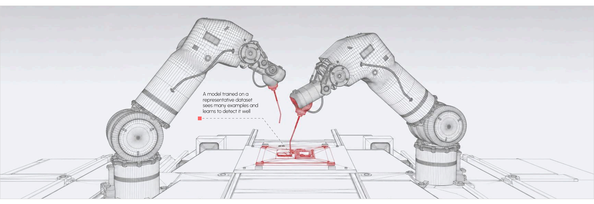
Case Study – Industrial Manufacturing

WHEN A WELD-INSPECTION MODEL IS MOST RELIABLE ON THE DEFECTS THAT MATTER LEAST



CASE STUDY — AUTOMATED RADIOGRAPHIC WELD INSPECTION

An automated inspection system X-rays each weld and passes the image to a detection model. The model looks for the defect types the weld standard defines: porosity, slag inclusions, lack of fusion, incomplete penetration, undercut, and cracks. Models like this reach around 85% accuracy on test data. On the strength of that number, the model is trusted to sort welds on the line, passing the ones it reads as sound and sending the rest to a radiographer.

The number is accurate. What gets concluded from it is not.

WHAT THE AVERAGE LEAVES OUT

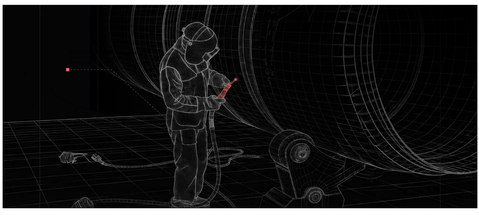
Weld defects differ in how often they occur and in how much damage they do.

Porosity is gas trapped in the weld metal. It happens often, and in a radiograph it shows up as rounded, high-contrast marks. A model sees plenty of examples and learns it well. Lack of fusion is similar: common enough and distinct enough to be learned reliably.

Cracks are the opposite on every count. They are rare, because welding procedures are designed to prevent them. They are hard to see, often appearing as a thin, faint line barely wider than a pixel or two, with branching that blends into the surrounding texture. And they are the most dangerous defect there is, because a crack grows under load and is where catastrophic fracture starts.

ISO 5817 treats them accordingly. Cracks are not permitted at any quality level.

So the model is best at the defects that are common, visible and comparatively tolerable. It is worst at the one that is rare, faint and never allowed. The 85% figure is an average carried by porosity and fusion. Underneath it sits a much lower detection rate on cracks.



LOW CONFIDENCE DOES NOT WARN YOU

The obvious hope is that the model will at least be unsure when it meets a crack it cannot handle, and that a low score will send the case to review.

It does not work that way.

When a crack falls in a part of the model's learned space with few examples nearby, the model does not register doubt. It matches the image to the closest thing it knows. For a faint crack against textured weld metal, the closest thing it knows is often sound weld. So it returns a confident pass.

Confidence measures how strongly the model prefers one answer over the others. It says nothing about whether the image resembled anything the model was trained on. A confident pass on a good weld and a confident pass on a missed crack look identical at the output.

Retraining helps, but only where the new examples land. Cracks vary in orientation, length, branching, base material and exposure. The next one missed is the one the expanded dataset still did not cover, and the model will be confident about that one too.

MAKING THE BOUNDARY VISIBLE

The team cannot see where the model's competence ends, so they cannot tell a safe pass from a dangerous one.

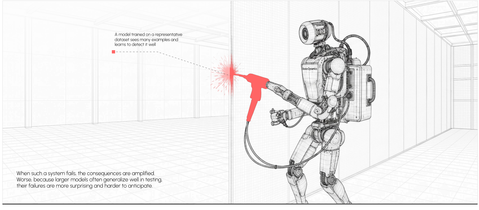
Squint Vision Studio maps the model's decision space and separates the regions where its reasoning is well supported from the regions where inputs are sparse or overlapping. For weld inspection this surfaces what the accuracy figure hides: thin and unusual cracks sit in regions where the model's judgment cannot be relied on, whatever its confidence says.

From that map, the team builds a watchdog and exports it into the deployed system. When the model calls a weld sound from inside one of those unreliable regions, the watchdog flags that specific decision for the radiographer. Porosity and fusion cases keep flowing through automatically. The rare, faint, zero-tolerance cases go to the person best placed to catch them.

WHAT THIS DOES NOT SOLVE

This catches cracks that fall in sparse or ambiguous parts of the model's decision space, which is where the dangerous misses concentrate.

It does not guarantee catching a crack that looks, to the model, like ordinary sound weld. If a defect falls squarely inside a region the model handles well and still gets read as normal material, a map of where the model is unsupported will not necessarily flag it. That is a different and harder problem.



WHAT CHANGES

Without this, an inspection programme is relying on an average to stand in for a guarantee it cannot give. The figure says the model is usually right. It does not say the model is right on cracks, and it is reliably worst there.

With it, the inversion becomes something you can manage. The model keeps the throughput on the cases it handles. The cases it cannot be trusted on are identified and sent to a person. The inspection stops being governed by an average and starts being governed by a map of where the model can be believed.

This scenario is representative, drawn from published research on automated radiographic weld inspection. The defect types, detection difficulty, the ISO 5817 rule on cracks and the reported porosity precision reflect the state of the field. Squint Vision Studio's decision-space mapping and exportable watchdog modules are described as designed.