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 computationally expensive and not optimised for fast, iterative design exploration, which can be partly attributed to the input geometry complexities (unstructured mesh files, point clouds, etc.) inhibiting 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 methods such as grid search, Monte-Carlo methods, or gradient-based optimisation to be performed more efficiently. This research aims to identify, investigate, and compare different approaches to addressing this high-dimensional challenge.

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.

Objectives:

- Investigate techniques and approaches for representing geometry in a low dimensional form to enable sample efficient design space exploration and optimisation and carry out a literature review collating existing methods for geometry representation, and design space exploration.
- Develop approaches for representing a 2D geometries in in a lower dimensional space 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, for example gradient-based, genetic algorithms, and Bayesian optimisation.
- 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.

Generative models, such as encoder-decoders networks and generative adversarial networks, are finding use cases across engineering. Despite auto-encoders emerging in the 1980s, their applications across engineering are increasing to enable conventionally high-complexity and compute-resource-hungry processes to be approached in an intermediary low-dimensional search space. Additionally, design space exploration and optimisation are a critical enabler to design innovation, and doing so in a cost-efficient manner is sought in industry whereby conventional numerical simulations are intractably resource hungry. In addition 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; however, this approach could be extended to other industry problems,

such as optimising the lift to drag coefficient for an airfoil. The quantify of interest to be optimised will be labelled as compactness, and described by:

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

For example, for a circle, this is trivial as a circle can be described by a single parameter r – therefore the 'design space' is a single variable over r . The below equation for compactness of a circle allows us to make statements such as 'if we increase radius, r, the compactness will decrease'. This allows a targeted modification of the design by adjusting just r (assuming we constrain ourselves to circles).

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

In contrast, 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 highly complex. For this reason, the above 'compactness' cannot be easily interpreted in the same way as the circle statement. The motivation of this research is to enable a concise descriptor of the way in which arbitrary shapes relate to each other by realisation of an intermediatory, low-dimensional representation learned from data that provides latent variables, i.e., tuneable variables, that describe a smooth transition between arbitrarily complex and different shapes. In doing so, we provide a means to reduce the high-dimensional design space which better lends itself to optimisation approaches.

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

The project will be conducted by firstly identifying 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. Add sources of potential data sources.

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.

A literature review will be carried out that will identify techniques for representing high dimensionality data into a low dimensional space, alongside design exploration and optimisation methods that can be used for engineering design. This review will aim to identify state-of-the-art techniques to latent space representation and identify opportunities to build on or leverage technologies for design space exploration.

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.

The optimisation of the quantity of interest may include a search the design space via methods such as: grid search, monte Carlo methods, Markov chain monte Carlo with optimisation techniques such as gradient-based optimisation and genetic algorithms. Include references to this here and in list below

Python, PyTorch, Cuda, JAX, NVIDIA GPU, Viper HPC (get account)

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., Attaiki, 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

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
	Dataset creation
	Literature matrix created
	Write literature review
	Model identification & down selection.
	Construct models & code for geometry representation.
	Construct optimisation methods for quantity of interest
14/08/2025	Submission deadline.

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. ==‘Simplified to see what can be achieved in the space’== – change this.

Planned outputs/publications/research datasets/impact/dissemination

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

Proceeding 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.

The best performing model will be selected and produced as a tool that a practitioner could install locally to aid their design space exploration and optimisation.

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	