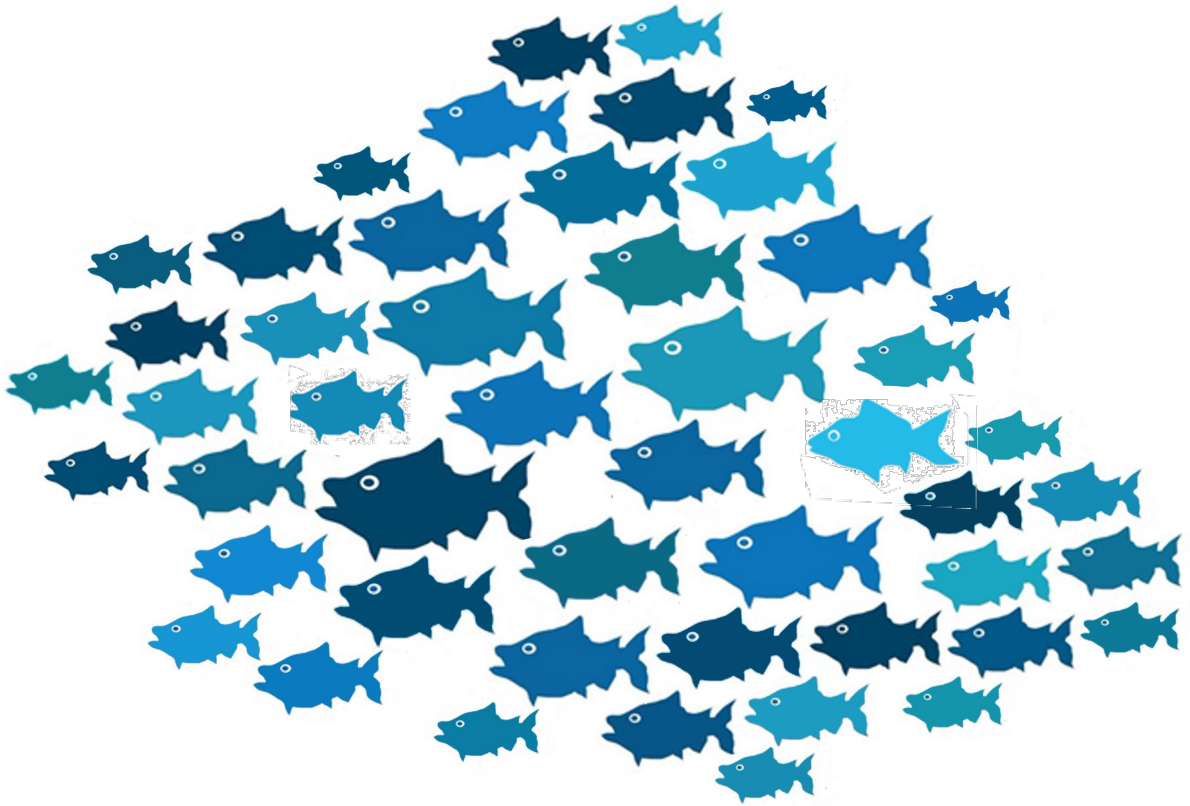
It’s Amazing How A Human Brain Does This

A little harder?

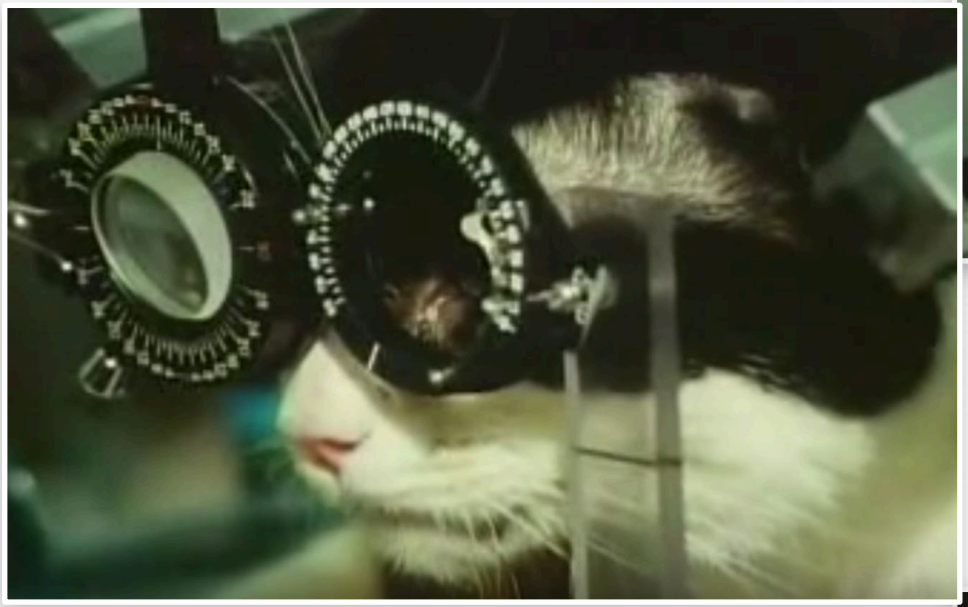


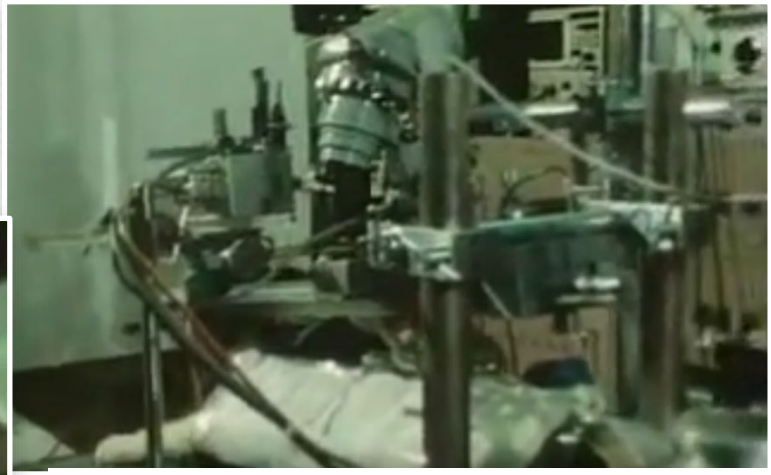
It’s Amazing How A Human Brain Does This

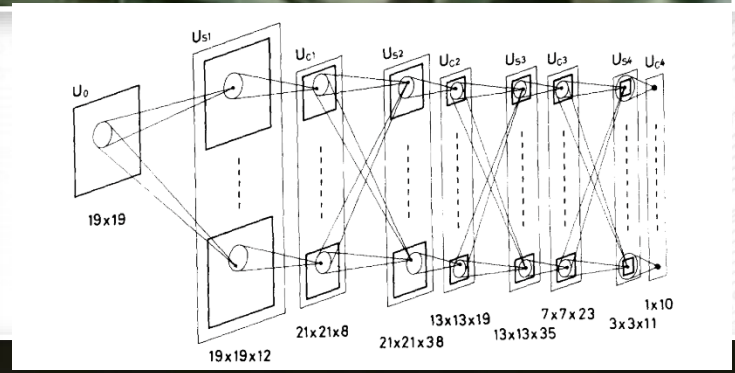
Even for a hard one, once you get it, it’s easy

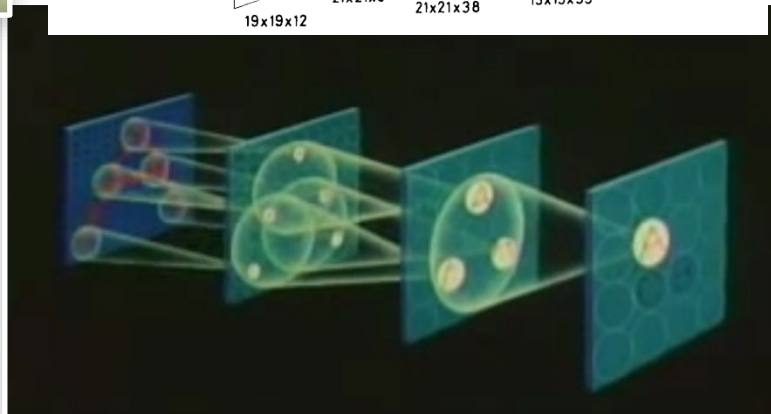


Actually, a Cat can do This, too









Dr. Kunihiko Fukushima
Then of NHK Science and Technology Research Laboratories
Now at http://flsi.cird.or.jp

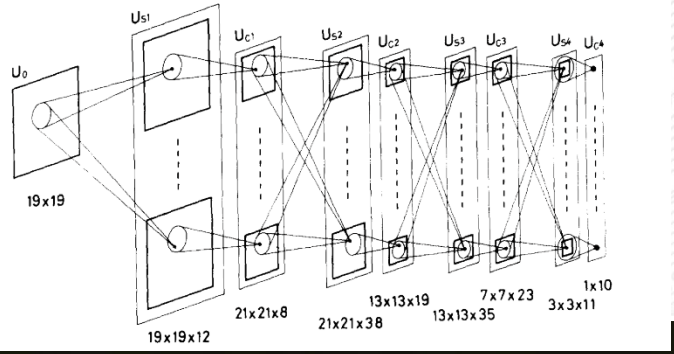
Biological Cybernetics
September 1975, Volume 20, Issue 3–4, pp 121–136 | Cite as

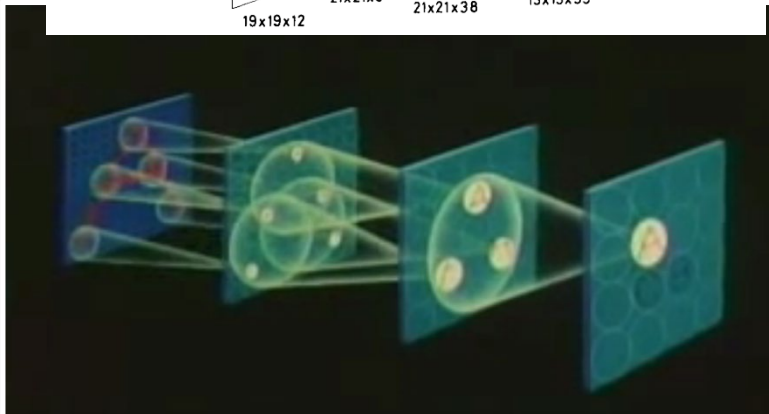
Cognitron: A self-organizing multilayered neural network

Authors | Authors and affiliations

Kunihiko Fukushima

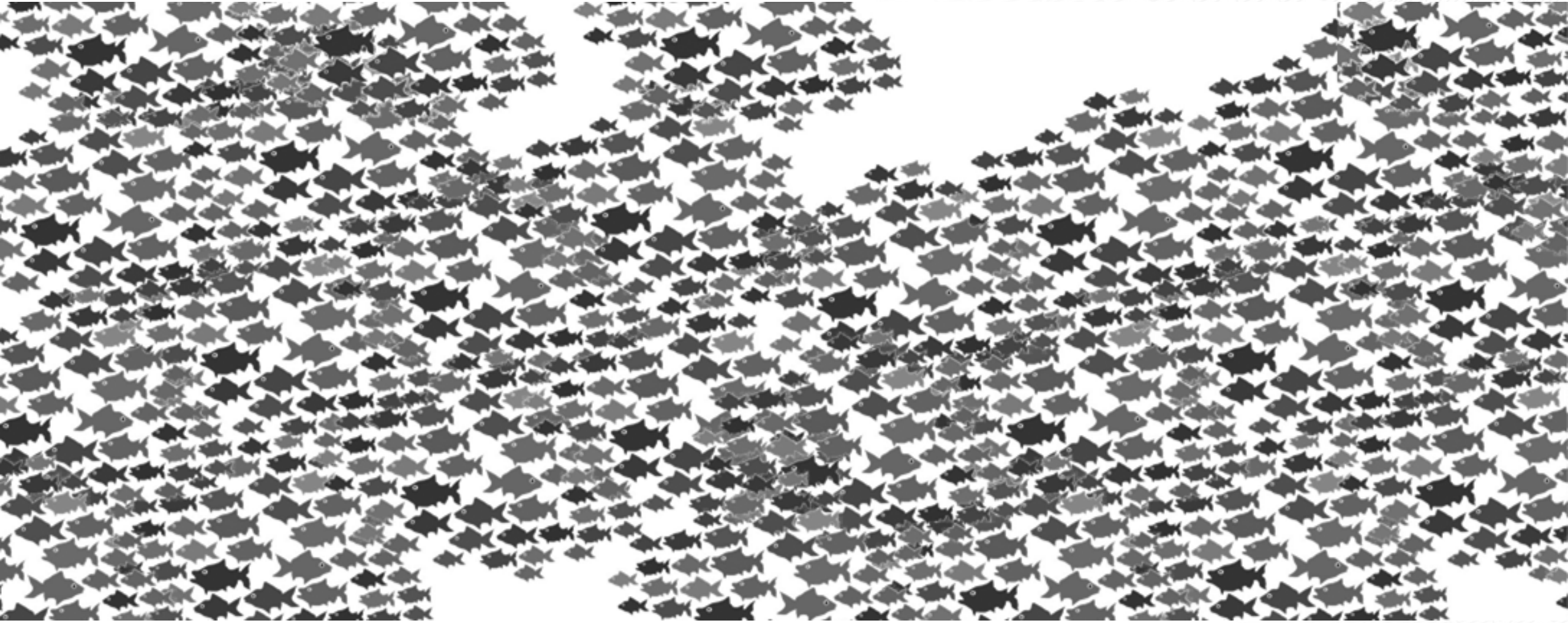
A new hypothesis for the organization of synapses between neurons is proposed: “The synapse from neuron x to neuron y is reinforced when x fires provided that no neuron in the vicinity of y is firing stronger than y”. By introducing this hypothesis, a new algorithm with which a multilayered neural network is effectively organized can be deduced. A self-organizing multilayered neural network, which is named “cognitron”, is constructed following this algorithm, and is simulated on a digital computer. Unlike the organization of a usual brain models such as a three-layered perceptron, the self-organization of a cognitron progresses favorably without having a “teacher” which instructs in all particulars how the individual cells respond. After repetitive presentations of several stimulus patterns, the cognitron is self-organized in such a way that the receptive fields of the cells become relatively larger in a deeper layer. Each cell in the final layer integrates the information from whole parts of the first layer and selectively responds to a specific stimulus pattern or a feature.





What's it Take to Program Anomaly Detection?

Even if Only to Solve for These Fish Pictures, Very Hard

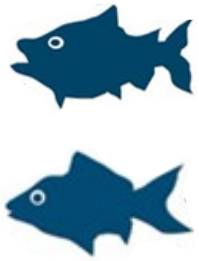


Without DL, We Would Write Code to "Classify"

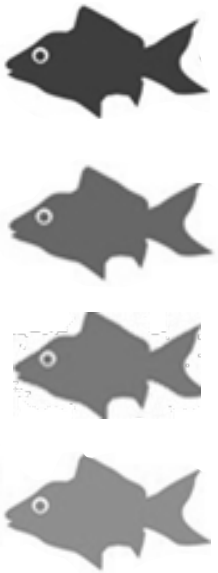
Compare and Count: Least Occurring is Anomaly

But that works only when you start with knowing all the fish

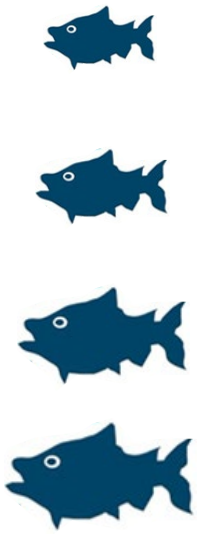
Two Shapes



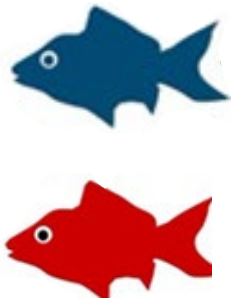
Four Shades



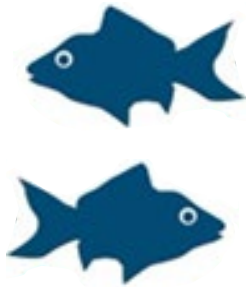
Four Sizes

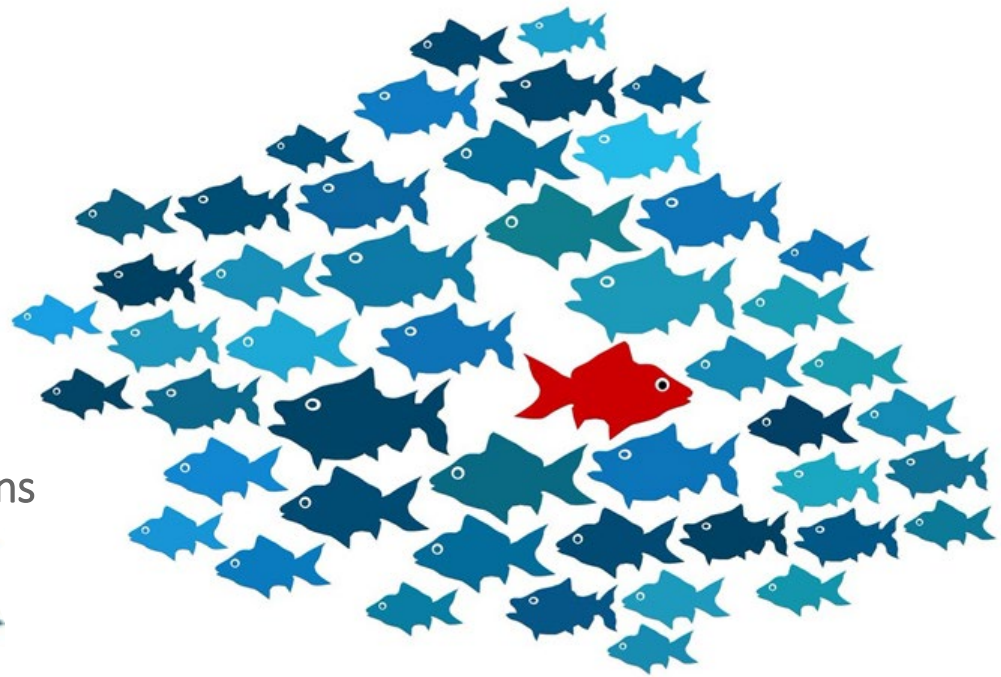


Two Colors



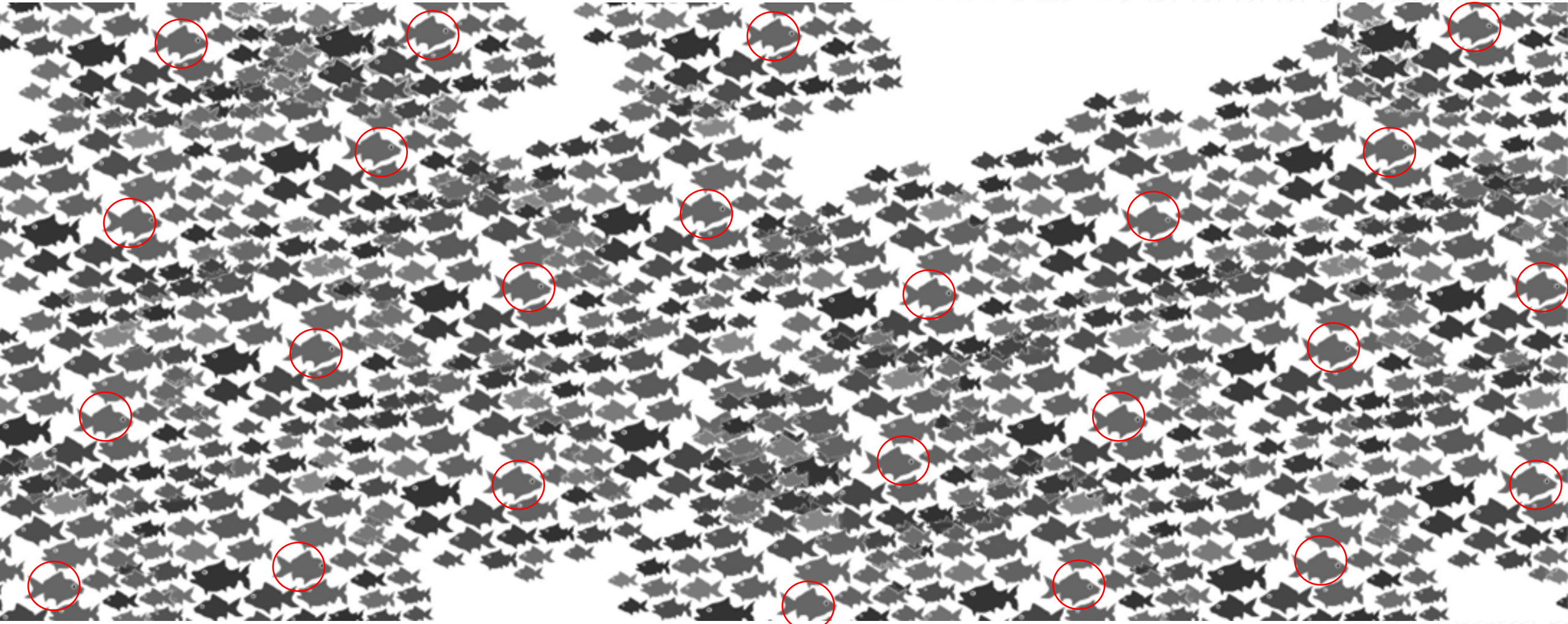
Two Directions





DL is Automatic Programming by Pattern Matching

Programmed by Data, so anomaly detection of birds is only data difference away



“Deep Learning can do all this by being nothing more than a fancy ‘grep’” – Steve Teig, CEO, Perceive

Isn’t the Anomaly That Some Fish Swim in Schools?

Human brain is more than just pattern matching

This is Normal for Sardines

Humans can reason that without learning; But DL can only learn that

GPU Computing Created Inflection Point for DL

A100 was announced May, 2020:
FP32 for PCIe : 19.5 TFLOPS

H100 announced March, 2022:
FP32 for PCIe : 48 TFLOPS

So, Mask Making Should Benefit from DL, too

There's been many papers

ASML

Canon

CENTER FOR DEEP LEARNING IN ELECTRONICS MANUFACTURING

D2S

Fraunhofer

HITACHI Inspire the Next

imec

MYCRONIC

NuFlare

SIEMENS

life.augmented

SYNOPSYS®

'TORAY' Innovation by Chemistry

TASMIT, Inc.

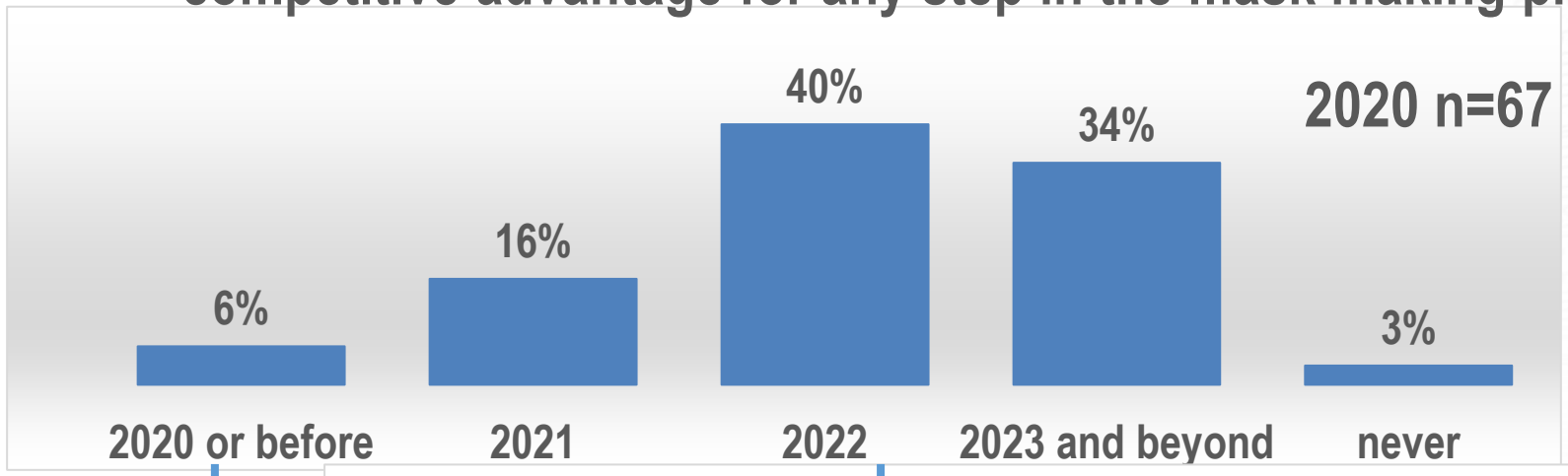
And yet…

D2S

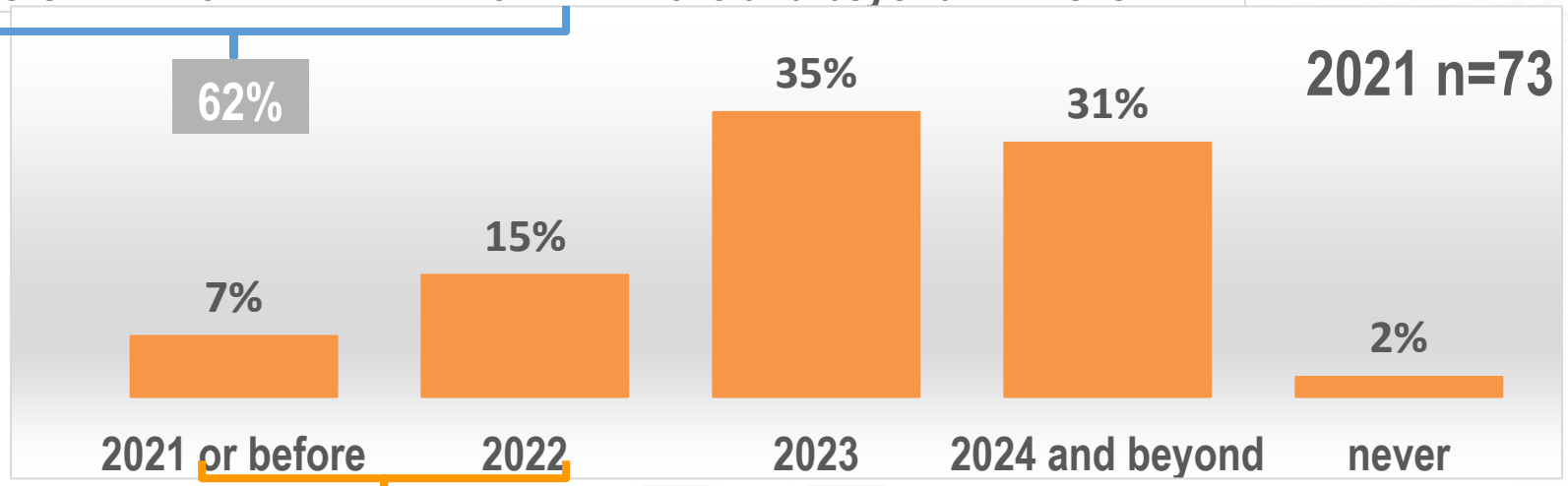
Deep Learning Predictions Shift to 2023 & Beyond

Only 22% say 2022 vs 62% in last year's survey

In the mask industry, when will capabilities based on deep learning become a competitive advantage for any step in the mask making process?



62%



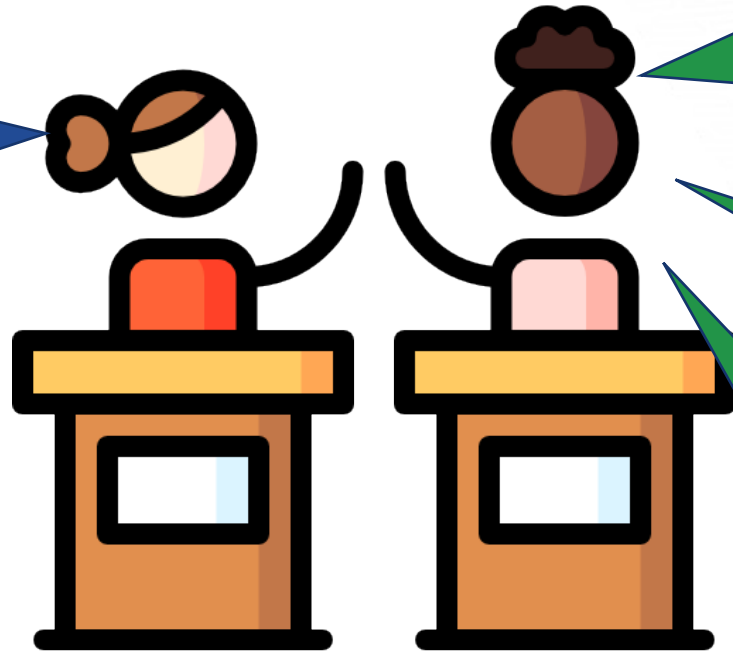
22%

Answer choices expanded in 2021 survey – see X axis in two charts above

19

So Why isn't DL Exploding in Mask Making?

We don't need DL. Mask shop is already designed and controlled to be automated. Not like driving a car where all sorts of conditions need to be anticipated



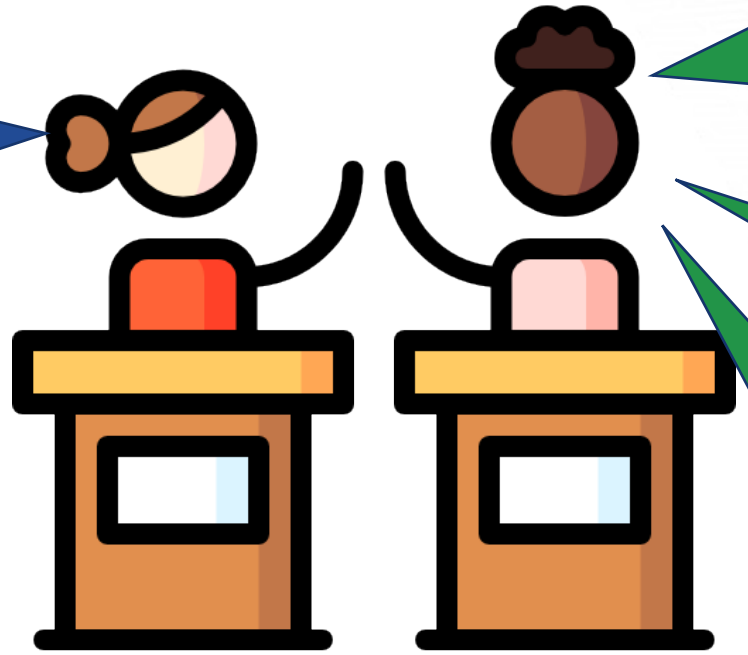
True that a mask shop is amazingly automated. But there's still need for inspection and repair, right? So not everything is fully automatic.

And as reliable as these machines are, they do fail and need maintenance over time.

Lots of what we do is software too, and they surely can use DL.

So Why isn't DL Exploding in Mask Making?

Even for software, mask effects are about physics , chemistry, and math. We can calculate the answers accurately and even faster than DL can guess.



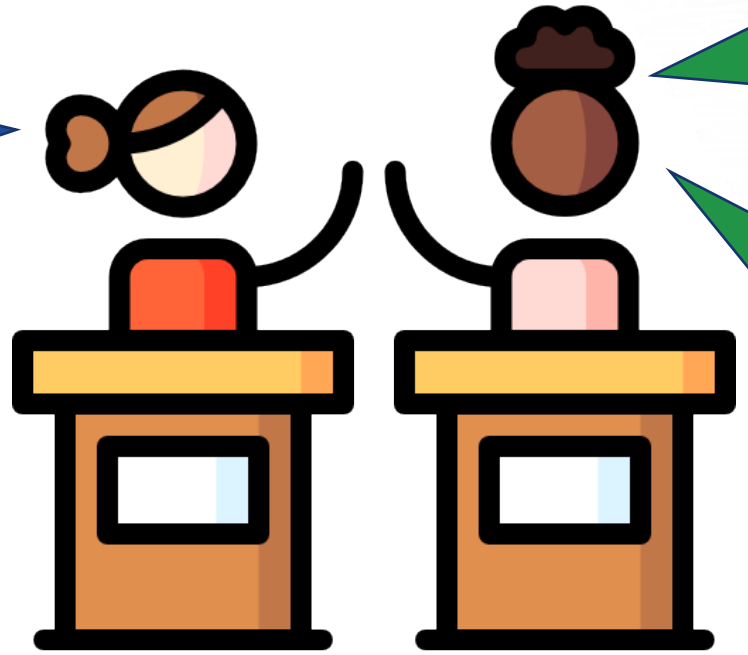
Many things are like that. An analytical computation that's fast like Gaussian convolution or FFT should just be computed.

But there are many effects like Variable Etch Bias where accurate calculation takes too long.

And computations that take iterative optimization can be accelerated by DL prediction.

So Why isn't DL Exploding in Mask Making?

But mask making needs too much precision. DL is a pattern-matching method, so accuracy is limited. Maybe 2-3nm 3 sigma for geometries?

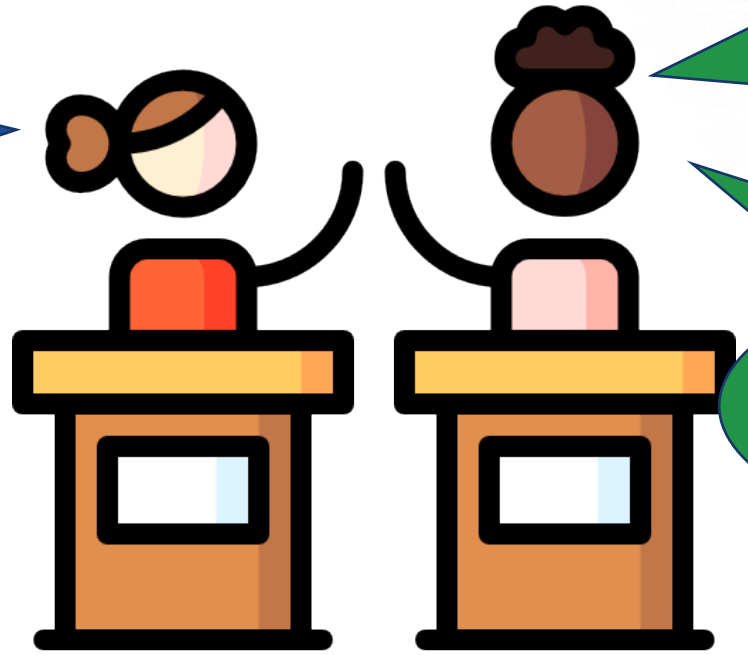


Right. Analytical computing is better than DL if it's fast enough. But DL is great for prediction of complex effects.

For example, Mask 3D effects are essential for ILT accuracy. Rigorous M3D takes too long. Since M3D is a secondary effect, "close enough" is accurate enough. It's a better speed-accuracy trade-off.

So Why isn’t DL Exploding in Mask Making?

Isn’t it also a problem though, that there’s no way to prove that DL is never going to make a big mistake?

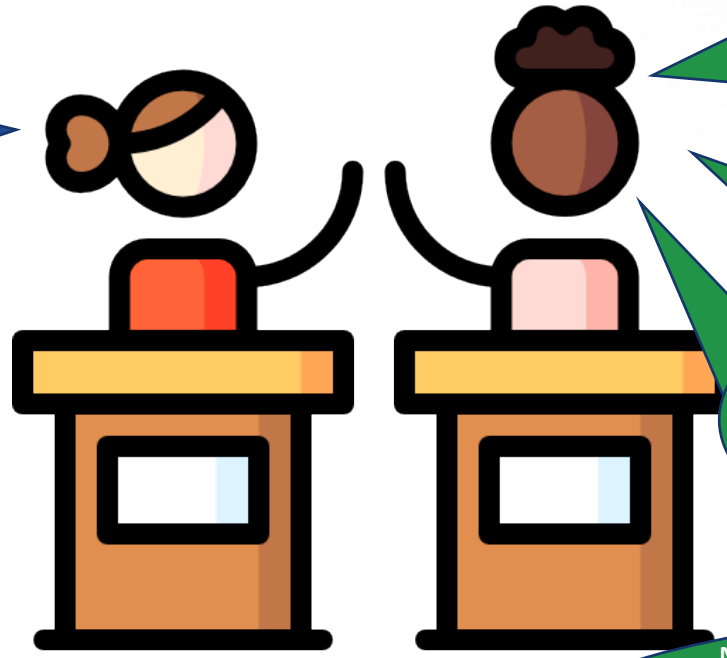


Well, that depends on the network architecture chosen. In general, convolution-based methods have predictable behavior.

Still, it’s an important point that constructive algorithms like ILT and MPC should only use DL for acceleration or full-chip model approximation.

So Why isn't DL Exploding in Mask Making?

Ok, so there are some uses for DL. But isn't it basically an overhyped shiny new toy that's not so shiny and new any more?



It's amazing what pattern matching alone can do. So the realization that this is possible was and still is exciting and worthy of hype.

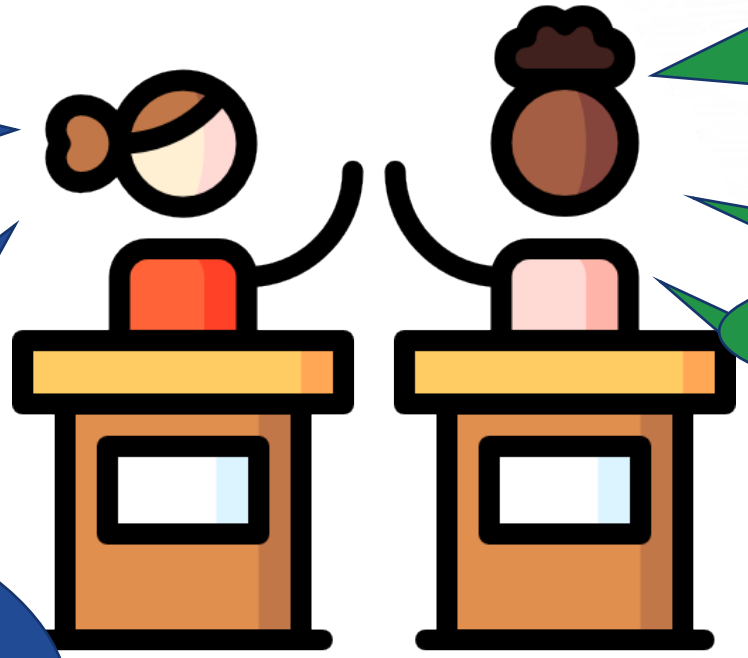
But you're right that it isn't general AI that can reason like humans can. DL doesn't reason, but has attention to detail with a tireless work ethic.

Much like gaming initially funded incredible growth in GPU computing, social media is continuing to fund incredible expansion in DL. Shiny new toys are coming out every day.

So Why isn't DL Exploding in Mask Making?

DL is for vendors to provide to us, right? They improve their products using DL and we get it and use it? DL is software, right?

Vendors would love to provide it, but DL is programmed by data. Customer data is confidential, so only you have the data.



Most amount of data wins in DL. So we need to collaborate.

We get the need to collaborate. But you're right that we can't give you the data. Most of it isn't even our data, it's our customer's or their vendor's. Does this mean that we have to re-train the DL network for each process?

Networks can be pre-trained and then only updated with data from a different process. But yes, we need to give you that ability.

So Why isn't DL Exploding in Mask Making?

We tried, but our DL didn't work. Maybe we didn't do it right, but our experience with DL is now tainted.

We've had great success with it. A number of DL applications are now in production use. But it took a lot of resources and time.

Yeah, data is the key. You have to invest in acquiring or generating data.

We were initially really excited by DL. We had an engineer show me a prototype she did by herself. So we made it an official project for her and tried to work with customers to get data. But ultimately we had to cancel the project. It just didn't work well enough for production use.

Prototypes being easy and cheap can mislead people. Production deployment needs commitment.

Protype: Easy; Production: Hard

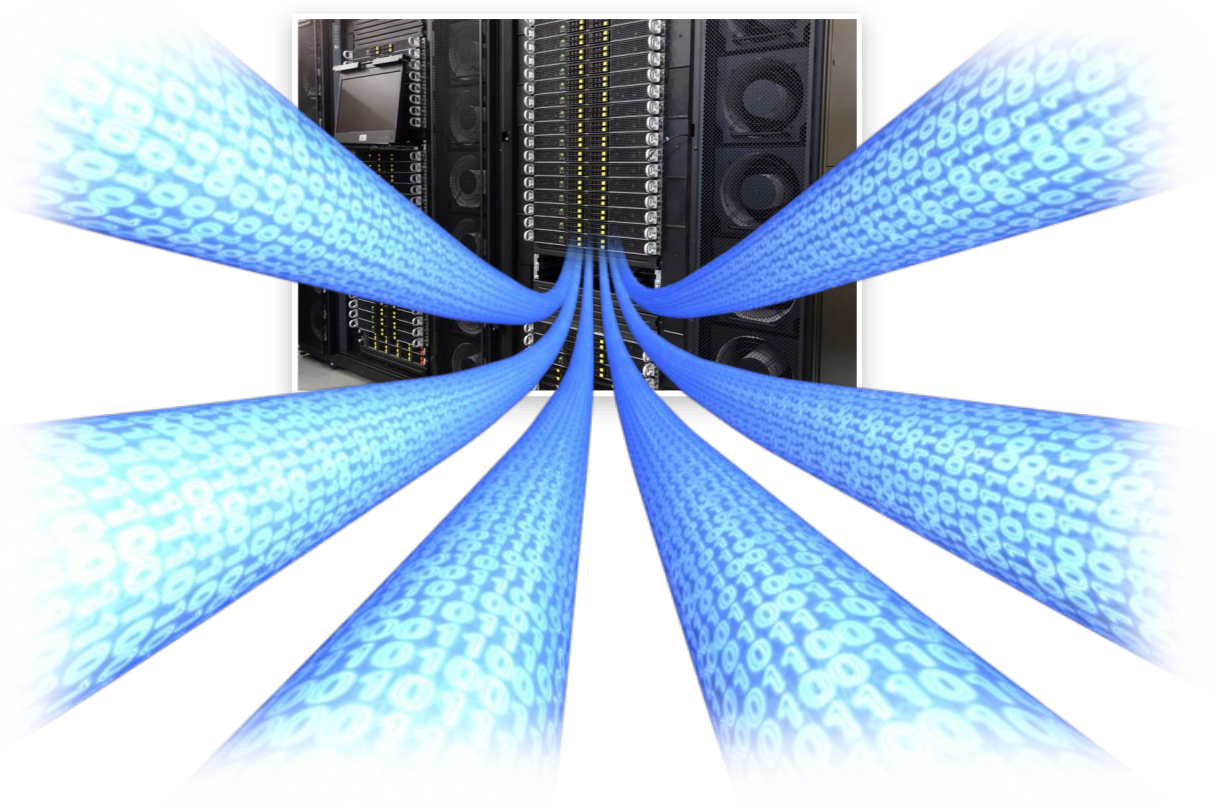
DL Application Prototype: Quick and Easy

DL Production-Quality Application: Much Harder!

No wonder there is some skepticism…

27

DL is “Programmed” with Data…Lots of Data



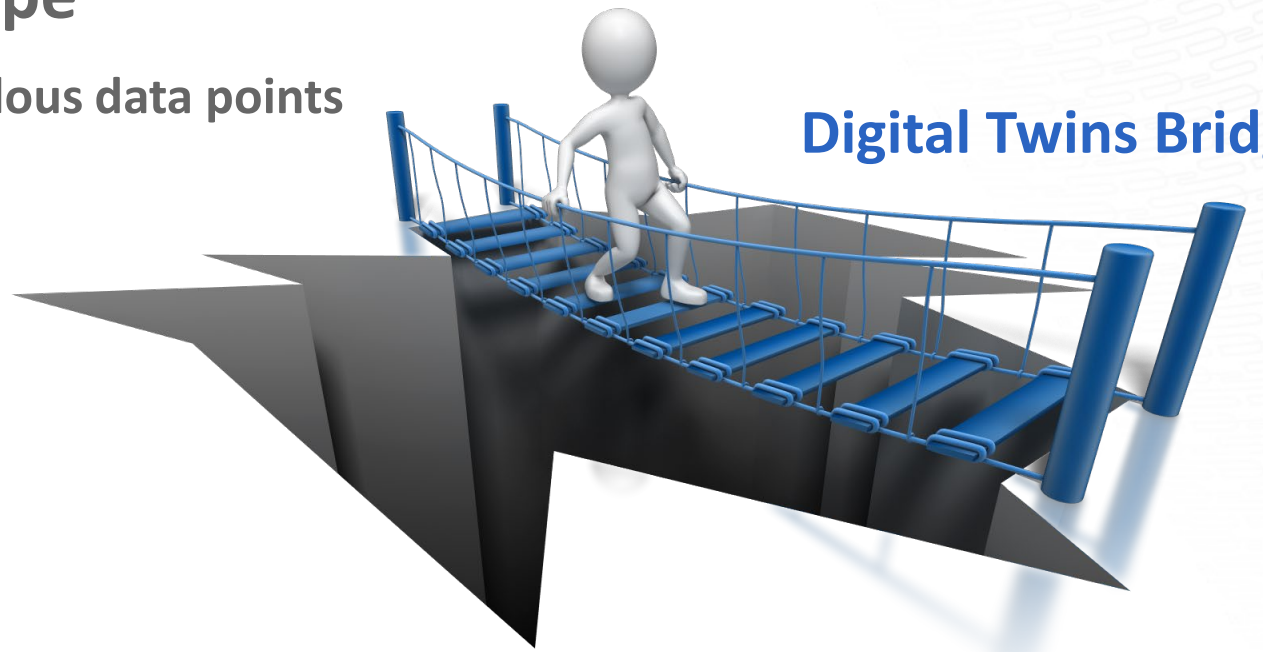
- DL is only as accurate as it’s trained to be = masses of data
- Data belongs to the customer
- Mask shops are good at not making too many samples of anomalous data

Simulation and digital twins are essential!

Digital Twins Generate Disambiguating Data for DL

Prototype

1000s of non-anomalous data points



Digital Twins Bridge the Data Gap

1,000,000s of data points covering all conditions

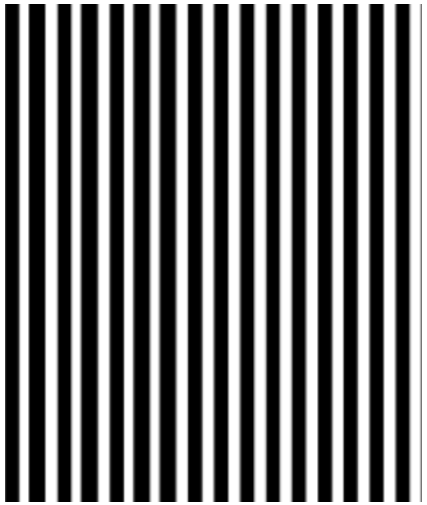
Production Quality

Digital Twins Are Custom Tools for DL Applications

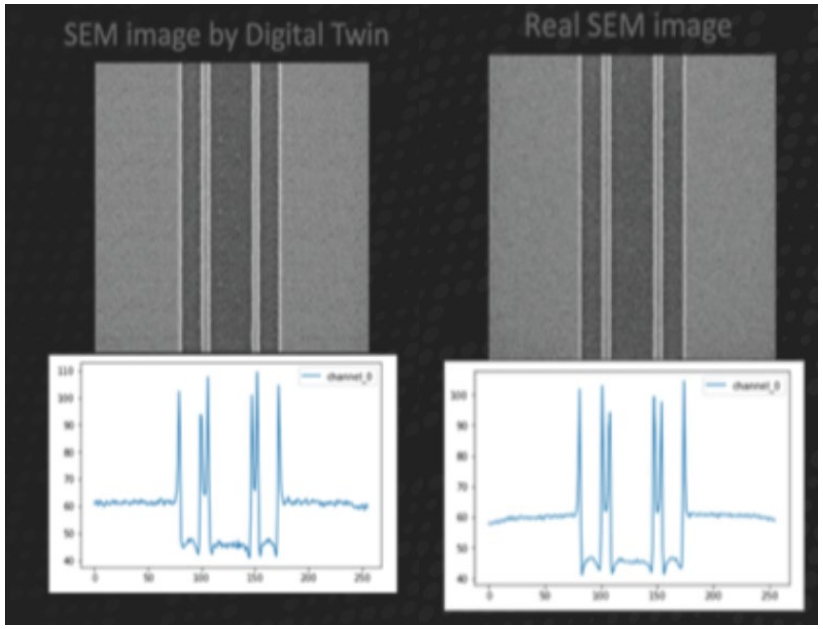
This Old House

SEM Digital Twins Help to Generate SEM Images

Generated SEM for training; real SEM to test



CAD

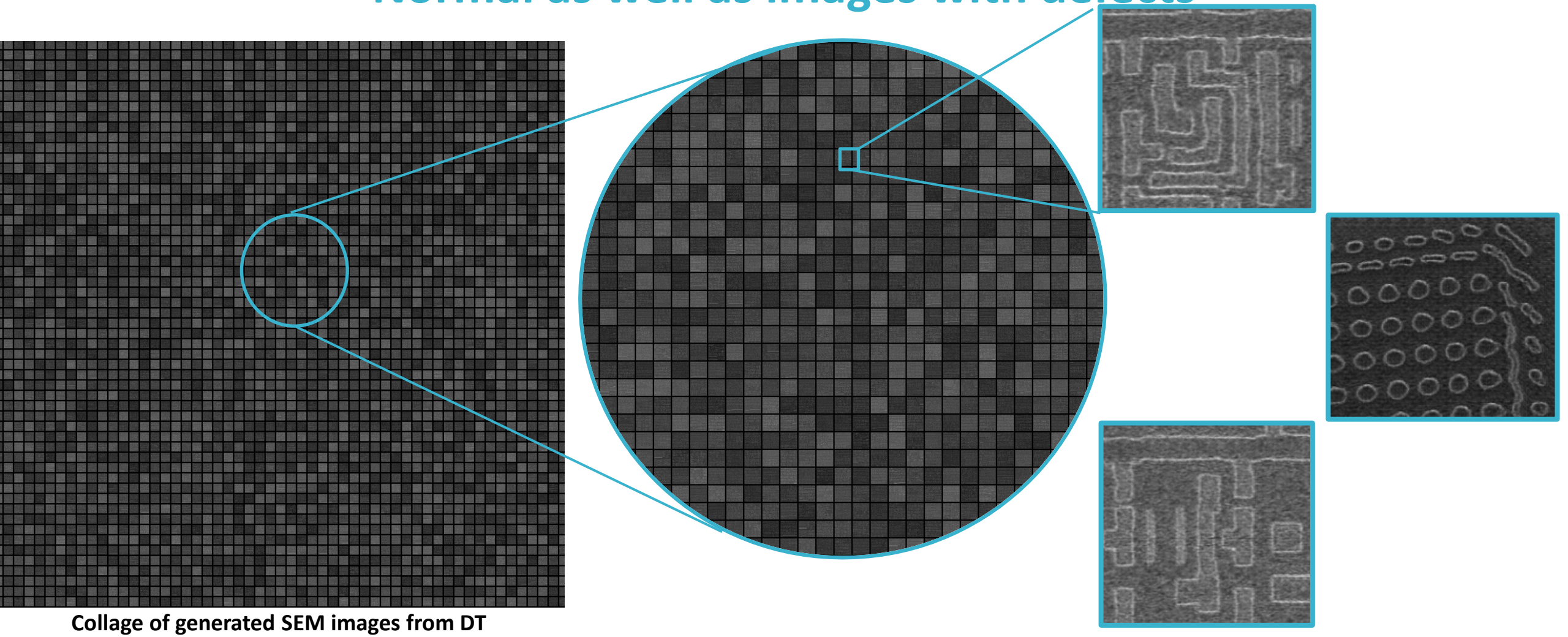


Digital Twins

Generated SEM

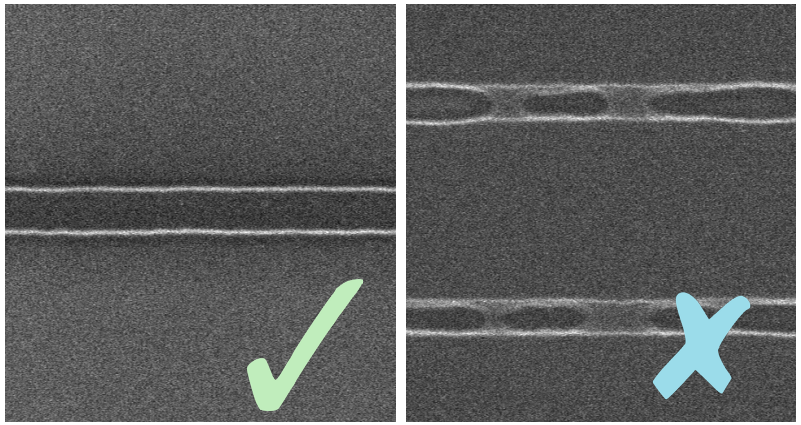
Millions of SEM images were generated by digital twins

Normal as well as images with defects

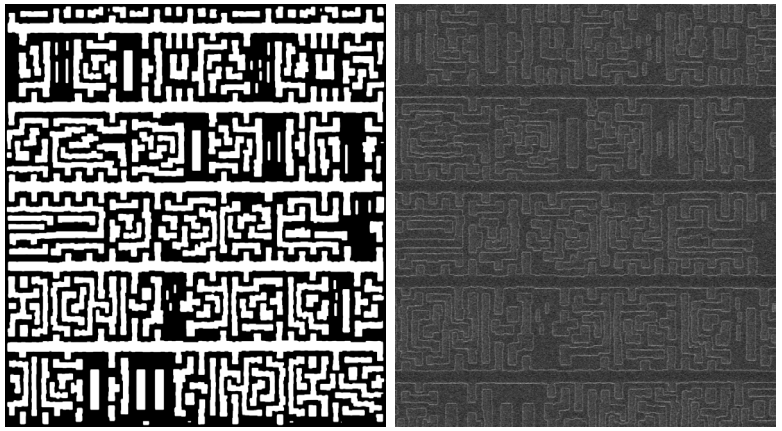


Collage of generated SEM images from DT

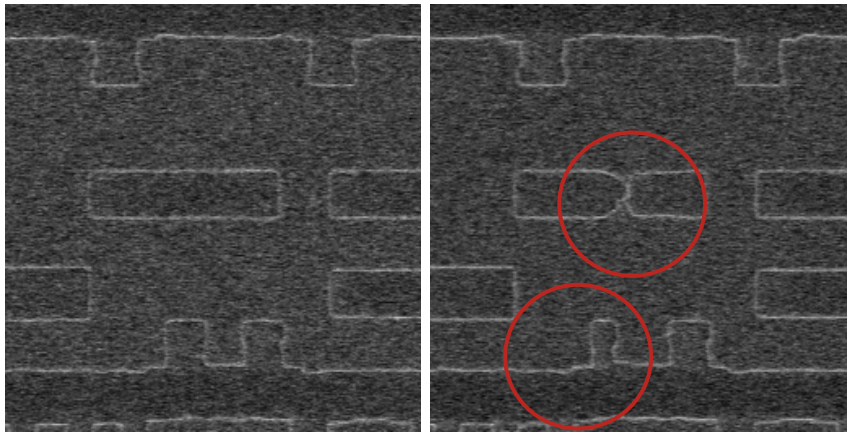
We Built Two More DL Tools for This Project



Automatically filter good quality SEM



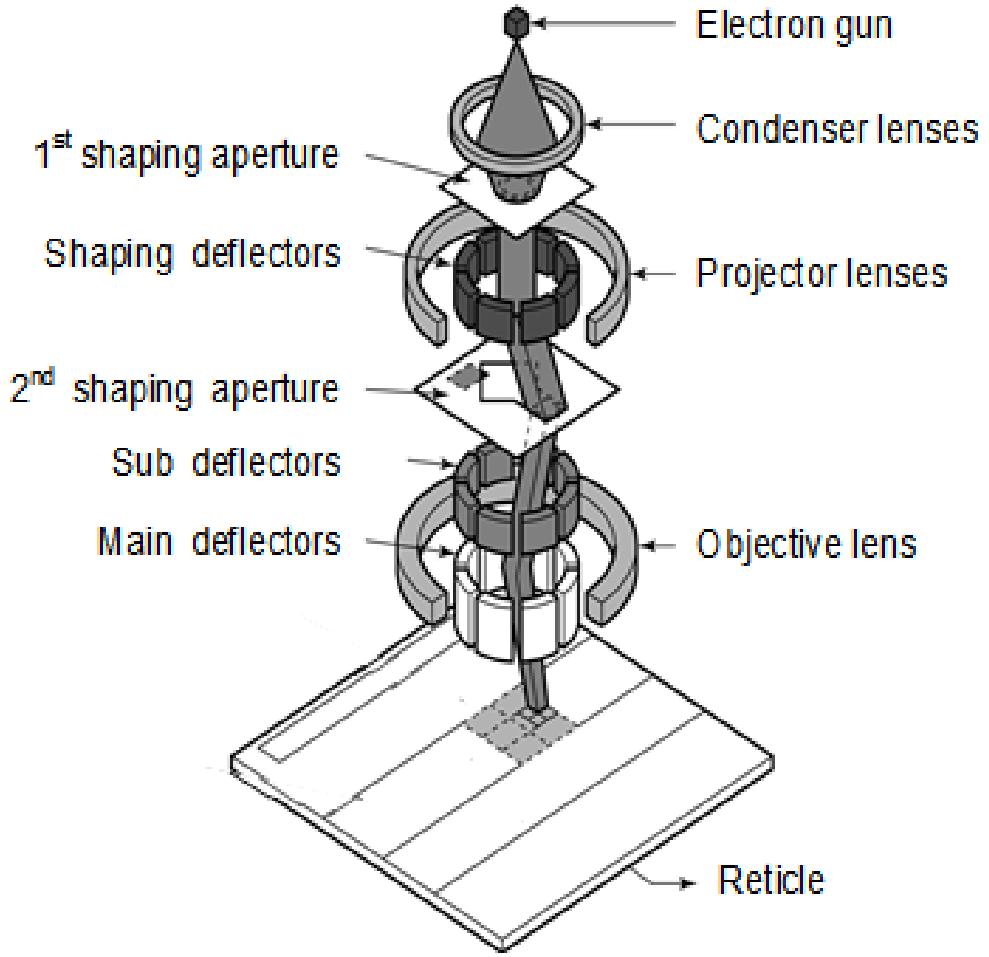
CAD to SEM image alignment



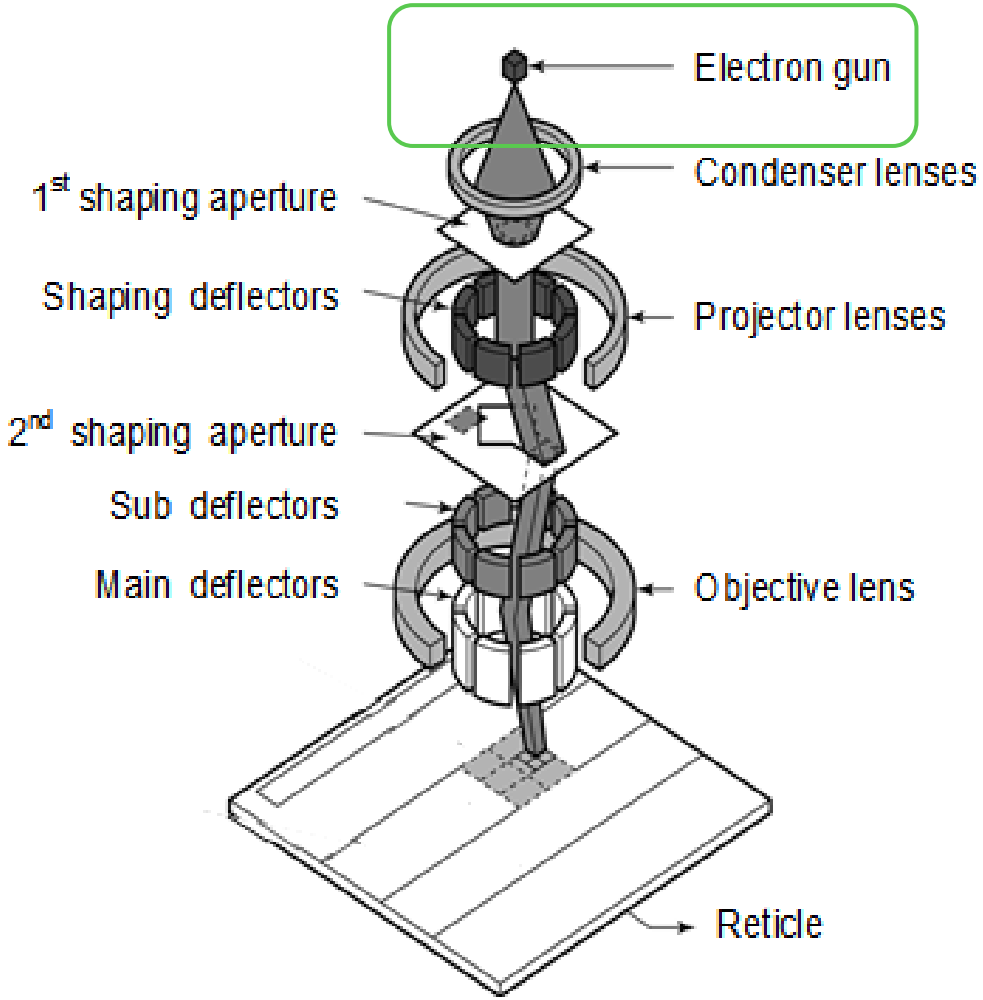
VSB writer defect classification

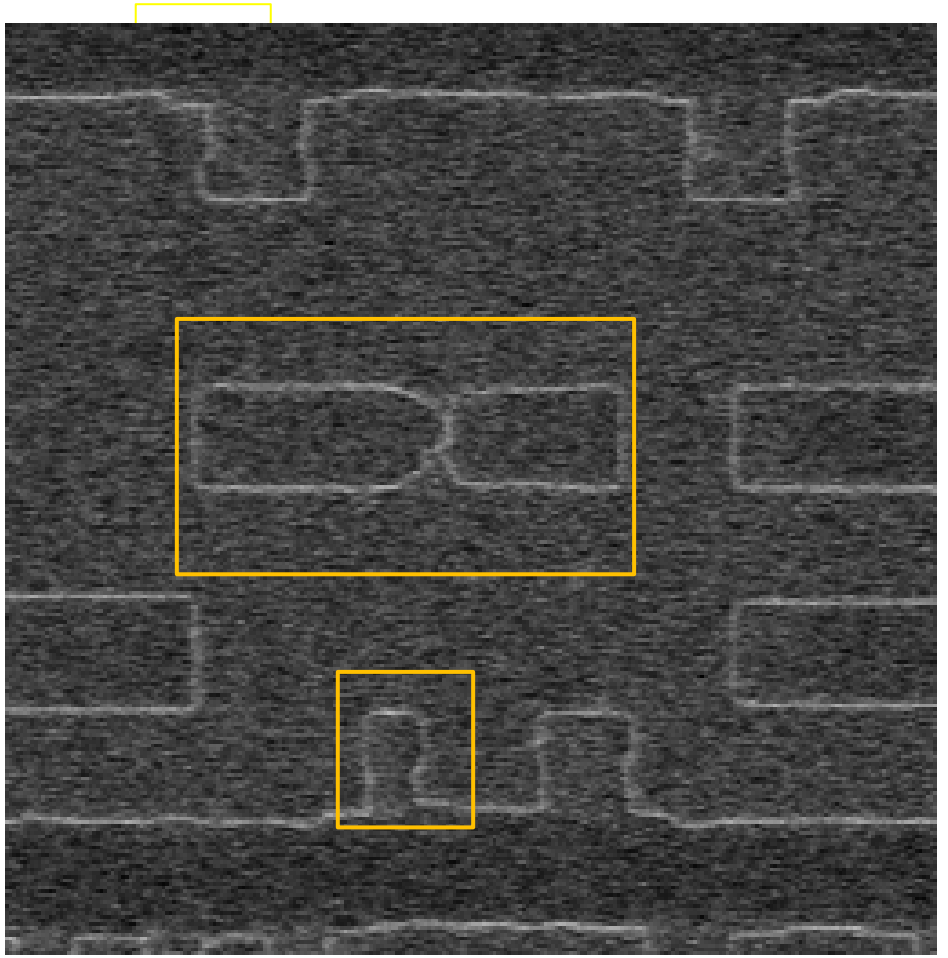
Architecture of a VSB Writer System

Components - Electron gun, apertures, deflectors; prints rect & triangle patterns



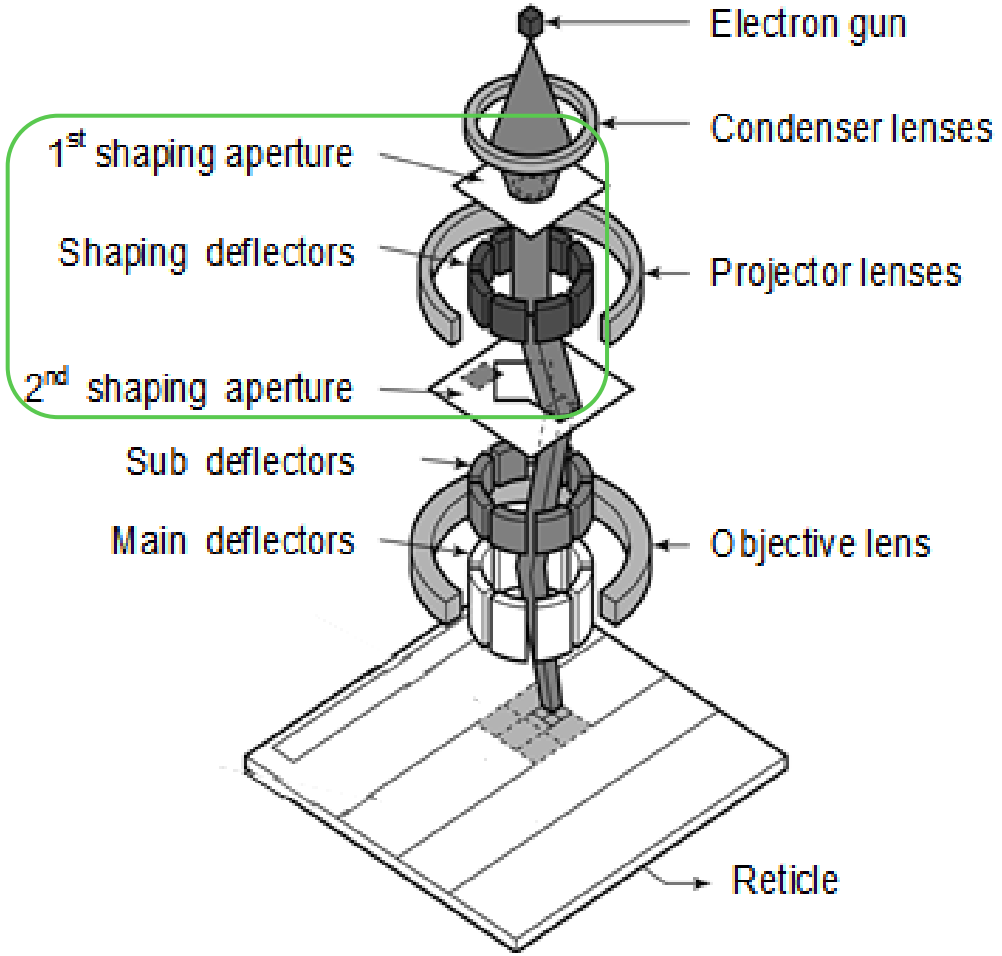
Dose errors: intensity variation of the electron gun

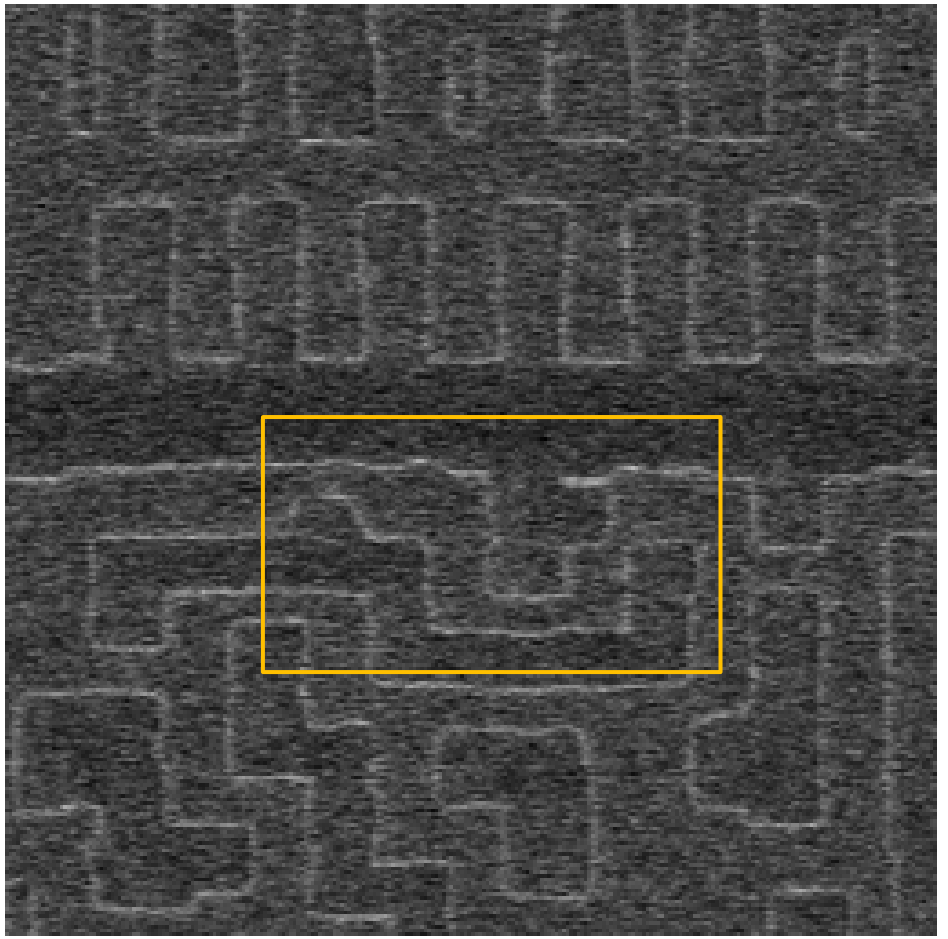




Dose errors

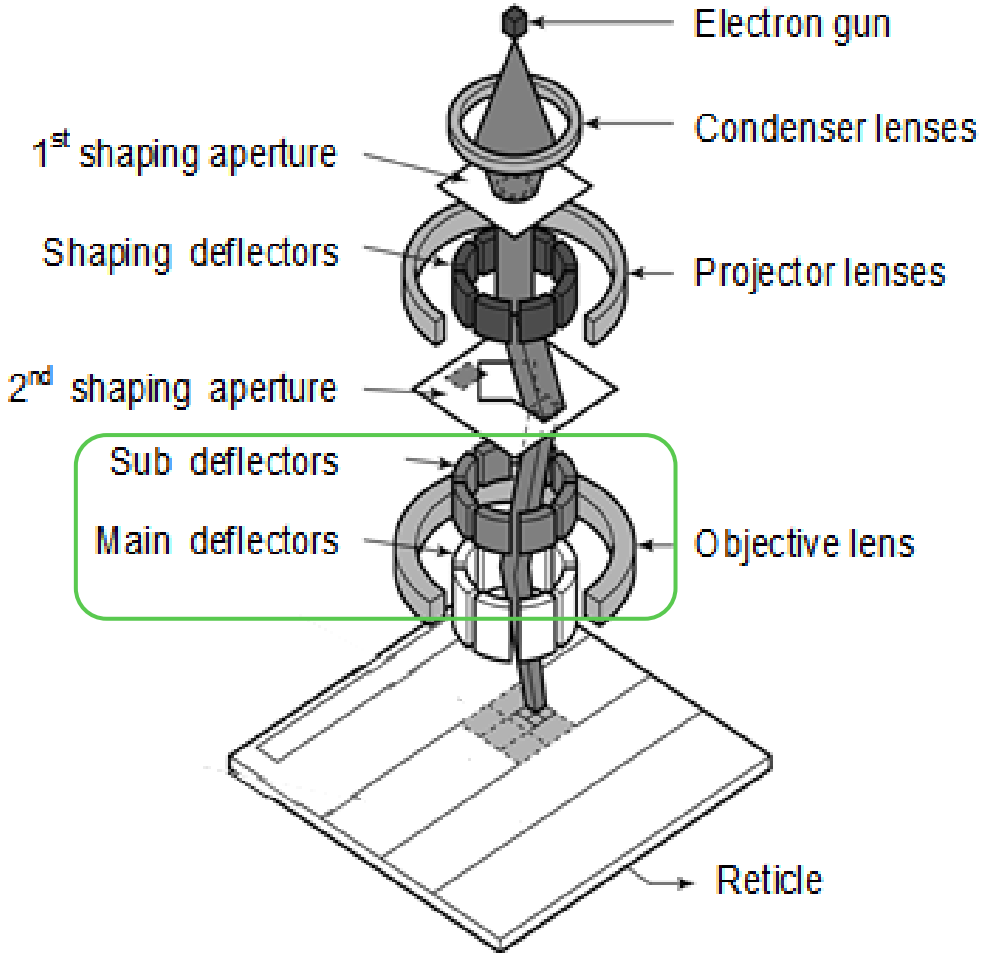
Shape errors: apertures & Shaping deflectors issues

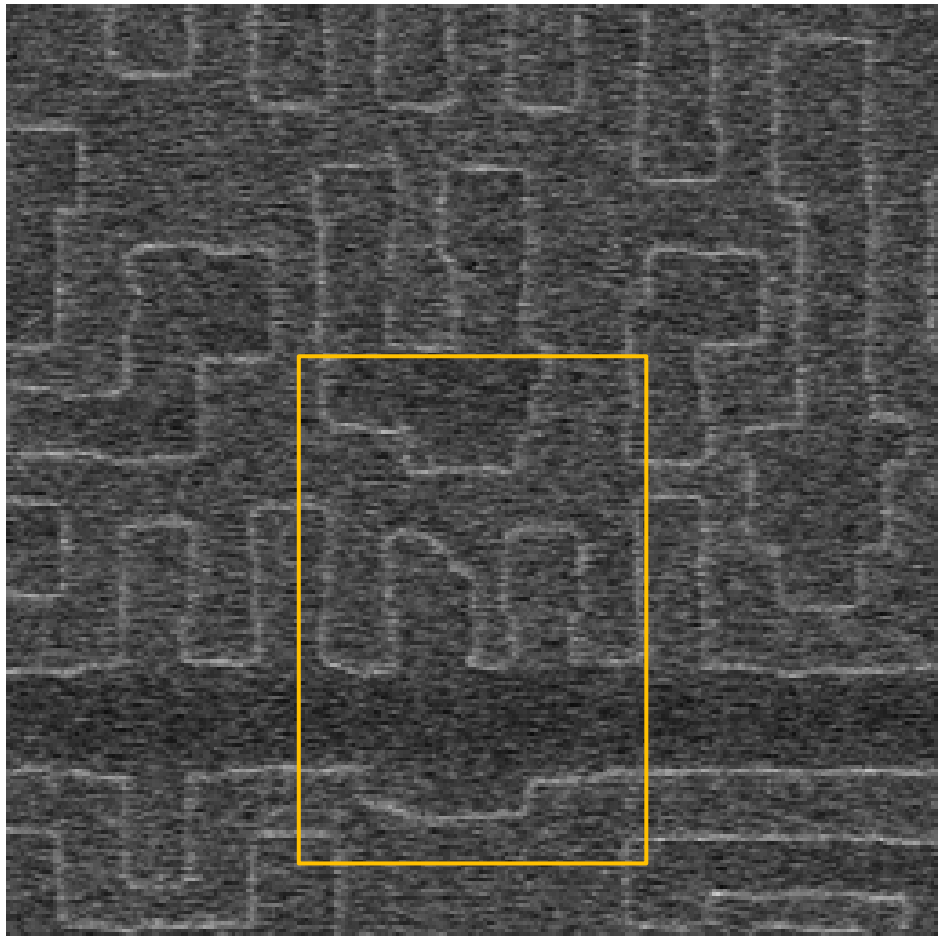




Shape errors

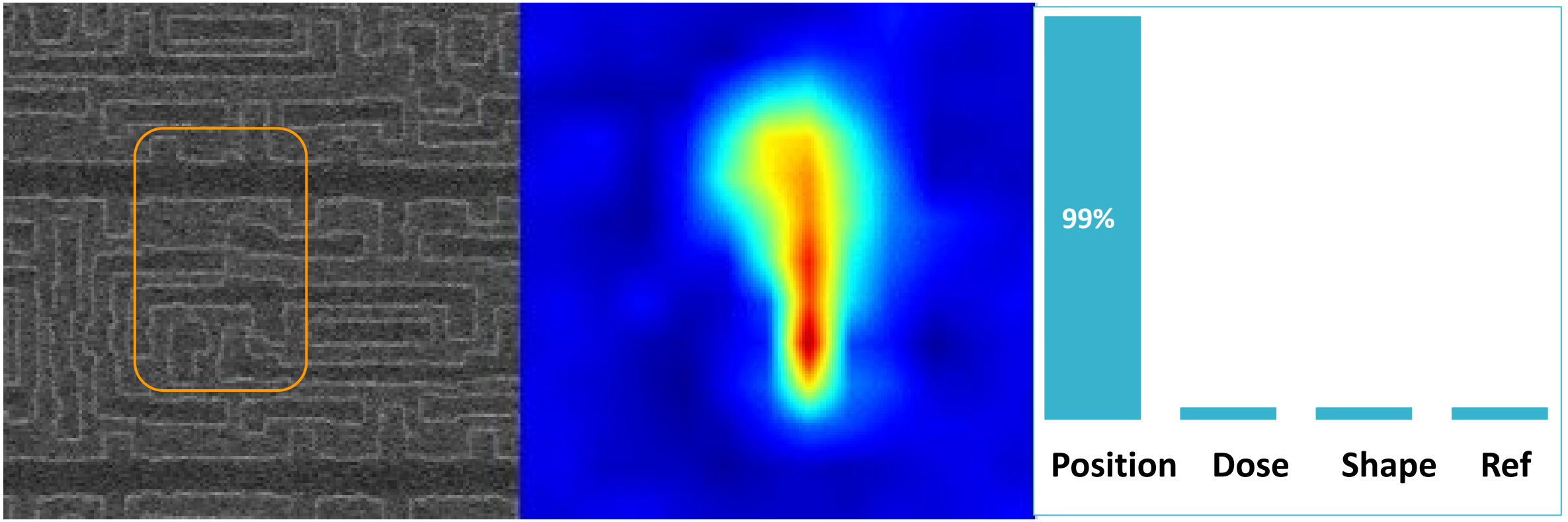
Position errors: Sub, Main deflectors concerns



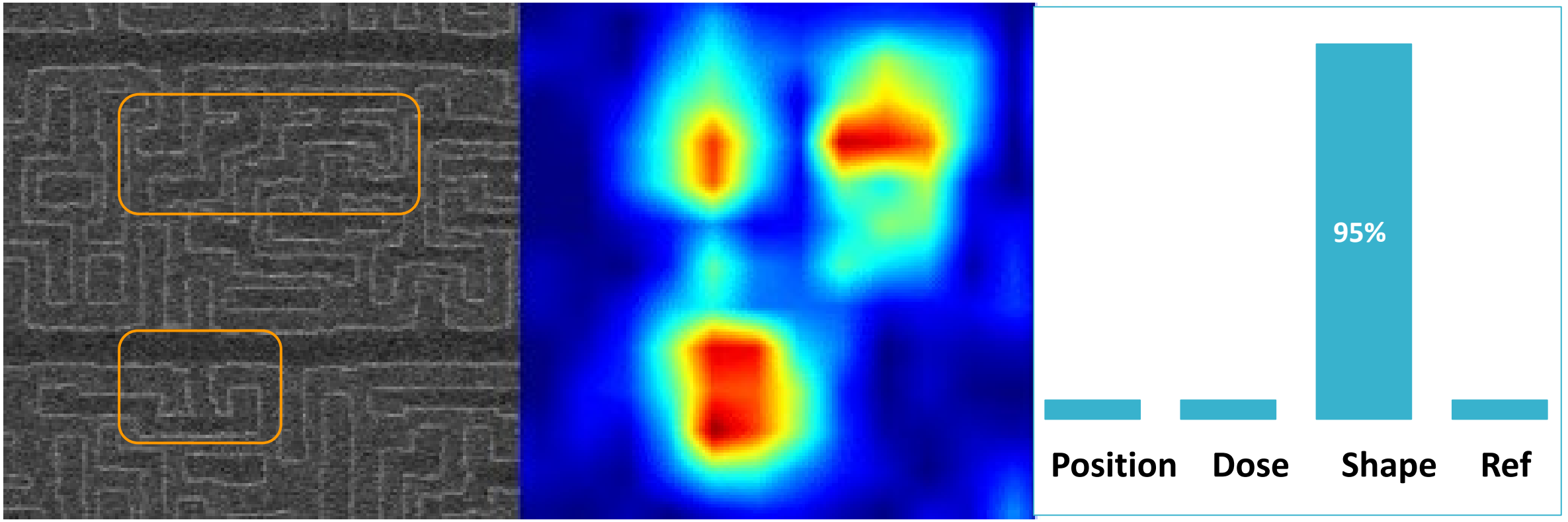


Position errors

DL sees the position error

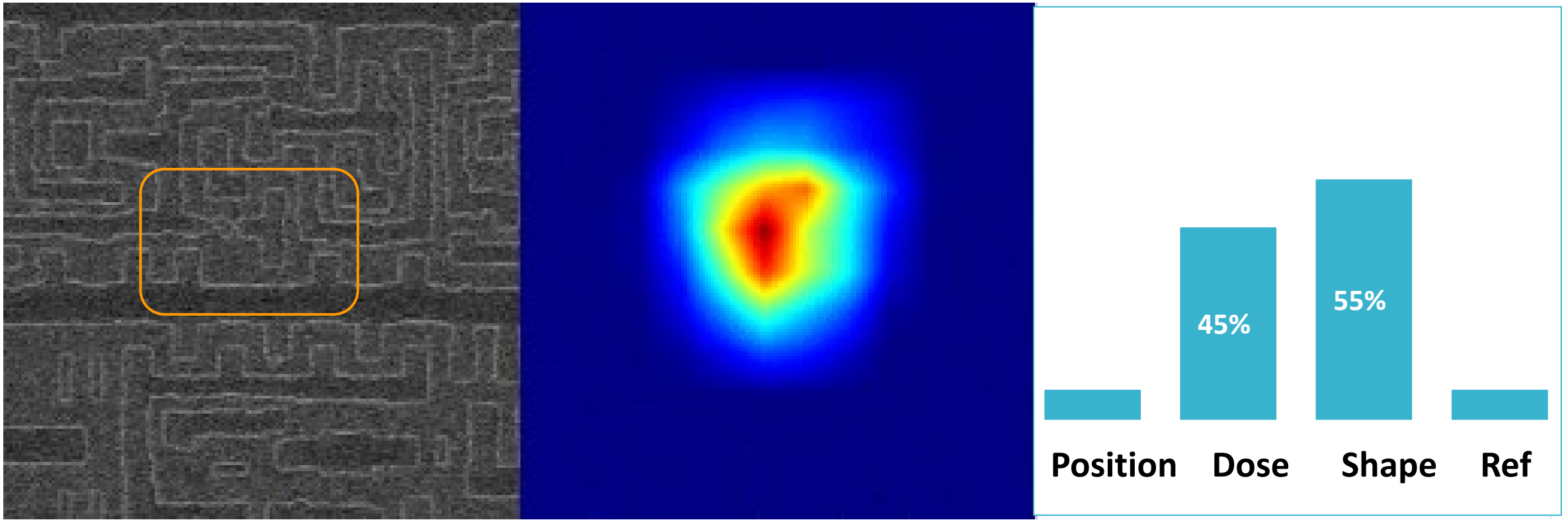


DL sees shape error



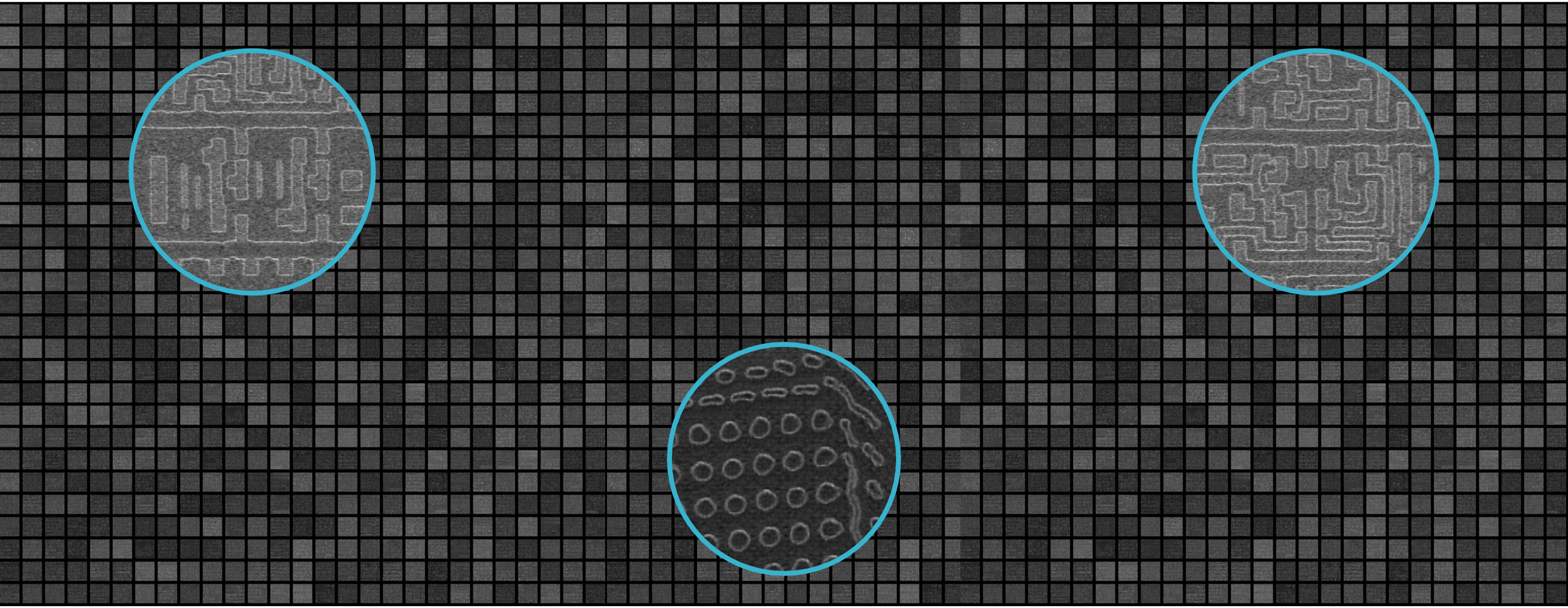
Some errors are difficult to discern

Hard to detect, even by human experts



Digital twins are essential for DL projects

Over two million “SEM” images were used



Also Essential : Selecting the Right Projects

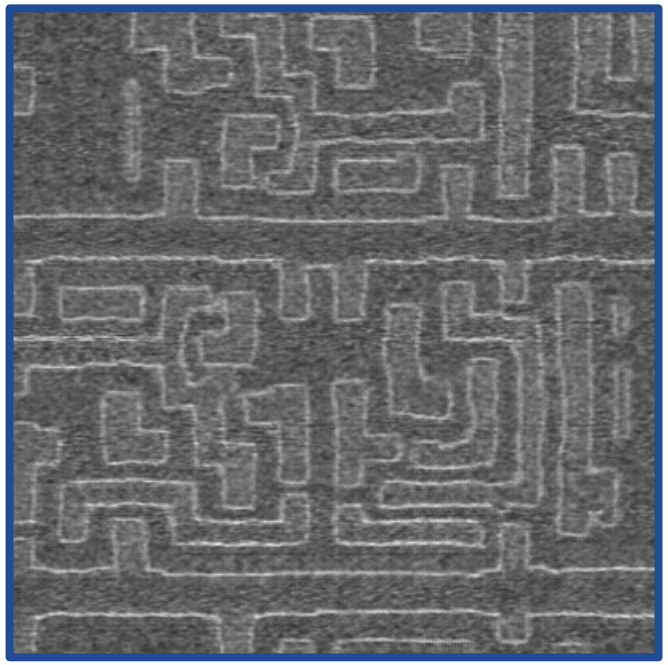
Not Suited to DL

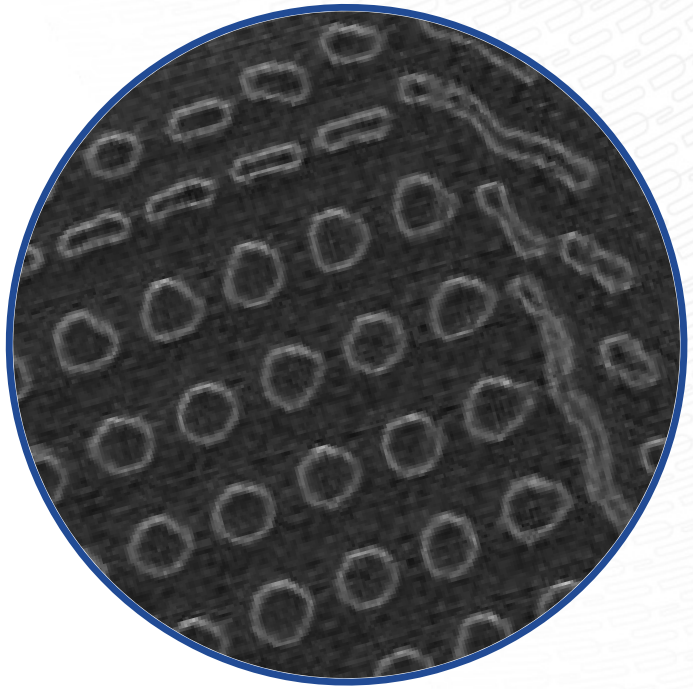
- DL-only constructive tools (ILT, MPC)
  DL is statistical: makes mistakes, though can be bounded
- eBeam or litho simulation
  Faster and more accurate analytically
- DL-only verification tools (MRC)
  DL is statistical: makes mistakes, though can be bounded

Good for DL

- Estimations for other tools
- Prediction applications
- Initial-condition acceleration for iterative optimization
- Categorization
- Quick prototype to prove feasibility
- Digital Twins : generate training data

DL is Shape-Agnostic



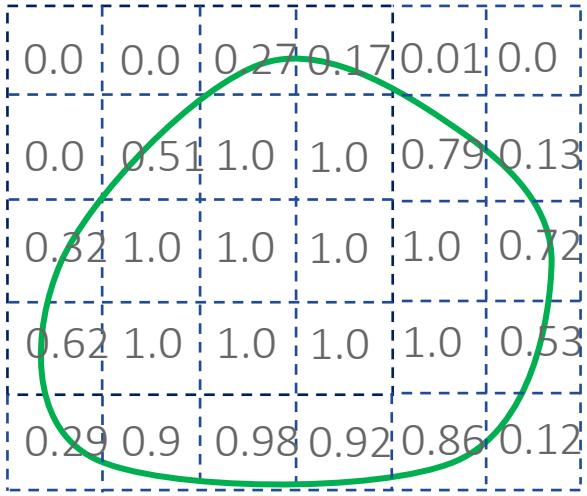


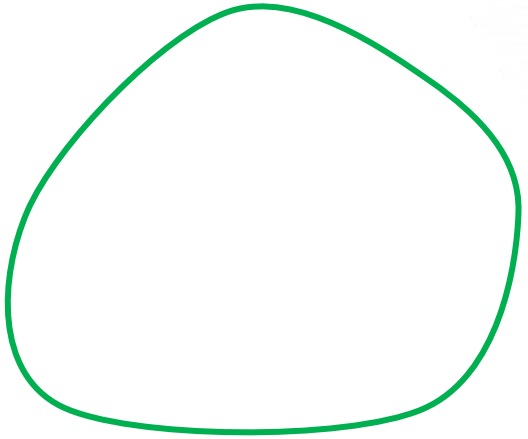
Deep learning inferencing runtime is the same, no matter what shapes are involved

DL Uses Rasterized Data

DL works just as well and just as fast on Curvy Masks

Magic of Rasterization





=

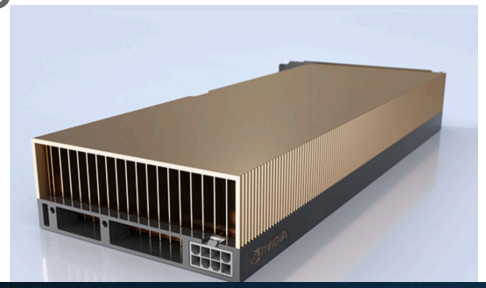
0.0	0.0	0.27	0.17	0.01	0.0
0.0	0.51	1.0	1.0	0.79	0.13
0.32	1.0	1.0	1.0	1.0	0.72
0.62	1.0	1.0	1.0	1.0	0.53
0.29	0.9	0.98	0.92	0.86	0.12

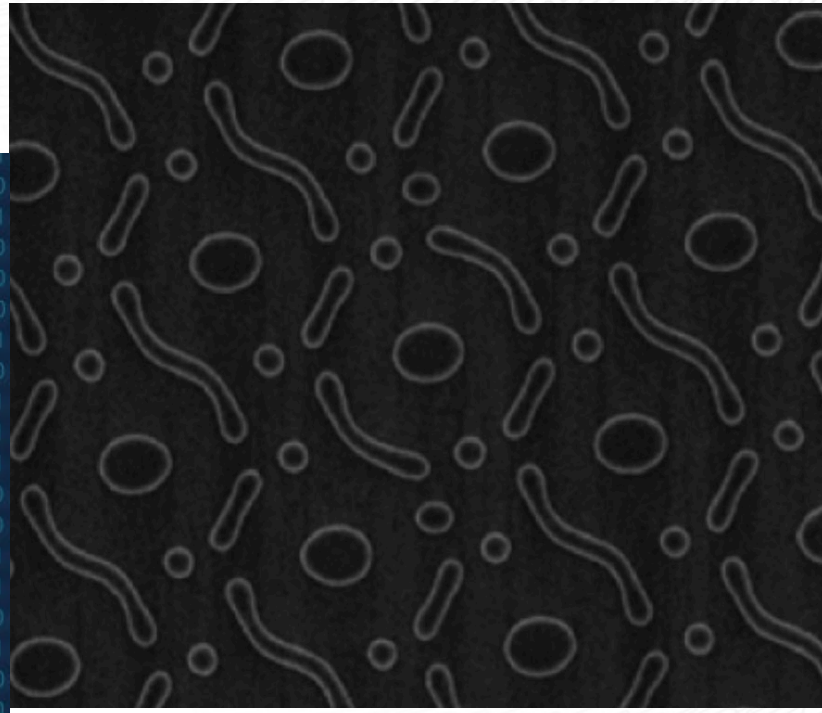
Mathematical duals, given a resolution limit

If your DL data is already in pixel domain, there is a huge advantage for inferencing time

Pixels + GPU + Curvy = The Future

Deep Learning is a Natural to Accelerate This



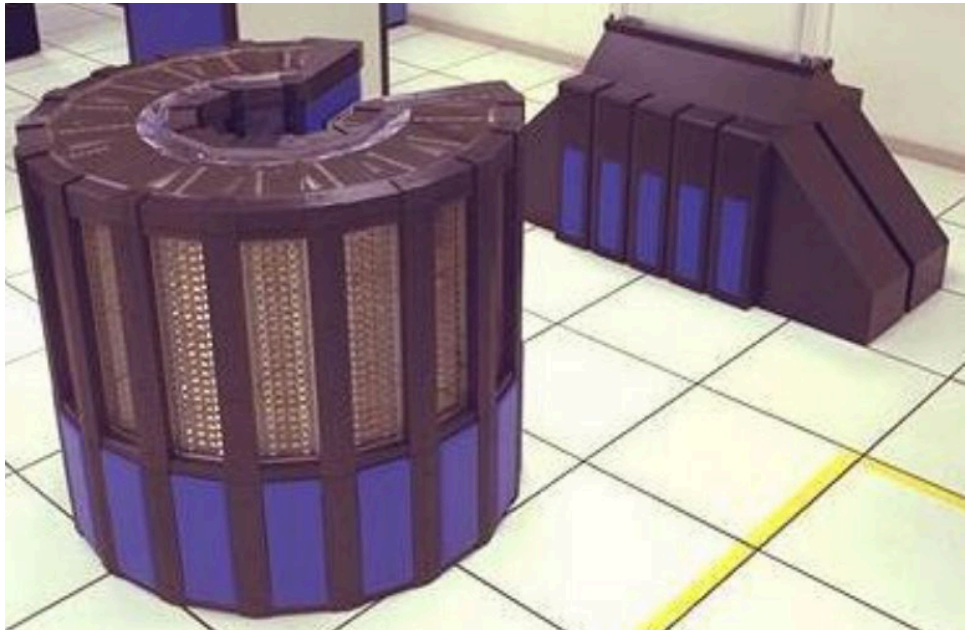


Today's GPU Workstation = 8,000 Cray-2s

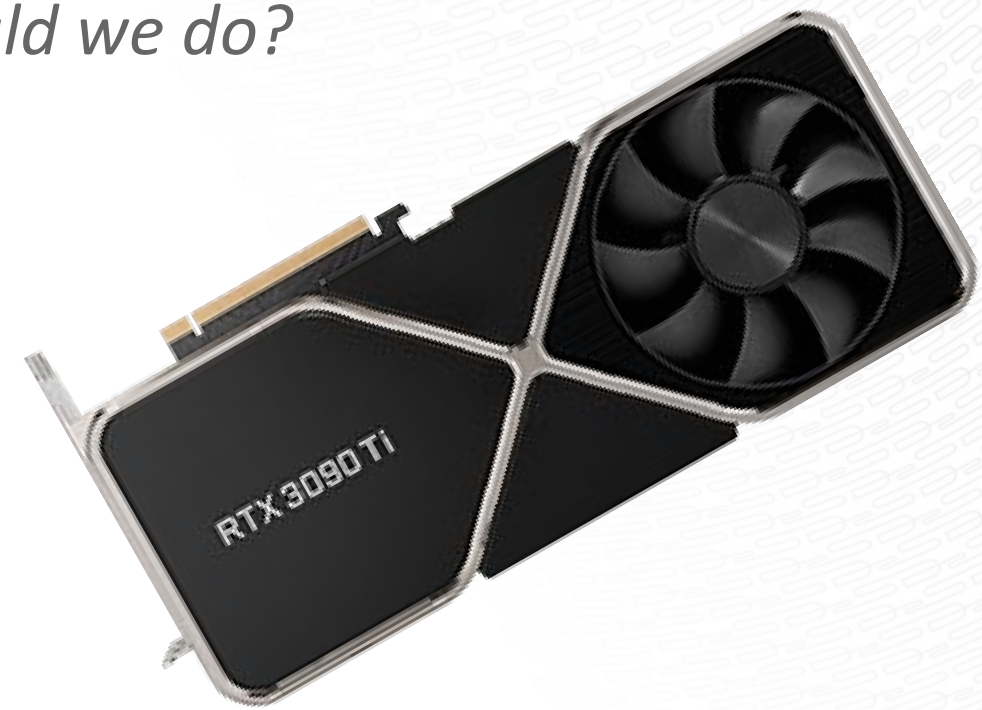
60,000,000x Price Performance

Deep Learning's "Useful Waste" was Enabled by This

What else could we do?



Cray-2 (1985)
1.9 GFLOPS w/500MB @ $15M



nVIDIA RTX 3090 Ti (2021)
15,300 GFLOPS w/24GB @ $2,000