

MSc AI Dissertation Project Proposal Form

Student number	201602374	Date	13/02/2025
Student name	Joshua Hellewell		
Supervisor name	Chandrasekhar Kambhampati		
Project title:	Enabling Sample Efficient Design Space Exploration through Machine Learning		
Project description – background, aims, objectives, research question(s), brief reflection on novelty/originality			

Engineering is propelled by effective design space exploration and optimisation. As is common, numerical tools are used to substitute expensive physical testing; for example, computational fluid dynamics can, by solving physical numerical equations, provide engineers with flow field information such as surface pressure fields and velocity fields. Such simulations are often computationally expensive and not optimised for fast, iterative design exploration, which can be partly attributed to the input geometries (unstructured mesh files, point clouds, etc.) not lending themselves to simple parameterisation that would enable an interpretable representation of complex features and geometries. In the absence of such a concise description of how shapes are related, design space exploration and optimisation is constrained to a high-dimensional landscape, inheriting computational challenges and falling victim to the curse of dimensionality, which suppresses efficient design exploration. Complex shapes, such as those found in typical aerodynamic applications, can be composed of millions of elements described by vertices, edges, and faces, and therefore exploring all combinations is computationally infeasible. Having a low-dimensional representation of complex shapes that concisely represents the defining features and nuances enables optimisation to take place over a lower-dimensional design space which better suits optimisation approaches such as genetic algorithm (heuristic approach), and gradient based methods. Enabling more efficient design space exploration and optimisation has the potential to significantly reduce the number of computationally costly simulations needed to realise the most optimal design amongst an initially highly complex and vast search landscape.

Despite many industry problems residing in the 3D space, the computational burden and data availability introduce unnecessary challenges that detract from the underlying concepts to be investigated in this research, and for that reason, this research will focus on 2D geometries and a simplified optimisation problem. Though the underlying principles are agnostic of the specific optimisation problem.

Objectives:

- #1 Investigate techniques and approaches for representing geometry in a low dimensional form to enable sample efficient design space optimisation and carry out a comprehensive literature review of existing methods, to both the representation problem, and design space exploration and optimisation.
- #2 Develop approaches for representing 2D geometries in a lower dimensional representation that can be used directly as part of a geometric shape optimisation problem and investigate the structure and smoothness of the learned geometry representations and their suitability for use in different optimisation approaches.
- #3 The research will assess which approaches to geometry representation provide the most interpretable and controllable representation that can be harnessed to conduct targeted design modification and optimisation for the selected optimisation problem.

In parallel to investigating methods for compressing the high-dimensional complexity of geometries, we will deploy the learned representations to a simple 2D optimisation problem, area to perimeter ratio – which may be perceived as the 2D equivalent of surface area to volume ratio, which is often used as a parameter to indicate heat losses. This approach could be extended to other industry problems, such as optimising the lift to drag ratio for an air foil. The quantity of interest to be optimised will be denoted as compactness, and defined as:

$$Compactness = \frac{Perimeter}{Area}$$

For example, for a circle, this is trivial as a circle's perimeter and area can be described by a single parameter r – therefore the 'design space' spans a continuum over the parameter r only (i.e., it is a 1D problem). The equation below for compactness of a circle allows us to make statements such as 'if we increase radius, r , the compactness will decrease' as the compactness is inversely proportional to the radius. This allows a targeted modification of the design by adjusting just r (assuming we constrain ourselves to circles).

$$\text{Compactness} = \frac{2 \pi r}{\pi r^2} = 2 \frac{\pi}{r}$$

This is a simple example, however, some arbitrarily complex shape with a combination of smooth edges and straight sides with no constraint on being a specific shape group (quadrilateral, triangle etc) is not easily parameterised by a small number of variables and without significant constraints, the design space is complex. Imagine if we represented arbitrary shapes as 100 connected nodes, coordinates and we wanted to explore the design space that provided a shape that was optimised for a specific objective function. The design space in this case, assuming only a 2D canvas and no additional constraints, would include every combination of x,y coordinates for every 100 points. Often for complex shapes, 100 points would be insufficient to describe the intricate details and may easily extend to millions of nodal positions for 3D, complex bodies. Additionally, imagine adding freedom to explore the connectivity of how the nodes interconnect. This would introduce an additional dimension to the design space.

For this reason, the above 'compactness' cannot be easily interpreted in the same way as the circle statement for complex shapes. Another example, but this time representing arbitrary shapes as binary images, where the structure of shapes is represented by 1s and 0 denoting presence of shape material, it is not difficult to prove how over even a small canvas of 128x128 pixels, the design space has $2^{128 \times 128}$ possible combinations of pixels. Extending this into 3D ($2^{128 \times 128 \times 128}$) dimensions become computationally unfeasible.

The motivation of this research is to enable a concise descriptor for the way in which arbitrary shapes relate to each other by realisation of an intermediary, low-dimensional representation learned from data that provides latent variables that describe a smooth transition between different shapes. In doing so, we provide a means to reduce the high-dimensional design space which better lends itself to optimisation approaches such as heuristic approaches, like the Genetic Algorithm.

Exploration through this low-dimensional latent design space is also expected to be more efficient. Further, representing complex design spaces in just 2-dimensions would enable human-interpretable visualisation of how exploration through a continuous latent space relates to the quantity of interest to be optimised for.

Methodology – rationale, data selection and collection, recruitment of participants, analytical process

The project will be conducted by firstly identifying or producing a comprehensive dataset of 2D shapes that will be used to represent the geometry space, albeit simplified into the 2D domain to avoid computational limitations associated with data formats such as voxel grids, point clouds and mesh grids. Optionally, the UIUC 2D aerofoil dataset may be used, however, this dataset is smaller and spans a more similar library of geometries (i.e., all aerofoils). For 2D shapes, the representation of original data can be structured as pixels like in images – the sparse connectivity of such data makes the project appropriately tractable vs. computationally intensive structures such as point clouds, mesh and voxels. However, to more closely align with real-world engineering representations, we will also explore 2D shape representations as nodes, each comprising of an x,y position on a 2D grid.

All data will be stored locally and processed using Python. Specifically, for working with ML models the data will be stored as .h5 files for efficient storage and access. Any neural network models will be built using Metas' PyTorch architecture, and training using the CUDA implementation for efficient, parallelised model training. The 2D shape data will be represented as images to enable a structured representation of the shapes. These will be stored as numpy arrays, or tensors depending on selected models.

A literature review will be carried out that investigates techniques for representing high dimensionality data a low dimensional form in addition to methods for design exploration and optimisation methods that can be used for engineering design.

Theoretical framework and intended tools/technologies

Generative models may include Auto-encoders, Variational Auto-encoders, Denoising Auto-encoders, Sparse Auto-encoders, Contractive Auto-encoders. Generative Adversarial Networks. Implicit Neural Representations, Diffusion models.

Design space exploration may investigate methods such as: grid search, monte Carlo methods, Markov chain monte Carlo with optimisation techniques such as gradient-based optimisation and genetic algorithms (Gupta et al, 2007).

Python, PyTorch, Cuda, JAX, NVIDIA GPU, Viper HPC

Key literature identified for literature review (include references)

- Park, J. J., Florence, P., Straub, J., Newcombe, R., & Lovegrove, S. (2019). DeepSDF: Learning continuous signed distance functions for shape representation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2019-June*.
<https://doi.org/10.1109/CVPR.2019.00025>
- Zhu, Z., Wang, X., Bai, S., Yao, C., & Bai, X. (2016). Deep Learning Representation using Autoencoder for 3D Shape Retrieval. *Neurocomputing, 204*.
<https://doi.org/10.1016/j.neucom.2015.08.127>
- Li, P., Pei, Y., & Li, J. (2023). A comprehensive survey on design and application of autoencoder in deep learning. In *Applied Soft Computing* (Vol. 138). <https://doi.org/10.1016/j.asoc.2023.110176>
- Sun, JM., Wu, T. & Gao, L. Recent advances in implicit representation-based 3D shape generation. *Vis. Intell.* 2, 9 (2024). <https://doi.org/10.1007/s44267-024-00042->
- Sharp, N., Attai, S., Crane, K., & Ovsjanikov, M. (2022). DiffusionNet: Discretization Agnostic Learning on Surfaces. *ACM Transactions on Graphics, 41*(3). <https://doi.org/10.1145/3507905>
- Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., & Xiao, J. (2015). 3D ShapeNets: A deep representation for volumetric shapes. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 07-12-June-2015, 1912–1920*.
<https://doi.org/10.1109/CVPR.2015.7298801>
- Qi, C. R., Su, H., Mo, K., & Guibas, L. J. (2017). PointNet: Deep learning on point sets for 3D classification and segmentation. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017-January, 77–85*. <https://doi.org/10.1109/CVPR.2017.16>
- Tripp, A., Daxberger, E., & Hernández-Lobato, J. M. (2020). Sample-efficient optimization in the latent space of deep generative models via weighted retraining. *Advances in Neural Information Processing Systems, 2020-December*.
- Zhen Wei, Edouard R. Dufour, Colin Pelletier, Pascal Fua and Michaël Bauerheim (2024). Diffairfoil: An efficient novel airfoil sampler based on latent space diffusion model for aerodynamic shape optimization. *AIAA AVIATION FORUM AND ASCEND 2024*, July 2024.
- Zheng, L., Karapiperis, K., Kumar, S., & Kochmann, D. M. (2023). Unifying the design space and optimizing linear and nonlinear truss metamaterials by generative modeling. *Nature Communications, 14*(1). <https://doi.org/10.1038/s41467-023-42068-x>
- Long, L., Cartis, C., & Shustin, P. F. (2024). Dimensionality Reduction Techniques for Global Bayesian Optimisation. *arXiv preprint arXiv:2412.09183*.
- Wu, J., Zhang, C., Xue, T., Freeman, W. T., & Tenenbaum, J. B. (2016). Learning a probabilistic latent space of object shapes via 3D generative-adversarial modeling. *Advances in Neural Information Processing Systems*.
- Siivola, E., Paleyes, A., González, J., & Vehtari, A. (2021). Good practices for Bayesian optimization of high dimensional structured spaces. In *Applied AI Letters* (Vol. 2, Issue 2).
<https://doi.org/10.1002/ail2.24>
- Danhaive, R., & Mueller, C. T. (2021). Design subspace learning: Structural design space exploration using performance-conditioned generative modeling. *Automation in Construction, 127*.
<https://doi.org/10.1016/j.autcon.2021.103664>

Gupta, S., Tiwari, R., & Nair, S. B. (2007). Multi-objective design optimisation of rolling bearings using genetic algorithms. *Mechanism and Machine Theory*, 42(10).
<https://doi.org/10.1016/j.mechmachtheory.2006.10.002>

Sun, J., Frazer, J. H., & Mingxi, T. (2007). Shape optimisation using evolutionary techniques in product design. *Computers and Industrial Engineering*, 53(2).
<https://doi.org/10.1016/j.cie.2007.06.010>

Demo, N., Ortali, G., Gustin, G., Rozza, G., & Lavini, G. (2021). An efficient computational framework for naval shape design and optimization problems by means of data-driven reduced order modeling techniques. *Bolletino Dell Unione Matematica Italiana*, 14(1).
<https://doi.org/10.1007/s40574-020-00263-4>

Laurenceau, J., Meaux, M., Montagnac, M., & Sagaut, P. (2010). Comparison of gradient-based and gradient-enhanced response-surface-based optimizers. *AIAA Journal*, 48(5).
<https://doi.org/10.2514/1.45331>

Catalani, G., Agarwal, S., Bertrand, X. *et al.* Neural fields for rapid aircraft aerodynamics simulations. *Sci Rep* **14**, 25496 (2024). <https://doi.org/10.1038/s41598-024-76983-w>

Kang, E., Jackson, E., & Schulte, W. (2011). An approach for effective design space exploration. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6662 LNCS. https://doi.org/10.1007/978-3-642-21292-5_3

Selig, Michael S. (1996). UIUC airfoil data site. Urbana, Ill. :Department of Aeronautical and Astronautical Engineering University of Illinois at Urbana-Champaign,

UIUC. 2025. UIUC Applied Aerodynamic Group – UIUC Airfoil Coordinates Database. Retrieved March 4, 2025 from https://m-selig.ae.illinois.edu/ads/coord_database.html

Industrial partner (if appropriate)

Rolls-Royce,

Mark Hobbs – Mark.Hobbs@Rolls-Royce.com | Technical Specialist Future Methods

Project management

Table: Project timeline and key outputs (expand table as needed)

Week	Activity
Week 1 – w/c 10.03.25	Submit Proposal
Week 1	Ethics approval form.
Week 2-4	Review literature on geometry representation
Week 4-6	Review literature for shape optimisation in low-dimensional space.
Week 2-6	Identify main gaps and establish problem statement(s)
Week 5	Milestone: Produced literature matrix (Objective #1)
Week 5	Create MSc thesis report structure.
Week 5	Define methodology and research approach.
Week 4-7	Write literature review
Week 6-7	Write up methodology
Week 7	Milestone: Literature review draft completed (Objective #1)
Week 2-3	Collect or generate 2D shape dataset.
Week 3-4	Pre-process data (resizing, augmentation, normalisation).
Week 4-5	Define evaluation metric for low-dimensional representation.
Week 5	Milestone: 2D shape dataset & processed (Objective #2)
Week 6	Create GitHub project structure for model comparison.
Week 6-9	Create geometry representation models/methods in Python.
Week 8-10	Train & Validate techniques to obtain low-dimensional geometry embeddings.
Week 8-10	Assess smoothness, interpretability and effectiveness of representations.
Week 10	Milestone: Geometry Representation techniques tested (Objective #2 & #3)
Week 9-11	Create optimisation technique scripts.
Week 11-13	Integrate learned representations into compactness optimisation framework.
Week 13-14	Evaluate effectiveness of different representations for optimisation task.
Week 14	Milestone: Integrated geometry representation and design exploration (Objective #3)
Week 14-18	Draft final dissertation, including results & discussion.
Week 19-21	Review & edit final report.
Week 21 – w/c 11.08.25	Milestone: Dissertation submitted.

***See table on page 9 for full project plan.

Resources and Research Data Management Plan (Describe the resources required and the data you expect to acquire or generate during this research project, how you will manage, describe, analyse, and store the data and what mechanisms you will use to share and preserve your data.)

Project will be managed in a GitHub repository:

<https://github.com/jhell1717/MScProject2025/tree/Planning>

All models will be updated and shared via this repository with the final selected model for geometry encoding being packaged into a Python package that a user can deploy on their own local machine through a pip install from the GitHub repo.

Many industry-relevant problems exhibit complex failure behaviour influenced by multiple factors that may not be apparent or comprehensible. Consequently, to avoid distracting from the underlying principles, the data will be simplified to two-dimensional geometric shapes. Adding additional complexity cannot be reasonably justified in accordance with the objectives of this research and risks obscuring the predictive accuracy and performance of selected models in isolation of challenge-problem induced error sources.

Planned outputs/publications/research datasets/impact/dissemination

The output from this research will be a comprehensive study of methods that enable sample efficient design space exploration.

A comparison of different methods for reducing high-dimensional challenges to low dimensional sample space, deployment to a simple optimisation will be performed. This is likely to be compactness of a 2D shape, as a pseudo metric for heat losses in physical components.

A literature review of existing approaches will be performed, along with a substantial research paper documenting methodologies, results, discussions, limitations, conclusions and recommendations for methods that should be extended to industry specific use-cases.

If successful, I undertake to carry out the research according to the University's Ethics code of Practice. I realise that I will not proceed into data collection without an ethical approval in place

(Applicant's signature required)



Date and signature of Supervisor approval

	Activity	March				April				May				June				July				August			
		Week-1	Week-2	Week-3	Week-4	Week-5	Week-6	Week-7	Week-8	Week-9	Week-10	Week-11	Week-12	Week-13	Week-14	Week-15	Week-16	Week-17	Week-18	Week-19	Week-20	Week-21	Week-22	Week-23	
Admin	Submit Proposal																								
	Ethics approval form																								
	Review literature on geometry representation																								
	Review literature for shape optimisation in low-dimensional space.																								
Research & Planning	Identify main gaps and establish problem statement.																								
	Milestone: Produced literature matrix.																								
	Create MSc thesis report structure.																								
	Define methodology and research approach.																								
	Write literature review																								
	Write up methodology																								
Data Collection & Pre-processing	Milestone: Literature review draft completed.																								
	Collect or generate 2D shape dataset.																								
	Pre-process data (resizing, augmentation, normalisation).																								
	Define evaluation metric for low-dimensional representation.																								
Model Development & Implementation	Milestone: 2D shape dataset & processed.																								
	Create GitHub project structure for model comparison.																								
	Create geometry representation models/methods in Python.																								
	Train & Validate techniques to obtain low-dimensional geometry embeddings.																								
	Assess smoothness, interpretability and effectiveness of representations.																								
	Milestone: Geometry Representation techniques tested.																								
Optimisation Model Integration	Create optimisation technique scripts.																								
	Integrate learned representations into compactness optimisation framework.																								
	Evaluate effectiveness of different representations for optimisation task.																								
Report Writing & Submission	Milestone: Integrated geometry representation and design exploration.																								
	Draft final dissertation, including results & discussion.																								
	Review & edit final report.																								
	Milestone: Dissertation submitted.																								