Advanced nonimaging optical design via physics-informed machine learning

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.

Dataset name / ID	Description	New or reuse	Digital or Physical data	Data Type	File format	Data volume	Physical volume
		Indicate: N(ew data) or E(xisting data)	Indicate: D(igital) or P(hysical)	Indicate: **A**udiovisual **I**mages **S**ound **N**umerical **T**extual **M**odel **SO**ftware Other (specify)		Indicate: <1GB <100GB <1TB <5TB >5TB NA	
1. Python code and data							
1.1. Design models	Machine learning models for predicting freeform nonimaging surfaces.	N	D	M, SO	.pth, .py, .ipynb	<100GB	NA
1.2. Raytracing models	Machine learning models for modeling phase space raytracing.	N	D	M, SO	.pth, .py, .ipynb	<100GB	NA
1.3. Dataset generation	Code for generating training and validation data for machine learning models.	N	D	SO	.ipynb, .py	<1GB	NA
1.4. Generated datasets	Generated 4d phase space datasets for model training.	N	D	N	.npy	<1TB	NA
2. Experimental results							
2.1. SLM measurements	Hyperspectral images of the resulting irradiance distributions in the SLM setup.	N	D	I	.raw	<100GB	NA

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

Question not answered.

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, refer to specific datasets or data types when appropriate and provide the relevant ethical approval number.

- No

Will you process personal data? If so, please refer to specific datasets or data types when appropriate and provide the KU Leuven or UZ Leuven privacy register number (G or S number).

- 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.

- Yes

The machine learning models and all involved code/datasets could be directly exploited to rapidly generate nonimaging freeform surfaces, for e.g. commercial illumination design or solar concentrators.

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material or 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

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).

All generated design code is accompanied by comments describing the reasoning of each line or sequence of code. The generated models and datasets will also be accompanied by a .txt file containing information about size, dimensions and usability.
The SLM images are accompanied by relevant code to process these hyperspectral images, and will be further detailed/described in a manuscript which is currently in the works.

Will a metadata standard be used to make it easier to find and reuse the data?
If so, please specify which metadata standard will be used.

If not, please specify which metadata will be created to make the data easier to find and reuse.

- No

Data Storage & Back-up during the Research Project

Where will the data be stored?

- Shared network drive (J-drive)

The data will be stored on the online KUL OneDrive (onedrive.kuleuven.be).

How will the data be backed up?

- Standard back-up provided by KU Leuven ICTS for my storage solution
- Personal back-ups I make (specify below)

Aside from the automatic OneDrive update (windows synchronization), I will also personally upload finalized code, models and datasets onto the team folder of our research team.

Is there currently sufficient storage & backup capacity during the project?

If no or insufficient storage or backup capacities are available, explain how this will be taken care of.

- Yes

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

All data is on the KUL OneDrive, thus requiring KUL authorization. Furthermore, access is only granted to persons with the specific permissions, i.e. the members of our research team.

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

No expected costs: all storage methods are standard for KUL staff members.

Data Preservation after the end of the Research Project

Which data will be retained for 10 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 will be preserved for 10 years according to KU Leuven RDM policy

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

- Other (specify below)

KUL OneDrive.

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

No expected costs: all storage methods are standard for KUL staff members.

Data Sharing and Reuse

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

- Yes, as restricted data (upon approval, or institutional access only)

All the data (code, models, datasets) will be on the OneDrive of KUL, and access can be granted through permission.

If access is restricted, please specify who will be able to access the data and under what conditions.

We would like to (for now) restrict access to the data, since we plan to bulid further on this in the coming years. In the future, once the full prediction framework is realized, we will make it public, e.g. on GitHub.

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 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.

- Other data repository (specify below)

KUL OneDrive.

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.

- Data Transfer Agreement (restricted data)

Do you intend to add a persistent identifier (PID) to your dataset(s), e.g. a DOI or accession number? If already available, please provide it here.

- No

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

No costs.

Responsibilities

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

Jeroen Cerpentier (main researcher).

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

Jeroen Cerpentier (main researcher).

Who will manage data preservation and sharing?

Jeroen Cerpentier (main researcher).

Who will update and implement this DMP?

Jeroen Cerpentier (main researcher).

Created using DMPonline.be. Last modified 25 February 2025

5 of 5