DEVELOPMENT OF A HYBRID DATA-DRIVEN / MODEL-BASED REGULARIZATION METHOD FOR X-RAY CT-METROLOGY

Project Identifier: G0C2423N
Principle Investigator: Wim Dewulf
Abstract
Dimensional quality control of hidden or hardly accessible features of high added-value products, produced by e.g. additive manufacturing, is becoming feasible using X-ray Computed Tomography (XCT). Mathematically, X-ray CT reconstruction is however an ill-posed inverse problem when only corrupted or incomplete data is available (i.e., there is no unique solution varying continuously with the data due to model and data uncertainties), then the output dimensional measurements are subject to errors. Regularized model-based or data-driven strategies are used to overcome such ill-posedness, exploiting prior information or specific properties on desired reconstructions. The selection, or design of a new regularizer is however strongly problem-dependent. A methodology for both deciding the strategy and amount of minimal regularization needed to achieve a desired quality; and for quantifying the associated uncertainty for XCT-metrology has not been established yet. This forms the main objective of this project.

DATA MANAGEMENT PLAN

1. RESEARCH DATA SUMMARY

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

We will generate new data obtained with XCT scanners available at the Manufacturing Metrology Group of KU Leuven. The measured objects will serve for validation purposes of the algorithms developed in the project; therefore, no personal or confidential data will be generated.
One standard XCT scan generates 2000*2000*3172 pixels with 16-bit of numerical precision and 2000*2000*2000 voxels for the processed volume with the same precision. At least three processed volumes will be generated with three different algorithms. This means around 24 + 3 * 16 = 72 GB per experiment.
We estimate to perform around 400 experiments (100 per year) therefore 30TB of storage will be needed. Computational codes / manuscripts will also be produced.

Experimental 3D image data:

- raw 16-bit data obtained from XCT scans (24 GB each, .vol)
- Procesed 16-bit data computed from the XCT data (16 GB each, .vol / .npy)

Software / codes:

- Python + CUDA + C++ computational scripts (.py, .cu, .cpp) : 1GB
- Latex files (tutorials, articles, presentations, documentation) (.tex, .bib, .pdf) : 10GB

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.

Research regards algorithm development for X-ray CT, which is on the list of potential dual-use items in the sense of Regulation 428/2009. Since there is no actual military purpose, nor collaboration with a sensitive partner/funder/end user, a confirmation and positive advice was obtained from the KU Leuven Ethics Committee on Dual Use, Military Use and Misuse (EC DMM).

Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.

No

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.

The obtained algorithms on xCT data processing could be exploited commercially. E.g., XCT technology companies could incorporate the software to their machines, or even companies using XCT for inspection e,g, quality control / metrology, could apply the produced algorithms for better / faster results.

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/Data transfer agreements/ research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

No

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

No

2. DOCUMENTATION AND METADATA

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g., in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, Codebook.tsv etc. where this information is recorded).

XCT experimental data: Metadata will be saved automatically including all the parameters used in that measurement (geometry configuration, kEV, power, scanning time, ...) with a brief description of the object measured (dimensions, materials if known). This can be saved in a .log file

Processed data will be uniquely identified with the corresponding XCT data and annotated with the algorithm used to obtain it and its parameters. Also saved as a .log files

Codes will include README files explaining installation and how-to-use. In addition, .pdf files will be created with complete documentation and examples.

Will a metadata standard be used to make it easier to find and reuse the data? If so, please specify (where appropriate per dataset or data type) which metadata standard will be used. If not, please specify (where appropriate per dataset or data type) which metadata will be created to make the data easier to find and reuse.

Yes. XCT metadata will be saved following the same conventions of .xteck files from NIKON XCT machines.

3. DATA STORAGE & BACK-UP DURING THE RESEARCH PROJECT

Where will the data be stored?

During the execution of the project, researchers will make use of the research group's NAS storage capacity (with back-up) for flexible daily access. All experimental and processed data will be annotated with metadata, which will be saved as a readme file with the dataset. When datasets are no longer used on a daily basis (e.g. at the end of a task or work package), the datasets will be archived on KU Leuven Large Volume Storage until at least 5 years after the end of the research.
All software / codes will be kept and maintained regularly on KU Leuven Gitlab repository. Manuscripts will be conserved on Overleaf for the latex codes, and on the Arxiv repository / KU Leuven Onedrive cloud server for the final pdfs files.

How will the data be backed up?

Regularly, using different disk / partitions

Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely.
If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.

No. Costs for extra storage have been budgeted for the capacity to be enough.

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

WIth KU Leuven network security and policies for digital data at KU Leuven servers.

What are the expected costs for data storage and backup during the research project? How will these costs be covered?

2.000 euro per year on project budget

4. DATA PRESERVATION AFTER THE END OF THE RESEARCH PROJECT

Which data will be retained for at least five years (or longer, in agreement with other retention policies that are applicable) after the end of the project? In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

All data described above that was at the basis of (internal or public) reports or publications.

Where will these data be archived (stored and curated for the long-term)?

On KU Leuven Large Volume Storage.

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?

Costs were budgeted.

5. DATA SHARING AND REUSE

Will the data (or part of the data) be made available for reuse after/during the project? In the comment section please explain per dataset or data type which data will be made available.

Yes, in an Open Access repository. Codes on KU Leuven Gitlab for the first versions. Manuscripts on the Arxiv repository.

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)? Please explain in the comment section per dataset or data type where appropriate.

No

Where will the data be made available? If already known, please provide a repository per dataset or data type.

Codes on KU Leuven Gitlab for the first versions. Manuscripts on the Arxiv repository.

When will the data be made available?

Upon publication of research results

Which data usage licenses are you going to provide? If none, please explain why.

No license is pertinent to the XCT data we will produce. Codes will have GPLv3 license for the initial versions.

Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.

No

What are the expected costs for data sharing? How will these costs be covered?

Arxiv is a free repository, KU Leuven Gitlab license will be used. No extra costs are expected here.

6. RESPONSIBILITIES

Who will manage data documentation and metadata during the research project?

Dr. Patricio Guerrero

Who will manage data storage and backup during the research project?

Dr. Patricio Guerrero

Who will manage data preservation and sharing?

Prof. Wim Dewulf and Dr. Patricio Guerrero

Who will update and implement this DMP?

Prof. Wim Dewulf and Dr. Patricio Guerrero

GDPR

Have you registered personal data processing activities for this project?

No

DPIA

Have you performed a DPIA for the personal data processing activities for this project?

No