School of Electronic Engineering

CB54: Machine Learning Algorithms for EM Wave Scattering Problems

Project Design Plan

Anthony James McElwee

ID Number: 20211330

~~May 2023~~

August 2023

MEng in Electronic and Computer Engineering

Supervised by Dr Conor Brennan

Contents

Project Design Plan 3
Research Question 3
Primary Research Question 3
Secondary Research Questions 3
Project Scope 4
Design Approach 4
Timeline 8
Success Criteria 9
Bibliography 9
Approval 9

Update Note: This plan reflects the position of the project on August 12th 2023. Original text has been left with strike-through convention to illustrate developments in the research as the project has progressed since May 20th 2023.

Project Design Plan

Research Question

The student proposes to sub-divide the research question into two stages to allow for iterative assessment of plausibility of research opportunities and increased agility around the resource and time constraints available.

- The primary stage is aimed at establishing elements of the work of [1]–[5] in a reproducible workflow called SolverEMF2 and creating a supervised regression model called Prescient2DL to test the various hypothesis associated with the Primary Research Question, as explained in the Design Approach section.
- The secondary stage is based around investigating the rudiments of Prescient2DL. Possible aspects for analysis are covered in the various hypothesis associated with the Secondary Research Questions, as explained in the Design Approach section.

Primary Research Question

"Can the manner in which deep learning has been shown to solve two-dimensional, forward electromagnetic scattering problems applied to the problem of predicting EM wave propagation over rural terrain be improved upon?"

As already reported in the Literature Review, various sources have described using deep learning to tackle forward electromagnetic scattering problems, however, to the knowledge of the student, none have provided a public, reproducible, open-source workflow or a model to the research community. The student proposes to approach the integration of the developed deep learning model, Prescient2DL, into SolverEMF2 through the use of Prescient2DL to generate initial guesses for the Krylov Iterative Solver. By establishing the SolverEMF2 workflow, this primary research question will be approached through the investigation of simulations with several segmented statistical hypothesis tests in lieu of qualitive mathematical proofs. Implicitly, the primary aim of the project is to implement a solver with a deep learning model that optimally shifts calculation metrics to towards the lower left corner of the Residual Error versus Iterations/Time graph when solving permittivity contrast source only Volume Electric Field Integral Equations.

Secondary Research Questions

The secondary research questions below are based around trying to expand knowledge around whether the application of deep learning to this domain is fundamentally underpinned by the attributes of the training data set or whether a hierarchical approach to model generation exists. The questions offer a rich range of potentially publishable findings and opportunities to contribute to the field of scientific machine learning.

"Are all models equal in the framework of the performance analysis conducted in the primary research question? Can improvements be made to model development approaches? Can model deployment be improved so that there is an increased opportunity/impact by the model on SolverEMF2 performance?"

Based on the literature review, and echoed in the more recent [6], the models reported in the available literature are either developed in a U-net based architecture or through a Physics Informed Neural Network (PINN). The student proposes to conduct studies into the performance between siloed models to illustrate advantages and changes in performance given a consistent testing environment.

From the insights arising from the model comparison stage, the student believes that these approaches can initially be expanded upon and ensembled. Stemming from this investigation will be the final Prescient2DL model that should demonstrate a more mature application of deep learning to the problem domain. This facilitates the exploration of deep learning attributes and features specifically refined for the problem domain with the possibility of reporting on refined architectures or properties of the model development purpose previously not recorded in journals.

As reflected in the literature review, there is almost no diversity in the deployment strategies for deep learning in this problem domain. The process of emulation is the baseline approach to harnessing deep learning models. The model is expected to resolve problems whose inputs are confined to the parameter ranges associated with the input data used to train the model in the first place. The online-stage of the process thus mirrors the use of Look-Up Tables (LUTs) in conventional problem approaches. The primary stage aims to expand this baseline approach by wrapping Prescient2DL solutions in the Method of Moments framework. In addition to this expansion, the literature review has highlighted sub-algorithmic adjustments to Krylov Iterative Solvers that may yield improved performance metrics. The area of probabilistic numerics may also offer alternative ways that deep learning can be integrated into SolverEMF2.

The benchmarking and characterisation of existing model formulations, the expansion of the model development approaches and the deployment techniques used in SolverEMF2 should facilitate investigation into the generalisability of the deep learning in the domain. The use of Prescient2DL as a basis for transfer learning may also be investigated. These secondary research questions will also be approached through the investigation of several segmented statistical hypothesis tests in lieu of qualitive mathematical proofs.

Project Scope

- The electromagnetic formulation that underpins the area will remain as the two-dimensional, permittivity contrast source only Volume Electric Field Integral Equation as derived at equation (6.5) in [2] or equation (3.86) in [3]. These are scalar integral equations. The derivations in [2] and [3] differ in their uses of time harmonic convention, however, the results from the equations should be equivalent. The Time Harmonic Dependence Convention (j=-i) can be used to translate between them. In [2], the Real-Value Transform with complex notation "i" is used while in [3] the Laplace Transform with complex notation "j" is favoured with the Laplace variable $s = -i\omega$.
- Establish a fresh Github version control repository for code maintenance and sharing.
- Develop the initial SolverEMF2 by migrating VEFIE code from MATLAB to Python and validating the code on a canonical problem. The student has purchased a new laptop with i7-11800H @ 2.3GHz CPU and NVIDIA GeForce RTX 3070 GPU. Access has also been granted to remote DCU GPU rigs and these will be utilised as the project matures.
- Establish and follow the dataset generation pathway as outlined in the Design Approach section below.
- Develop the initial iteration of Prescient2DL using Python deep learning packages.
- There are a number of advanced research areas outlined in the Literature Review, such as the expansion of models via Monte Carlo integration. Due to resource/time constraints, the student proposes to only approach these areas if the potential of the research questions, as outlined in the previous sections, has been exhausted.
- The visualisation and flexibility of the solution information conveyed to the user of SolverEMF2 was initially of much interest to the student, however, since the literature review was conducted the student has discovered a project hosted at (https://github.com/chiuhans111/fdtd-html). That project tackles a time-domain, rather than frequency-domain, problem but the student proposes to focus on the more technical performance metrics and deep learning model development due to the increased probability of generating publishable content.

Design Approach

Code Development Environment

Code will be developed in a manner where all outputs are reproducible through input documentation and seeding. Metrics around the non-DL simulations will be recorded. Time of training and time of inference with initial error (for hybrid models incorporated into conventional methods) will also be recorded.

Code Development Validation – Initial MATLAB migration

Verify code through the solution of canonical toy problems in MATLAB and Python, comparing results to a satisfactory degree of accuracy.

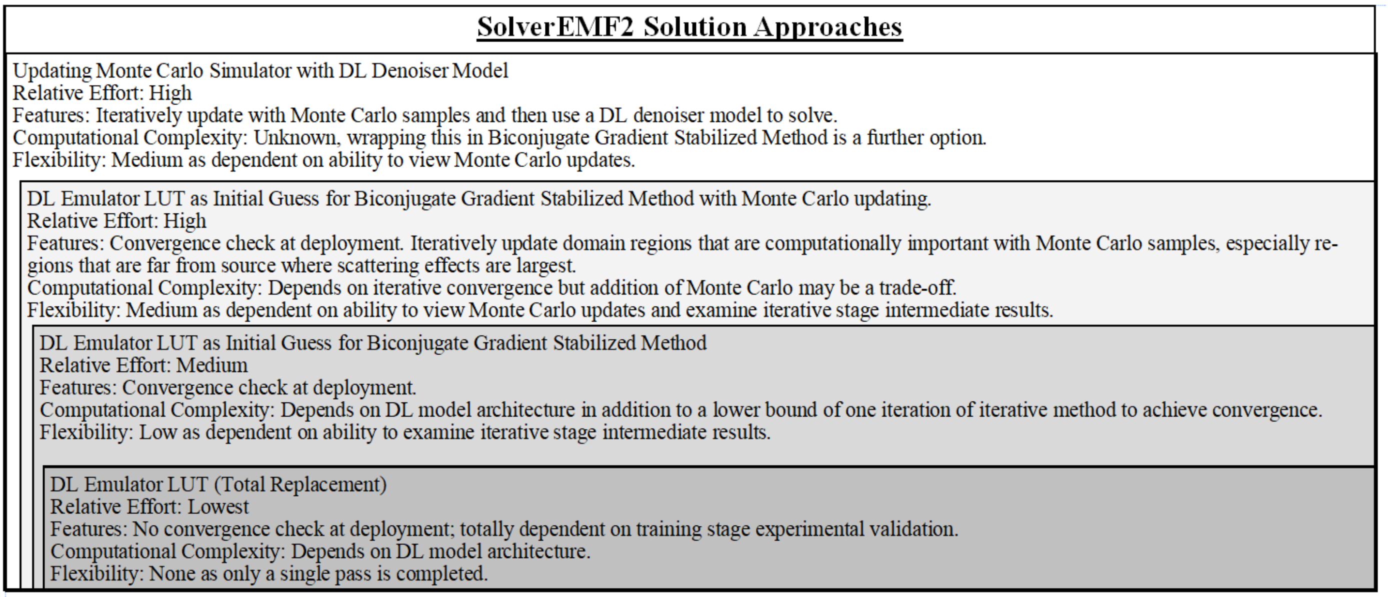
Deep Learning Dataset Development– Pathway

The creation of a benchmark dataset would be a publishable accomplishment in itself. The student proposes to start with a simple, narrow parameter settings database and as the feasibility of a model develops, the student plans on adding network architecture complexity.

- In the first instance, only train the model on data generated from canonical problem formulations with Bessel-function type solutions. Then benchmark an independent conventional model (MoM) against the Bessel-Function solutions and the ML model. If all three agree then the ML model can be trained next on the conventional model.
- Generate non-canonical dataset using MoM python code.
- Generating the training dataset will be computationally intensive. The student proposes to squeeze value out of the simulations by using data augmentation such as rotation and reflection. A large number of simulations should be possible by having a range over the input parameters.

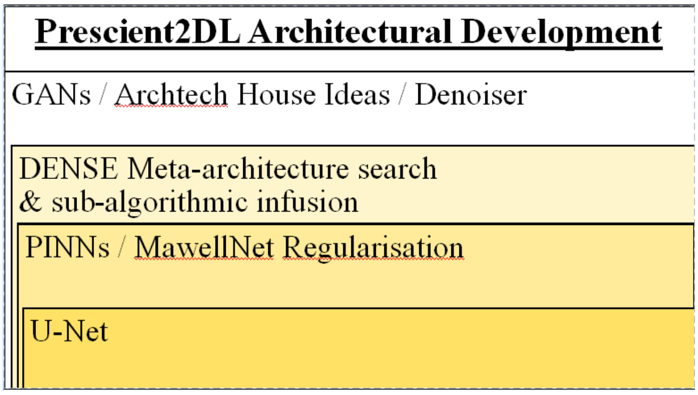
SolverEMF2 – Architecture

The architecture of SolverEMF2 initially sets all guesses to the Krylov Iterative Solver as an array of zeros. This step is later replaced with Prescient2DL informed guesses. The student proposes following the development of the solver methodology using the pathway illustrated in the image below. The student does not anticipate that the later Monte Carlo stages will be achieved in the time limitations of the project.



Prescient2DL – Architecture

The student proposes to follow the pathway outlined in the image below when expanding the model architecture and physics infusion.



Primary Research Test 01 – Initial Solution Conveyance t-Test

Null Hypothesis H_0: The initial error (Residual Norm) in the Krylov Iterative Metrics in SolverEMF2 is the same as for the non-DL assisted conventional solver.

Alternative Hypothesis H_A: The initial error (Residual Norm) in the Krylov Iterative Metrics in SolverEMF2 is lower than the non-DL assisted conventional solver.

Note: Conveyance is used to mean an indication of the level of information from the informative guess, be it Prescient2DL or a more vanilla approach, conveyed to SolverEMF2. The closer the Kyrlov Iterative Solver is to the be within the acceptable solution threshold, with respect to the naïve guess, then the more useful information has been conveyed via the guess. In the Test 01 situation, this manifests as simply the lower initial error.

Primary Research Test 02 –Solution Convergence t-Test

Null Hypothesis H_0: A linear approximation of the slope of the curve for plot Residual Norm versus Iteration Count, labelled as convergence rate, in the Krylov Iterative Metrics for SolverEMF2 is the same as for the non-DL assisted conventional solver.

Alternative Hypothesis H_A: A linear approximation of the slope of the curve for plot Residual Norm versus Iteration Count, labelled as convergence rate, in the Krylov Iterative Metrics for SolverEMF2 is the not equal to the non-DL assisted conventional solver.

Note: Further tests to establish if the absolute value of the slope is greater for Solver EMF2 may be required, however, it is expected that the convergence rate will remain unless sub-algorithmic integration of the model is successful in the secondary stage.

Primary Research Test 03 – Solution Conveyance t-Test

Null Hypothesis H_0: The area under the curve for plot Residual Norm versus Iteration Count, labelled as AbsementKIM, in the Krylov Iterative Metrics for SolverEMF2 is the same as for the non-DL assisted conventional solver.

Alternative Hypothesis H_A: The area under the curve for plot Residual Norm versus Iteration Count, labelled as AbsementKIM, in the Krylov Iterative Metrics for SolverEMF2 is smaller than for the non-DL assisted conventional solver.

Explanation: If both the Residual Norm and the Iteration/time vary, it may be difficult to judge the difference in performance based on a single parameter. This naïve approach to combining the parameters by finding the area under the curve intersecting the two axis may be a way to compare the computational expense expended by the various SolverEMF2 variants. AbsementKIM stands for Absement Krylov Iterative Method and is analogous to Absement in the domain of kinematics.

Secondary Research Tests – General

All tests conducted in the Primary Research Test stage will be applied to the secondary stage. To illustrate the idea of the primary metric from the Primary Research Test, the diagram below shows the baseline permutations that could arising when comparing the output metrics of the SolverEMF2 activities. The diagram on the left indicates an impact on initial errors while the right diagram indicates an impact on the rate of convergence of the solver. Both of these changes of parameter could change simultaneously so a third graph showing the exhaustive list of such graphs could be produced but is deemed too busy to be informative. The areas in green indicate the AbsementKIM that may arise. A smaller AbsementKIM in the final metrics would indicate that more information in general was passed through the initial guess relative to whatever configuration that was used to establish the null.

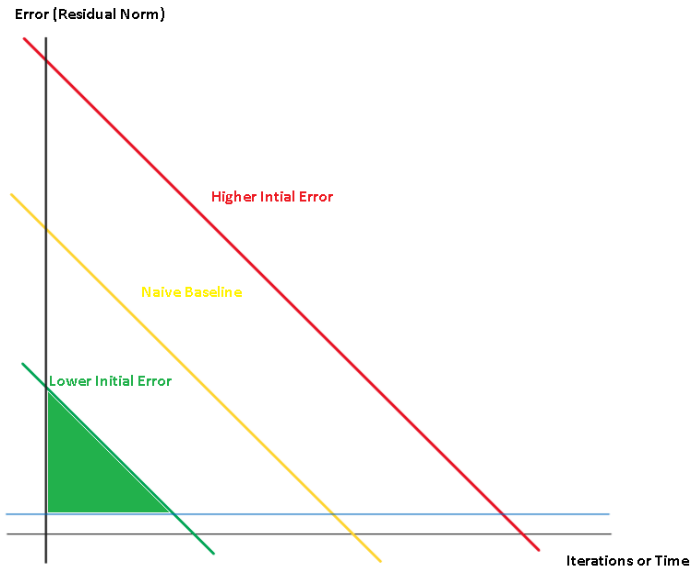


Figure 1: Impacts on Initial Error. Slopes are constant.

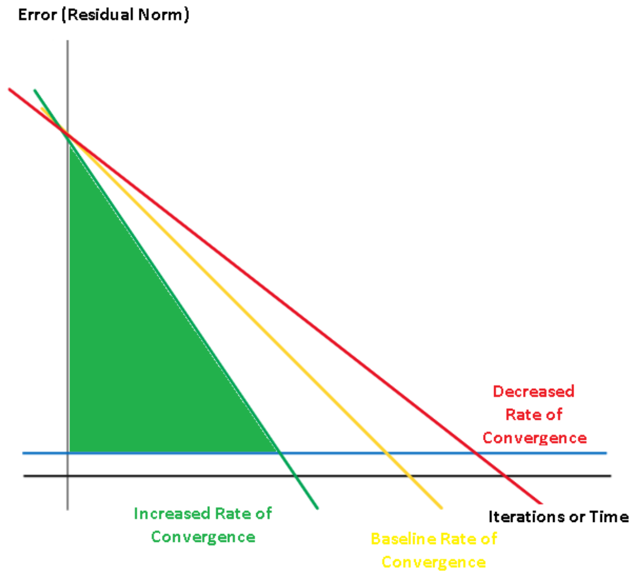


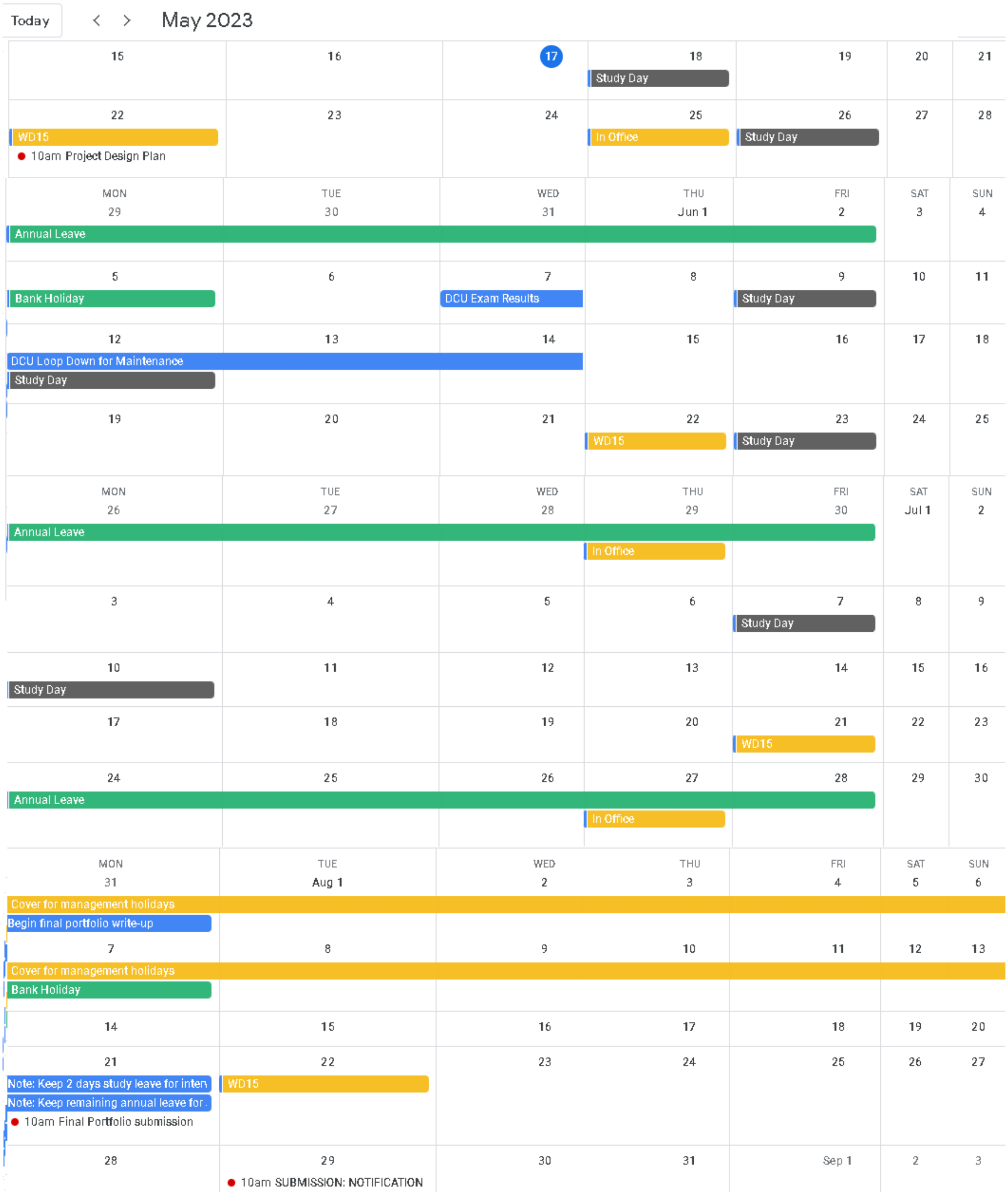
Figure 2: Impacts on Rate of Convergence. Initial error is held constant.

Secondary Research Tests – Model Development & Integration

The metrics used in the development process for Prescient2DL will be the residual sum of squares normally used in such regression problems. The developments will follow the baseline established in [1], [4] and expand to include aspects highlighted in the Literature Review. The generalisability of the model will be tested by generating a random shapes dataset, as opposed to the terrain dataset used in the rest of the project. Machine learning models are largely data-driven but the data in this experiment will be totally synthetic. As a result, the developed ML model will be derived from governing equations that are an idealization and representative of that centric world-view. This is the same idea as the Helio-centric versus Geo-centric interpretation in inter-planetary motion. The objective is to build an ML model of an existing conventional method. Testing generalizability could be done by seeing how far Prescient2DL diverges from canonical solutions when initially only trained on canonical datasets, then trained on general datasets, then compare the model on canonical validation set and general test set to see if performance degrades on the canonical test in order to generalize to the generalized inputs. A decision around metrics may also be required, for example, tolerance of some extra resources for time of inference may be tolerated for lower error and vice-versa.

Timeline

As of the date of the submission of this project proposal document (2023/05/21) there are only 92 days to the final submission deadline for the completed project (2023/08/21). In this period, there are 64 weekdays and 28 weekend days. The student has 7 study leave days and up to 23 annual leave days from work. Due to the constraints of work, it is highly likely that not all of the annual leave can be taken in this period. In August, the student will be required to act as cover for management. There is also a strong bias towards taking leave towards the end of the month to avoid impact on the bulk of regulatory reporting between working days 8 and 15. The student is populating the Google Calendar facility, available with their DCU email account, with key milestones and constraints. This is illustrated below. Due to the iterative and unknown time requirements for the tasks outlined in the Project Scope and Design Plan, the student has avoided pinning dates to specific actions with the exception of "Final Portfolio Write-Up". Due to severe constraints in August, the preparation for the final portfolio will need to commence at the start of August and while experimentation can run in parallel in the background, it is highly unlikely that new developments or additions to the model development architecture will be completed beyond this date.



Success Criteria

The student proposes that the criteria for success remain open-ended with the aspiration of completing all items on the list below while acknowledging that the time and resource constraints may curtail its completion. Chapter 13 of [6] offers some open problems in the domain of applying DL to electromagnetic problems. The relevant ones for this specific subdomain can be summarised as trying to generate more generalisable models using less data in a more efficient manner without overfitting. The suggestion is to enhance the model architecture with physically-based loss functions and generate foundation-type models that can be adapted and fine-tuned via transfer learning approaches.

- Validated literature review. See log entry relevant to [6].
- 10 research log submissions updating work progress and developments in the field.
- Python implemented SolverEMF2 workflow that can be used and expanded in future by students of the field.
- Mature Prescient2DL model architecture that can be deployed to accelerate the generation of solutions to these types of problems.
- A dataset of solved simulations for future use by researchers in this field.
- A completed final report portfolio for submission to DCU for masters accreditation that provides answers to the outlined primary and secondary research questions.
- A paper on (https://arxiv.org/) or in a peer-reviewed journal reporting some finding from the second stage of research questions.

Bibliography

[1] C. Brennan and K. McGuinness, "Site-specific Deep Learning Path Loss Models based on the Method of Moments." arXiv, Feb. 02, 2023. doi: 10.48550/arXiv.2302.01052.

[2] V. Pham-Xuan, "Accelerated iterative solvers for the solution of electromagnetic scattering and wave propagation propagation problems," doctoral, Dublin City University. School of Electronic Engineering, 2016. Accessed: Mar. 20, 2023. [Online]. Available: https://doras.dcu.ie/20951/

[3] P. M. van den Berg, *Forward and inverse scattering algorithms based on contrast source integral equations*. Hoboken, NJ: Wiley, 2020. [Online]. Available: https://onlinelibrary.wiley.com/doi/book/10.1002/9781119741602

[4] Q. Ren, Y. Wang, Y. Li, and S. Qi, *Sophisticated Electromagnetic Forward Scattering Solver via Deep Learning*. Singapore: Springer, 2022. doi: 10.1007/978-981-16-6261-4.

[5] J. Lim and D. Psaltis, "MaxwellNet: Physics-driven deep neural network training based on Maxwell's equations," *APL Photonics*, vol. 7, no. 1, p. 011301, Jan. 2022, doi: 10.1063/5.0071616.

[6] A. P. M. Li, M. Li, and M. Salucci, *Applications of Deep Learning in Electromagnetics: Teaching Maxwell's Equations to Machines*. Institution of Engineering & Technology, 2023.

Approval

Signature of Project Worker: Anthony James McElwee Date: 2023/05/20

Print name of Project Worker: **ANTHONY JAMES MC ELWEE**

Signature of Project Supervisor: Conor Brennan Date: 2023/05/21

Print name of Project Supervisor: **DR CONOR BRENNAN**