Inspection Simulation to Enhance the Reliability and Validity of Automated X-ray Data Analysis

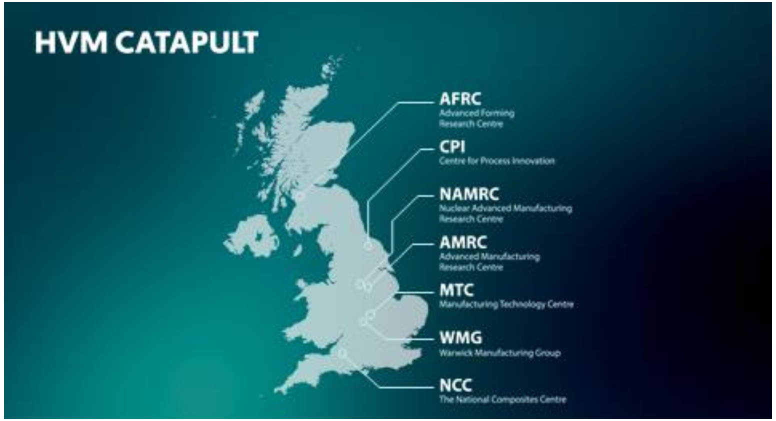
Robert Culver

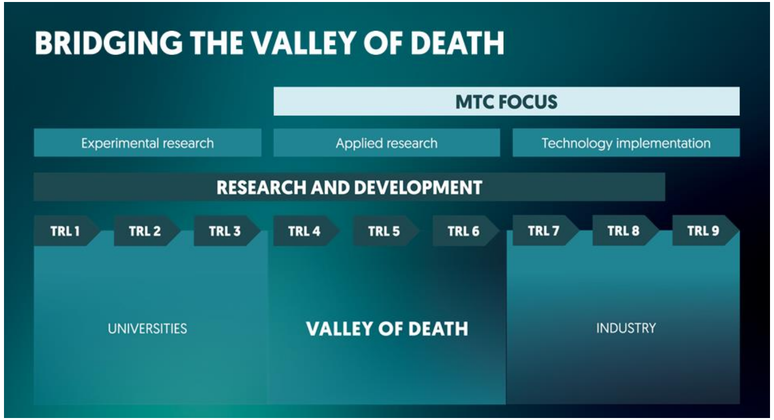
2022 / 11 / 03
ICAM 2022

Introduction to the Manufacturing Technology Centre (MTC)

MTC background:

- The MTC is a research organisation founded to bridge the gap between academia and industry. Part of the UK’s High Value Manufacturing (HVM) catapult.
- Founded in 2010 by four founding partners: University of Birmingham, University of Nottingham, University of Loughborough and TWI.
- Wide variety of Manufacturing Technology projects:
    - Additive Manufacturing
    - Automation and Robotics
    - Electronics Manufacturing
    - Manufacturing Simulation
    - Data Informatics
    - Non-conventional machining
    - Metrology & NDT…
- More than 100 industrial members.





Speaker Bio

Robert Culver

Working at the MTC for 5 years with the Metrology & Non-Destructive Testing group. Most of work involves developing techniques for **computer-aided inspection** and **NDT 4.0**.

Came to NDT from photonics background gained from PhD in atomic physics and laser cooling.

Current research includes:

- Developing tools for simulation and **optimisation of inspection procedures**
- Exploring novel routes to **automated data analysis**

Talk Contents

- Introduction to X-ray inspection simulation at the MTC
- Optimising Automated Defect Recognition (ADR) in X-ray Computed Tomography (XCT) with experimental data
- Optimising and mapping ADR performance in XCT with synthetic data
- Optimising and mapping deep learning ADR performance in 2D digital radiography.
- Considerations around using these tools in production
- Conclusions

Introduction

X-Ray Inspection Simulation Tools

Challenges for X-ray inspection and ADR

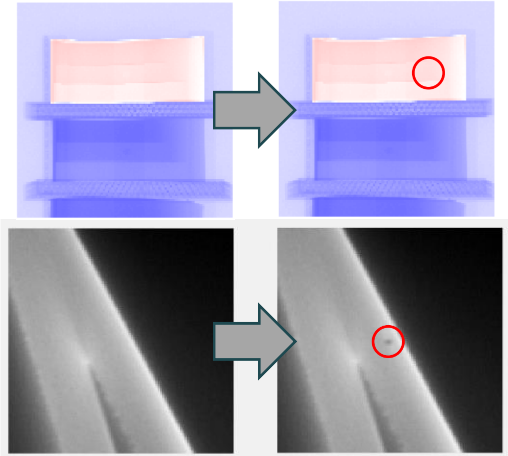
- X-ray inspection has **several sources of noise**
  - Acquisition noise sources
    - X-ray scatter
    - Poisson noise
    - Focal spot blur
    - Detector unsharpness
    - Electronic noise
  - 'Noise' from X-ray Computed Tomography (XCT) reconstruction
    - Polychromatic X-ray beam (beam hardening effects)
    - Reconstruction artefacts (cone beam artefacts, streaking etc.)
- The impact of **these effects will vary across the part geometry**.
- Without simulation tools, it is **difficult to predict how the impact** of these effects will vary across the geometry, particularly for complex AM geometries.
- Any challenges for X-ray will feed across into **challenges for Automated Defect Recognition (ADR)**.

Introduction

X-Ray Inspection Simulation Tools

- Regardless of the challenges, we must be certain that our inspection procedure (whether this includes automated analysis or not) is capable of finding the defects that we care about.
- The route typically taken is to build many test samples with seeded defects and verify that the inspection procedure is able to identify the defects. This is costly, produces a limited number of data and many scrap components.
- Using calibrated X-ray simulation tools, we can use **synthetic defect indications** to validate the inspection procedure.
- Synthetic defect indications can also be used to train Automated Defect Recognition (ADR) algorithms.

A 'conventional' validation approach
(manufacture many defects)

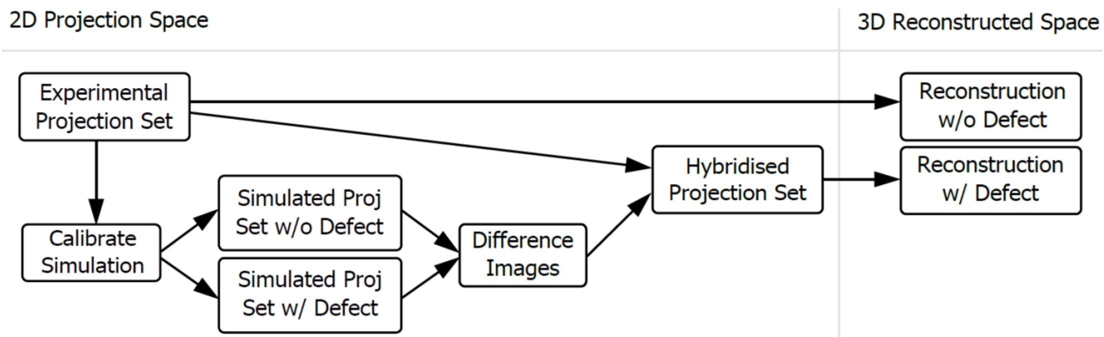


Simulation-driven validation approach
(simulate many defect indications)

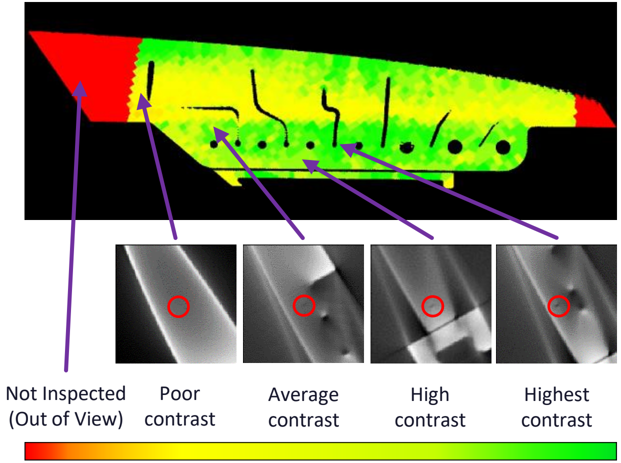
Introduction

Previous Work at the MTC (Presented at ICAM 2020)

- A previous project at the MTC was carried out, “Mapping Computed Tomography Performance” (MCTP) to map the **expected variability of X-ray Computed Tomography (XCT) performance.**
- Use hybrid simulation approach, which combines experimental data with synthetic defect indications.
- This approach works by simulating many thousands of defect instances and seeing **how CNR varies** across the component. The results can then be **mapped onto the component**.



Hybrid defect generation algorithm. This technique allows synthetic defect indications to be augmented into defect-free experimental datasets.



Predicted Contrast to Noise (CNR) values from MCTP analysis on ARA aerofoil component.

Brierley et.al. 2019, NDT & E International 107, 102127

Introduction

The aims of this work…

- While the MTCP approach can map expected CNR, it cannot say definitively whether a hypothetical defect will be detected or not by a human inspector (a human-led Probability of Detection (PoD) analysis would help answer this question...).
- However, feeding these defects into an Automated Defect Recognition (ADR) algorithm can give you a definitive answer of where defects are being detected successfully / unsuccessfully.
- The synthetic data can also be used to train and maximise the ADR performance.

Thus, this project had two main goals:

- Use synthetic data to **train and optimise the ADR performance** on an AM use case component
- **Map the final performance of the ADR** across the component

Introduction

Notes on this Project

- Reliability-Driven Automation of Radiographic Data Analysis (“RDA-Squared”)
- Three client project: BAE Systems, Rolls-Royce, GKN Aerospace;
- Concerned with ADR for 3D X-Ray CT (XCT) and 2D radiography;
- Use case component: AM ‘DITTO’ bracket, supplied by BAE;

‘DITTO’ AM Bracket, BAE Systems.

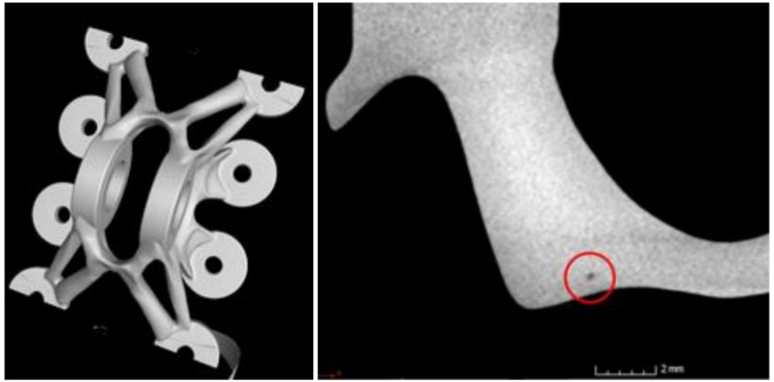
Task	X-Ray Modality	ADR Approach	Ground Truth Used for Training
1	XCT	Conventional image processing	Manual labels on experimental data
2	XCT	Conventional image processing	Simulated (hybrid) defect indications
3	DR	Deep learning (Mask R-CNN)	Simulated (hybrid) defect indications

Task 1: ADR Optimisation on Manually-Labelled XCT Datasets

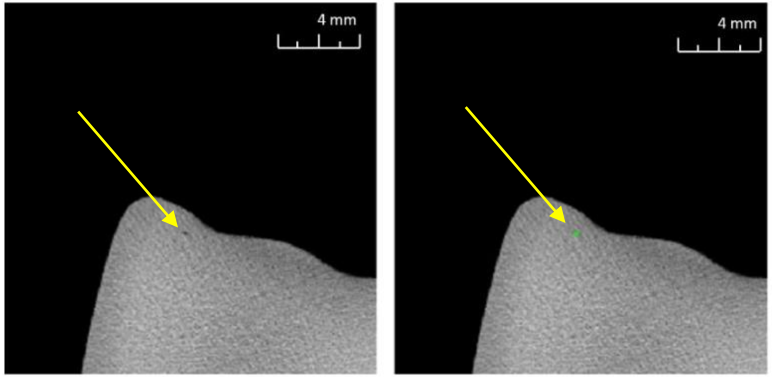
Task 1: ADR Optimisation on Manually-Labelled XCT Data

Data Creation

- The first task was used to establish a **framework for optimising conventional image-processing** ADR algorithms based on **manually-labelled experimental datasets**.
- The MTC was supplied with four DITTO bracket components:
  - 3 each with one seeded defect,
  - 1 with no seeded defects,
  - No other defects were detected in components
- All components were scanned using XCT, and corresponding Boolean defect mask datasets were created manually.



XCT dataset and highlighted defect



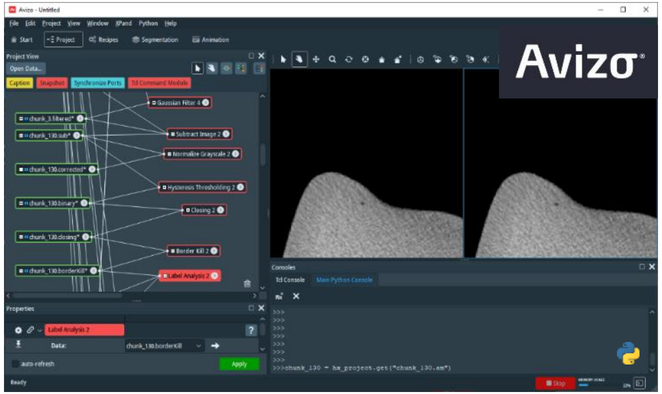
Original slice

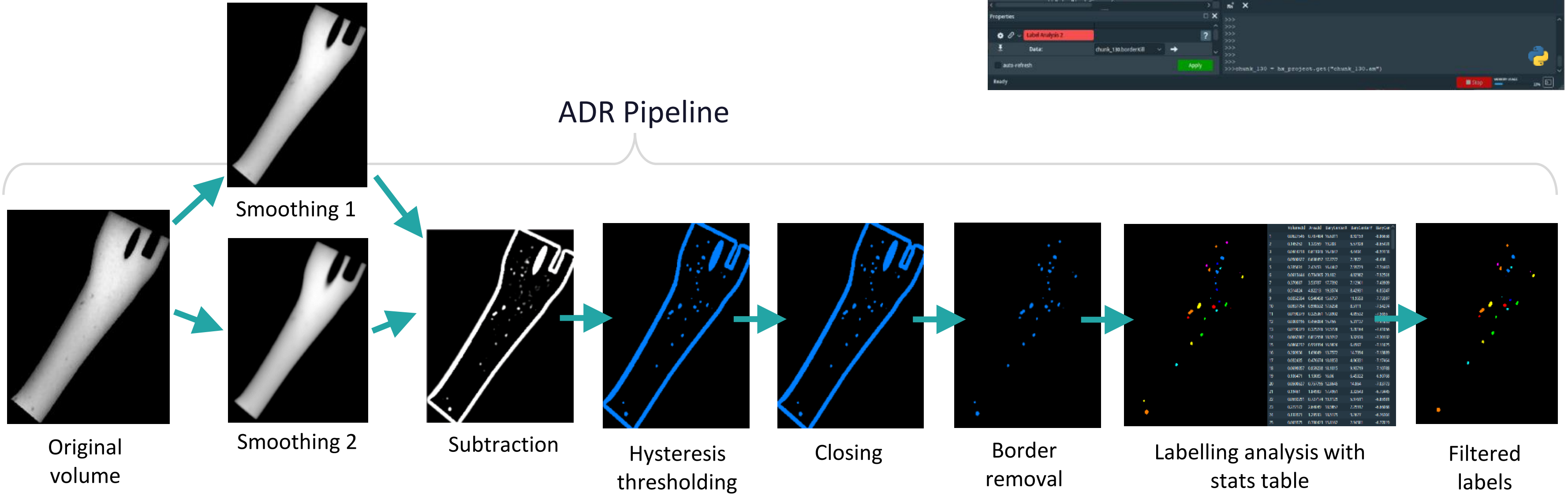
Labelled slice (Boolean mask highlighted in green)

Task 1: ADR Optimisation on Manually-Labelled XCT Data

Conventional Image Processing ADR

- Used Avizo Inspect software to create a custom ADR pipeline.
- Avizo Inspect comes with an integrated Python console, which allowed us to drive the software with an optimiser, automatically varying the 9 pipeline parameters and re-running the ADR.





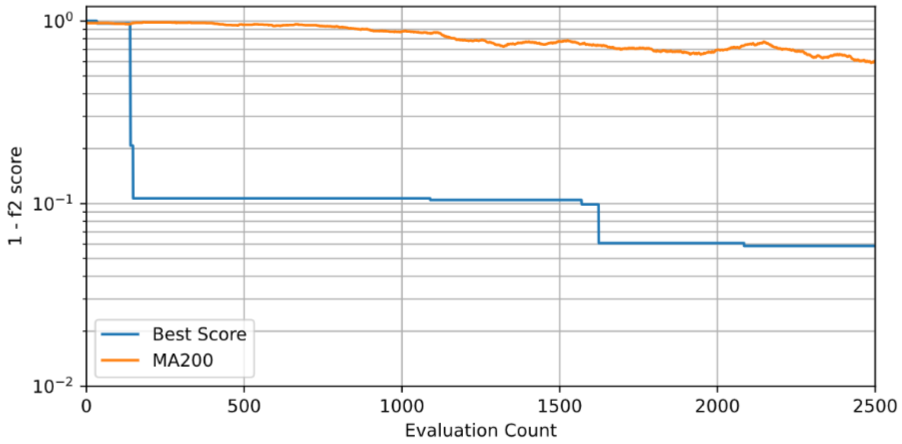
Task 1: ADR Optimisation on Manually-Labelled XCT Data

Conventional Image Processing ADR

- From Avizo Python environment, **ran optimiser on 9 ADR parameters** (Shuffled Complex Evolution with PCA, “SP-UCI”) algorithm.
- Used **F_2 score as the optimisation metric** (single objective optimisation)
- Trained ADR using defective volumes as well as clean, defect-free volumes, to make robust to false positives.
- Based on optimisation timeline, the optimisation was likely stopped prematurely (due to time constraints)

Step	Operation	Purpose	Parameter	Min	Max
1.1	Gaussian Smoothing 1	High frequency noise reduction	Standard deviation	0	3
			Kernel size (voxel units)	1	3
1.2	Gaussian Smoothing 2	Identification of general greyscale trends	Standard deviation	1	10
			Kernel size (voxel units)	1	10
2	Subtraction	Amplify dark regions in component	N/A	-	-
3	Normalisation	Desensitise ADE to inputs of differing brightness / contrast	N/A	-	-
4	Hysteresis Threshold	Produce binary image	Low threshold	-2000	2000
			High threshold	-1000	10000
			Fuzzy zone length (voxel units)	1	10
5	Closing	Remove small holes in binary image	Kernel size (voxel units)	1	10
6	Labelling	Identify individual regions	N/A	-	-
7	Label Filtering	Remove small (noise) regions	Volume (mm^3)	0	0.01

Summary of ADR pipeline parameters



Timeline of Optimisation

Yang et.al. 2017 “SP-UCI” Algorithm, Information Sciences 418, 302

Task 1: ADR Optimisation on Manually-Labelled XCT Data

Recall / Precision / F_2 Score

	Defect Exists	No Defect Exists
Defect Detected	True Positive	False Positive
No Defect Detected	False Negative	True Negative

"Confusion Matrix"

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

$$Recall = \frac{TP}{TP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$F_\beta = (1 + \beta^2) \frac{Recall \cdot Precision}{Recall + \beta^2 Precision}$$

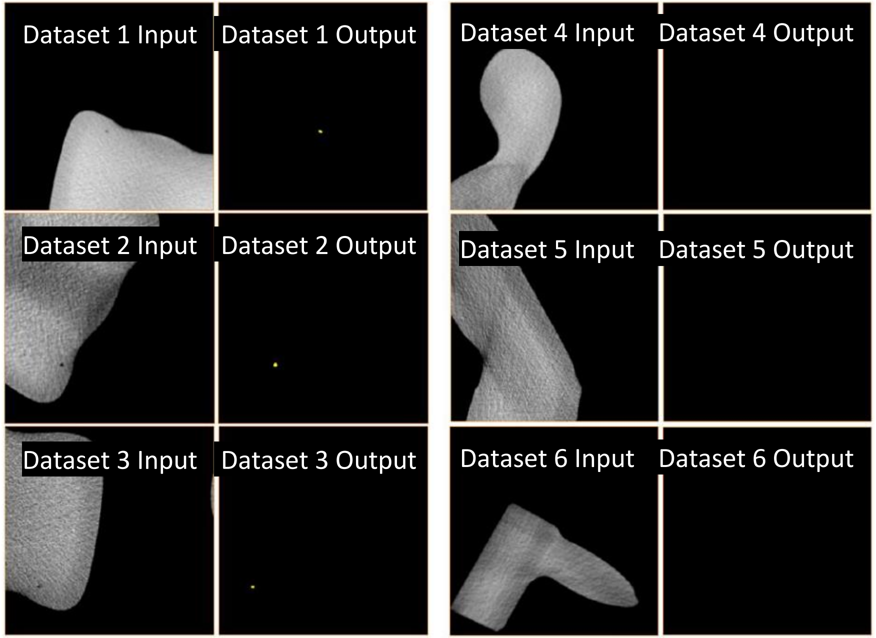
- Sensitive algorithms seek low FN at the expense of high FP
- High **Recall** at the expense of lowered **Precision**
- Trade-off can be managed by F_β score. We chose β=2, i.e., we value recall twice as much as we value precision.

Task 1: ADR Optimisation on Manually-Labelled XCT Data

Optimised ADR Performance

Dataset	TP	FN	FP	Recall	Precision	F_2 - Score
1	143	3	37	0.979	0.794	0.936
2	150	2	43	0.987	0.777	0.936
3	102	2	23	0.981	0.816	0.943
4	0	0	0	1.000	1.000	1.000
5	0	0	0	1.000	1.000	1.000
6	0	0	0	1.000	1.000	1.000

Optimised ADR performance



Comparison of inputs and optimised ADR outputs for each dataset. Datasets on right side are defect-free and were used to minimise false positives.

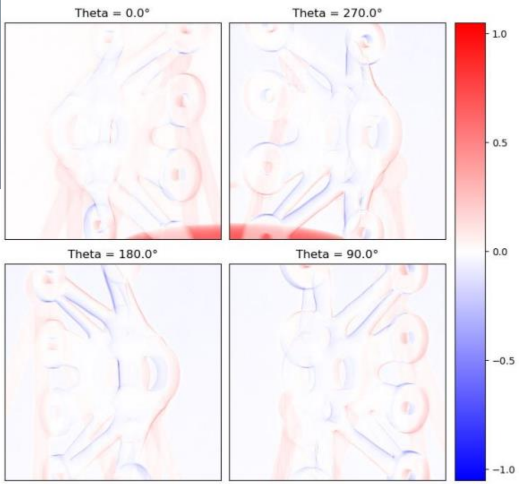
- This task demonstrated a framework for optimising a conventional image-processing Automated Defect Recognition (ADR) algorithm based on manually-labelled data…
- Now to perform the approach with synthetic defect indications…

Task 2: ADR Optimisation on Synthetic XCT Data

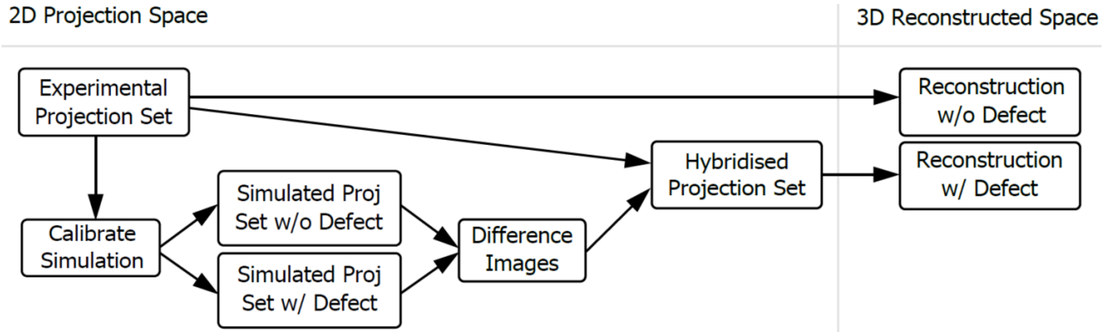
Task 2: ADR Optimisation on Simulated XCT Data

Calibrating the Simulation

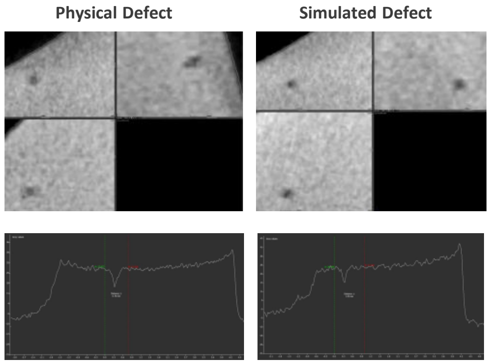
- X-ray simulation tools can be calibrated to experimental data
- Defects simulated with high-accuracy and added into experimental data
- Qualitative differences between the physical defects and simulated defects were small. Main difference was apparent defect size as-built vs specified in CAD.



Simulation calibration: difference images between simulated and experimental X-ray projections.



Algorithm for producing XCT datasets with simulated defect indications while retaining experimental noise

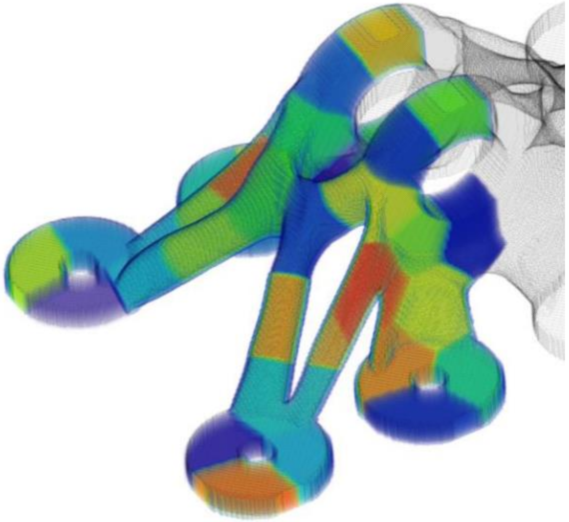


Comparison of physical and simulated defect indication

Task 2: ADR Optimisation on Simulated XCT Data

Simulated Defect Considerations

- Ideally need to consider ADR performance for:
  - Different defect positions
  - Different defect sizes
  - Different defect shapes
  - Different defect types
- **<u>Defect positions were addressed by separating component into zones</u>**. (useful for mapping results, as will be seen later).
- Only top half of the component considered (assume symmetrical behaviour)
- **<u>100 micron, 200 micron and 300 micron defects chosen</u>**.
- For Task 2, only considered spherical void defects (will use different shapes in Task 3).
- **<u>835 simulated defects generated across 9 XCT datasets.</u>**



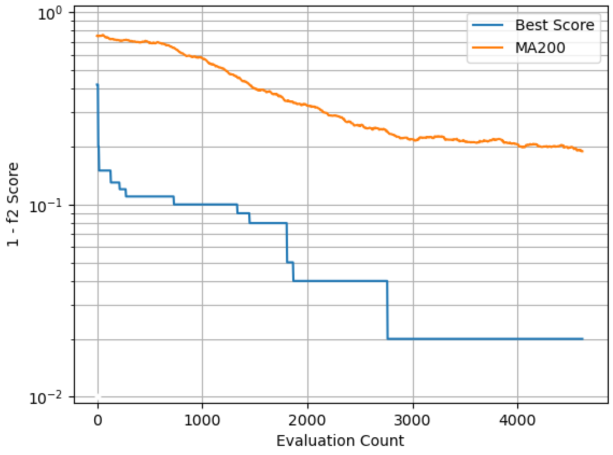
Component geometry segmented into zones

Task 2: ADR Optimisation on Simulated XCT Data

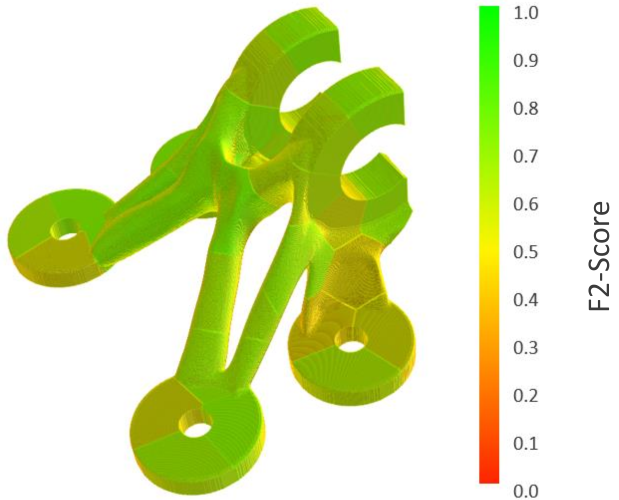
Simulated Defect Considerations

- Some changes were made to Task 1 approach for computational efficiency:
  - Code rewritten in pure Python environment
  - Only a subset of defects considered in each evaluation (while this allows more evaluations to take place, it introduces noise in to the optimisation)

Defect Size (micron)	Recall	Precision	F2 Score	Num Defects
200	0.91	0.92	0.91	259
300	0.95	0.68	0.88	259



Timeline of Optimisation

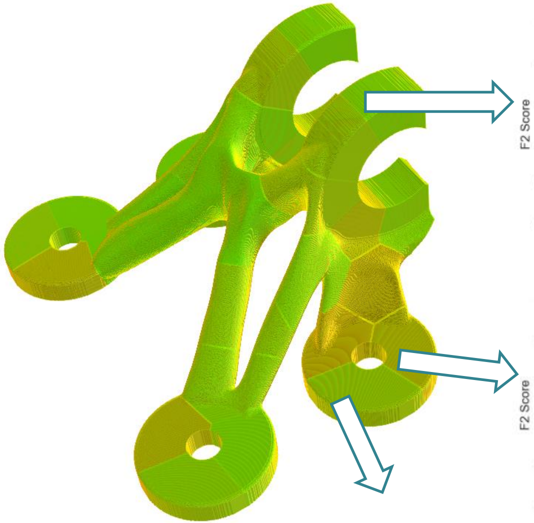


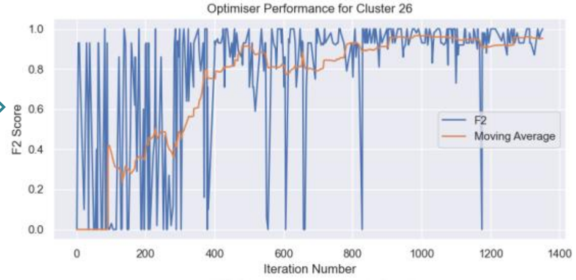
Optimised F2-Score values for each zone

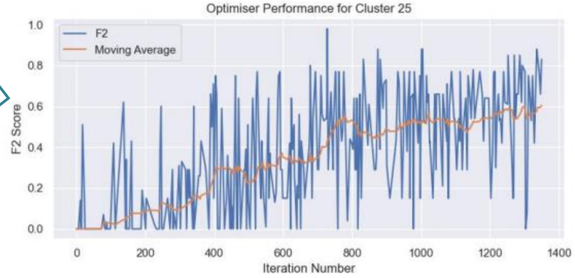
Task 2: ADR Optimisation on Simulated XCT Data

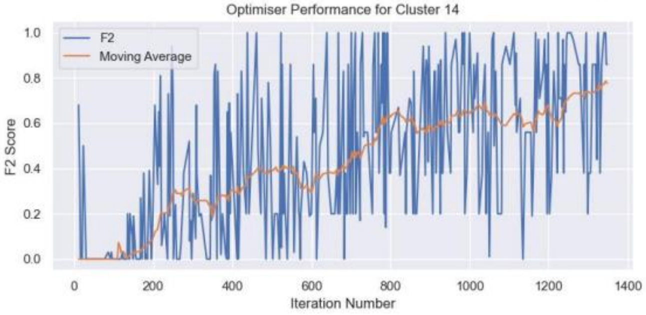
Region-Wise Performance Tracking

- Optimisation timelines for each region shows that some zones of the component arrive at good solutions quickly, while others will need more iterations to fully converge.
- Illustrates that there may be a **benefit to computing an optimal set of ADR parameters for each region**, rather than trying to optimise a single ADR for the whole component…
- The performance map could be used to inform where the automated analysis is trustworthy and where the analysis should be left for a human operator.







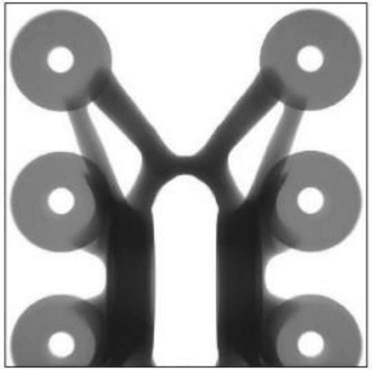


Task 3: CNN ADR Optimisation on Simulated DR Data

Task 3: CNN ADR Optimisation on Simulated DR Data

Applying Methodology to 2D Datasets

- Task 3 focused on whether these approaches could be applied to **2D X-ray data**
- For variety, the conventional image processing ADR was swapped for a **deep learning Convolutional Neural Network** (CNN) framework (Mask R-CNN)
- In this case, three poses for the radiographic projections were chosen.



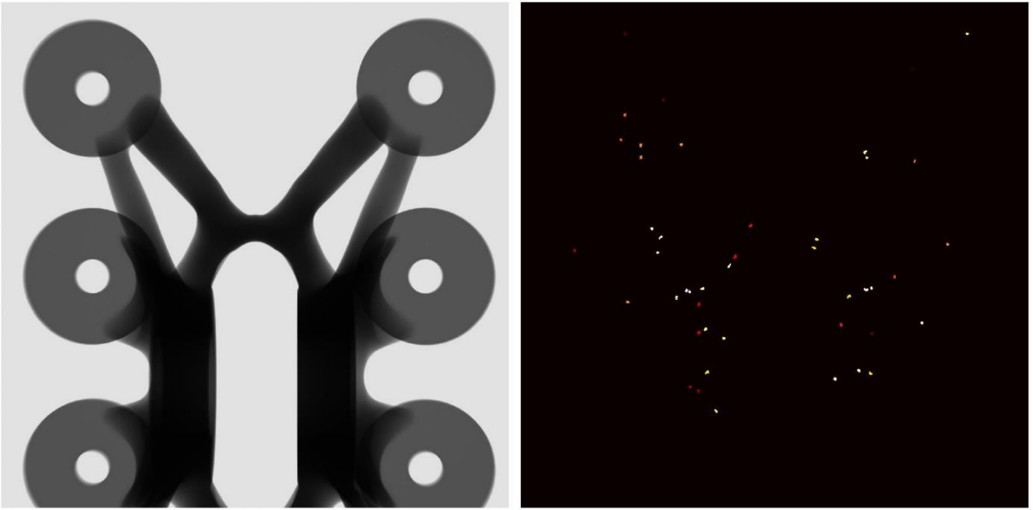
Top / Front Top / Left Top / Right

1 Dataset = 3 Projections

Task 3: CNN ADR Optimisation on Simulated DR Data

Synthetic Data Generation

- As with the XCT data, defective datasets were generated along with corresponding label mask images.
- In this case, a variety of randomised defect shapes and sizes were used (did not want CNN to only look for circular indications)
- 13600 defects generated across 272 datasets (272 × 3 = 816 images), average of 50 defects per dataset.

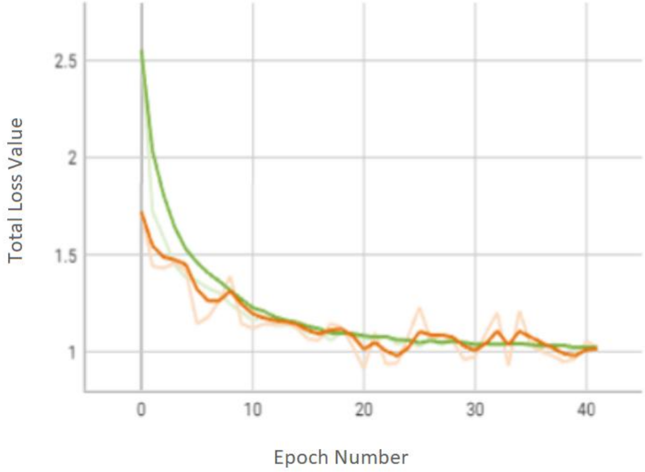


X-ray projection with several defects and corresponding defect label mask.

Task 3: CNN ADR Optimisation on Simulated DR Data

Mask R-CNN

- In order to use the Mask R-CNN framework, some adjustments were required to be made, which are often overlooked:
  - Move from 3-channel 8-bit RGB input to single channel 16-bit monochromatic input
  - Needed to define objective / loss function to be F_2 score rather than accuracy
- Trained on 65% of datasets, leaving 35% for test datasets.
- After 40 epochs of training, the ADR perform**ance appeared to converge**.



Timeline of Optimisation

Task 3: CNN ADR Optimisation on Simulated DR Data

CNN Performance

- The optimised CNN ADR performance follows what would be expected intuitively :
    - True positives in regions with low material path lengths and good contrast.
    - False positives and negatives in regions where material path lengths are longer and contrast is poorer.
- Each of the 3 projections in the dataset will have different rates of TP, FN and FP across the dataset.

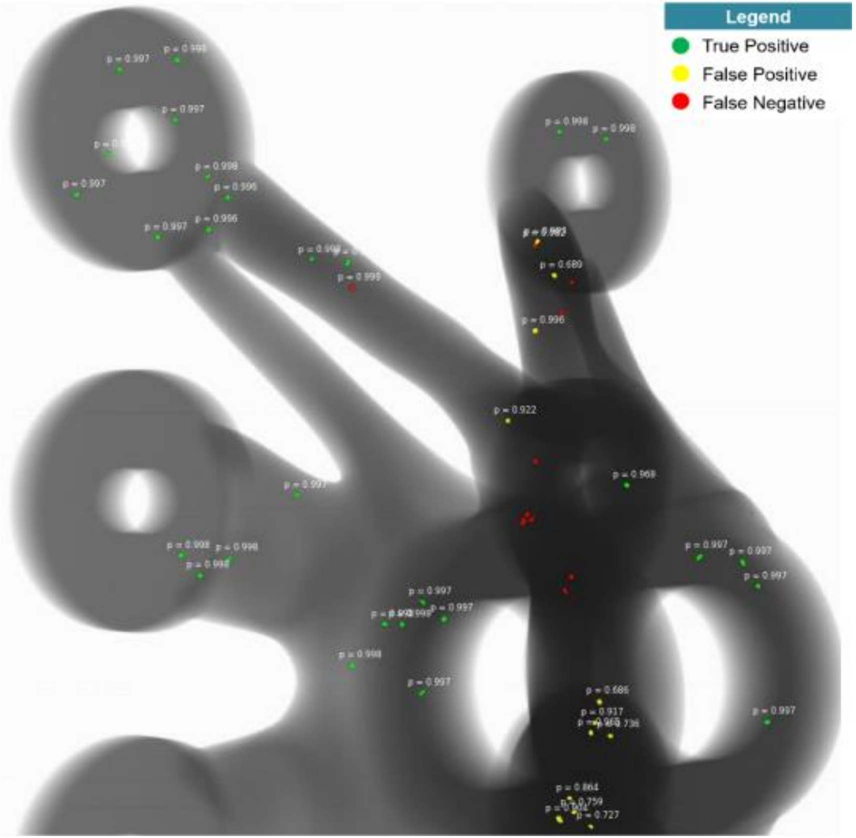
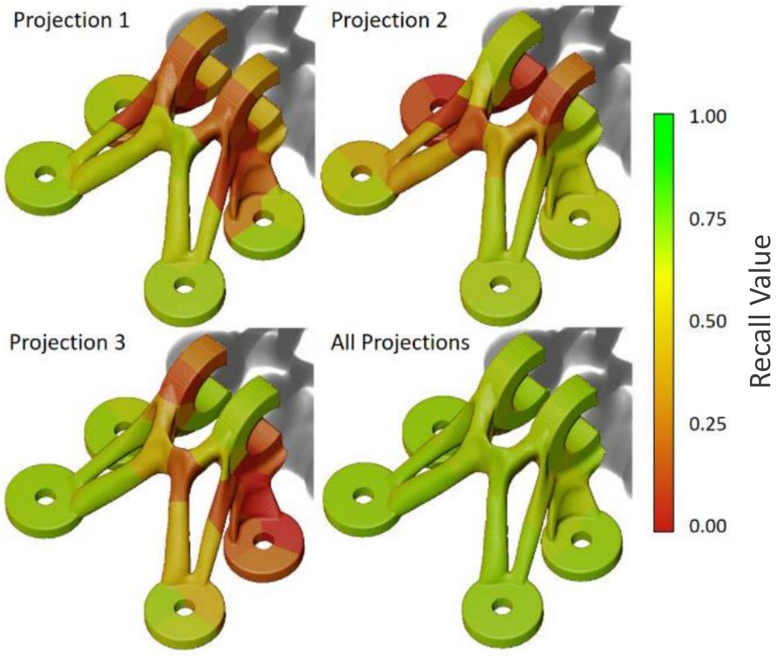


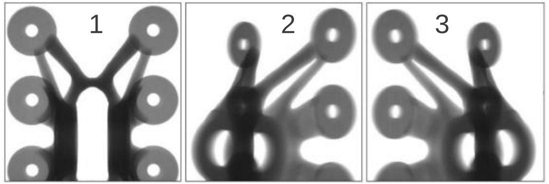
Image mapping TP, FP and FN for an example projection

Task 3: CNN ADR Optimisation on Simulated DR Data

CNN Performance



ADR Performance Map



Component poses used in projection set.

Projection	Recall	Precision	F2
1	0.659	0.734	0.672
2	0.792	0.740	0.780
3	0.802	0.760	0.792
All	0.924	0.638	0.848
2 and 3	0.857	0.709	0.823

Recall, projection and F2 score for each projection.

- Using the reserved test dataset (35%), were able to map performance for each projection individually, and all projections combined.
- Idea is that defect might be missed in projection 1, but detected in projection 2 (for example).
- Can assess **quantitatively** which projections offer the most **unique information** to the projection set.
- For example, can see that projection 1 provides less unique information, and can quantify the impact on expected recall, precision and F2 of only using projections 2 and 3.

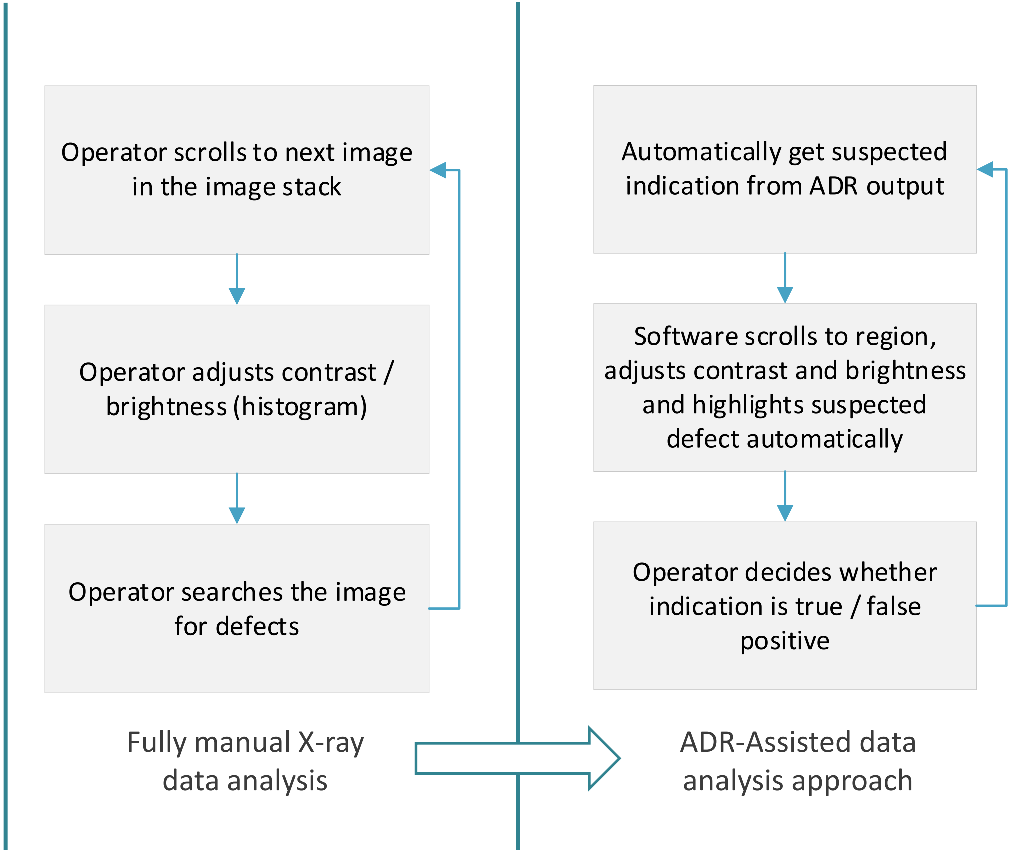
Considerations for ADR in Production Environments

Considerations for ADR in Production Environments

- For automated approaches to be trusted for production inspection:
  - The recall value of any ADR will need to be pushed sufficiently high (100% for critical applications) before they can be used. In many cases, **99% recall will not be good enough**!
  - 100% recall may still not be good enough, unless the training dataset is sufficiently robust to avoid unusual 'edge' cases.
  - While this project leant heavily on synthetic data, it's not clear how accurate a simulation needs to be to make it reliable for genuine experimental data. When comparing synthetic data and genuine data, there is no "data realism" metric. **To sidestep this issue, could use 100% genuine data for testing dataset.**
  - Decisions made by the automated procedure will need to be fully traceable (conventional image processing may be more transparent).
  - May need to combine multiple ADR models into a single automated procedure to enhance AI robustness.
- There are also ethical concerns around AI:
  - A starting point may be Rolls-Royce's Alethia Framework for ethical AI. This document describes 32 principles which should be considered.
  - For an algorithm to achieve 100% recall, high false positive rate is expected, at least to begin with. For this reason, we expect ADR tools will be used to simplify datasets to specific regions rather than remove the need for human operators.

Considerations for ADR in Production Environments

- A potential route to using an ADR tool in a 100% recall application is illustrated…



Conclusions and Next Steps

Conclusions and Next Steps

- The MTC developed a **simulation-driven solution to optimise and validate Automatic Defect Recognition** (ADR) approach for X-ray inspection;
- Both **conventional image processing** ADR and **deep-learning** approaches were considered.
- Applied to both **2D and 3D XCT X-ray** inspection.
- Method taken allows comprehensive assessment of ADR performance, **mapping quantitative performance onto regions of the component**. Were also able to quantify value of individual projections in a projection set.

Next steps:

- Rigorous validation:
  - Can an algorithm trained on synthetic data be trusted with genuine production data?
  - How accurate is the simulation?... Metrics for validation needed.
- Include scan parameter optimisation into ADR optimisation. ADR short comings can be fixed by modifying scan parameters (or vice-versa).
- Applications to other inspection modalities (Ultrasonic testing?)

Thanks for listening

robert.culver@the-mtc.org