CAN THE FOURIER NEURAL OPERATOR ACHIEVE SUPER RESOLUTION FOR THE PREDICTION OF STRESS AND DAMAGE FIELDS IN COMPOSITES?

Dylan Gray – MEng (Hons) Integrated Mechanical and Electrical Engineering
Department of Electrical Engineering, University of Bath, Claverton Down, Bath, UK

ARTICLE INFO

Keywords: Fourier Neural Operator; Deep Learning; Full-field Prediction; Composites; Stress; Fracture

ARTICLE DETAILS

MEng Final Project Report
Student Number: 209286171
Academic Year: 2023/24
Supervisor: Yang Chen
Assessor: Biagio Forte
Abstract Word Count: 166
Main Body Word Count: 7114

ETHICAL APPROVAL

Received: 08 Feb 2024

ABSTRACT

A novel deep learning technique is applied to create a surrogate model for the full-field predictions of stress and damage in composite materials. The model is developed using the Fourier Neural Operator (FNO) and outputs elastic stress fields and local damage maps. The model makes these predictions with high resolution Representative Volume Elements (RVEs) as inputs. These RVEs are randomly generated to produce diverse training and testing data containing geometric variabilities such as fibre volume fraction and porosity. The development and testing of the surrogate models demonstrate the power of the FNO compared to traditional physics-based systems of Partial Differential Equations (PDEs), and classical Convolutional Neural Networks (CNNs). The FNO models are evaluated using unseen data to prevent biased results from the training data. After training, the FNO produces accurate full-field predictions up to 1.5×10^4 times faster than the physics-based model while also demonstrating resolution invariance. The accuracy and computational efficiency of the FNO highlights it as a powerful method for multiscale simulation and microstructure optimisation.

TABLE OF CONTENTS

TABLE OF ACRONYMS

ACRONYMS	DEFINITIONS
CNN	Convolutional Neural Network
FFT	Fast Fourier Transform
FNO	Fourier Neural Operator
ML	Machine Learning
MSE	Mean Squared Error
PDE	Partial Differential Equation
RVE	Representative Volume Element
SSIM	Structural Similarity Index Measure

TABLE OF FIGURES

TABLE OF TABLES

1 Introduction And Context

Material development is key for the creation of high-performance, cutting-edge technologies which are always in major demand across various industries. These innovations work to shape the future of engineering applications. It is vital that the properties of these materials are well understood at a microstructure level so that they can be effectively optimised. The methods to accurately analyse these material properties and behaviours are in constant development, however, continue to face significant challenges especially when dealing with the inherent complexity of the microstructures required for modelling, for accurate real-world applications [1].

Recent advancements in the field of computational materials science have been impressive, with a range of novel solutions developed to improve modelling capabilities involving complex microstructures for material analysis and predicting behaviours. One such area of focus is the prediction of stress and damage fields in composite materials which is essential for developing the mechanical properties of components engineered using composites.

Current physics-based modelling of stress and damage in composites is computationally expensive, as it involves the solving of large systems of PDEs within high resolution meshes and for a significant number of small timesteps. Despite significant advances in modern computing, finding accurate and stable solutions becomes extremely time consuming, if not unfeasible, especially when the composite material demonstrates complex structures. Using traditional PDE based solvers, decreasing computation time by reducing mesh resolution leads to the compromise of reducing accuracy. Complex geometries present in microstructures require a high-resolution mesh to produce accurate solutions.

The continued search for increasingly efficient and accurate modelling methods has expanded past traditional methods and has begun to include the potential applications of machine learning techniques. These techniques can be used to produce surrogate models with the aim of significantly reducing computation time while still producing highly accurate predictions. The Fourier Neural Operator, which has proven its capability and performance in the field of fluid mechanics [2], stands out as a novel approach; however, its potential regarding the prediction of stress and damage fields in composites remains largely unknown.

The aim of this paper is therefore to explore the potential of the FNO in the field of solid mechanics by developing and training FNO based models for the prediction of stress and damage fields in composites. The proposed models look to offer a promising alternative to traditional physics-based solvers by utilising the power of deep learning and the FNO. Through comprehensive testing and analysis of the models, and by exploring a diverse range of complex microstructural geometries, the feasibility, efficiency, and accuracy of the FNO-based surrogate models is demonstrated.

The development and evaluation of these surrogate models significantly contributes to the field of computational materials science by advancing modelling capabilities within the field. This work allows for further development, and increasingly accurate and efficient models for the prediction of stress and damage fields in composite materials. This has implications for a wide range of real-world applications and industries, from sports to aerospace. Further to this, through thorough and comprehensive testing and analysis, the foundation is laid for future research efforts.

This paper presents the development and testing of FNO based surrogate models for the prediction of stress and damage fields in composites. The development explores composites with a diverse range of microstructure geometries and within these microstructures predicts damage maps containing significant, complex damage structures. To the best of the author’s knowledge this paper presents the first example of the FNO being applied to stress and damage predictions within composite materials.

2 BACKGROUND RESEARCH

Driven by the demand for high performance materials, a range of modelling techniques have been developed, each aiming to improve performance and accuracy. This section provides a review of existing literature and discusses their respective strengths and weaknesses. Through thorough critical analysis of this work, areas for future research are highlighted.

As previously mentioned, research involving stress and damage fields in composite materials typically requires physics-based modelling techniques for predictions. This often relies on complex systems of PDEs and while highly accurate when implemented correctly, they require large amounts of computational power and are time consuming, especially when trying to solve for complex microstructure geometries. These challenges of traditional solvers have led to the exploration of more novel approaches such as deep learning [3]. Recent research has shown the potential of producing surrogate models which utilise deep learning methods for creating full-field predictions for composites.

It is worth noting that terminology is not consistent amongst the literature, therefore there is a range of terms for stress and damage in composites. The outputs of the models are often interchangeably called fields or maps. Damage has a wide variety of terms including damage, crack, and fracture. Quantitative evaluation metrics are also inconsistent between studies which makes comparison of performance more challenging. Popular metrics include Mean Squared Error (MSE), Structural Similarity Index Measure (SSIM), and F1 score [4] [5] [6].

Sepasdar et al. [7] developed a deep learning model for the prediction of nonlinear stress distribution and failure patterns in composites. Their method involved the development of two fully convolutional networks which were trained sequentially. The first network acted to predict post-failure stress distribution, and then the second network was used to translate this to a failure pattern. A physics-informed loss function was also implemented to improve performance. This method was very promising and was capable of making predictions with around 90% accuracy. However, the diversity of the training and testing samples was limited. The data contained only a single fibre volume fraction of 57.5%, all fibres had the same diameter of 7μm, and no pores were included. This level of complexity is not sufficient to be applicable to real-world problems.

Similarly, Chen et al. [8] developed two sequential CNNs for full-field predictions of stresses and cracks in composite material microstructures. This work included a self-attention layer within the networks to improve the capturing of relevant global and local features. This implementation allowed for predictions of around 70% and 80% accuracy for stress fields and damage maps respectively. The training data was diverse and of high quality, including a wider range of fibre volume fractions, a range of fibre diameters, and a range of porosities. This inclusion of high-quality data ensures that the results of the testing are more accurate to real-world applications. To achieve the performance outlined in the research, a large data set of 7200 samples was required for training and testing. Producing such large data sets is time consuming and may not be feasible for all real-world scenarios. Also, the damage surrogate was unable to make accurate predictions using the outputs of the elastic surrogate as an input. This was likely due to error accumulation from the elastic surrogate. Instead, the damage surrogate required inputs directly from the training data set.

A CNN using a U-Net architecture [9] was developed by Bhaduri et al. [10] for the prediction of stress fields in fibre-reinforced composite materials. Their work also investigated the potential of reducing the computational cost of training by making predictions with larger numbers of fibres using models pretrained with samples containing a smaller number of fibres. Through this, they successfully demonstrated that their approach led to accurate predictions and the inclusion of pretraining further increased model performance and generalisation capability. However, the diversity of the training data was limited, and no pores were included. As well as this, the study was limited to stress at a single load step, as the training data did not contain stress until failure. This means that stress at final fracture and damage maps were unable to be modelled.

Sholrollahi et al. [11] investigated the application of deep learning techniques for the prediction of stress fields in composite materials. They also implemented a CNN with U-Net architecture however, their work included a more diverse dataset for training and testing. The dataset contained RVEs which had uniform fibrous structures, non-uniform fibrous structures, irregularly shaped fibrous structures, and a combination of the structures. Despite this range of training data, the model exhibited high levels of performance and achieved similarity index scores of 0.977 on test data. While the data was varied in respect to fibre structure, the data did not include any pores; this aspect is required to accurately reflect composite microstructures. In addition, the model developed for this study was limited to stress fields only, and this was within the elastic range; therefore, the complex behaviours that occur when the composite fractures were not evaluated.

Mohammadzadeh and Lejeune [12] developed a deep learning model using MultiRes-WNet architecture for full field predictions of crack propagation. This architecture was developed as a modification of the MultiRes-UNet architecture [13] to improve feature extraction and performance. For the prediction of crack paths, the approach achieved a mean F1 score of 0.87. The model demonstrated good performance for the prediction of long crack paths which is a challenging problem. However, this performance was only shown for small image samples of 28x28 pixels, which did not contain any pores. Therefore, the study does not show the capability of the model for the prediction of cracks in sufficiently complex microstructures.

Implementing a game-theory based conditional generative adversarial network (cGAN) with a focus on plastic deformation, Yang et al. [14] developed a Machine Learning (ML) based approach for the prediction of physical fields from material microstructure geometries. This model was shown to achieve high accuracy for predicted field data as well as properties such as overall residual stress. The input geometries used for training the model contained brittle and soft units that were arranged into a random geometry. This geometry generation technique led to a gaussian distribution of soft and brittle units centred around 0.5. Therefore, it is unlikely that generated geometries are representative of real-world geometries. Instead, real-world geometries will cover a wide range of constituent ratios and will contain assembled patterns as a result of microstructural interactions. Further development and evaluation of this model would require improvements to the diversity and quality of the training and testing data.

Ghomali et al. [15] developed a deep neural network for the prediction of mechanical properties within composite hydrogels [16]. The network was trained and evaluated on a large dataset of 9000 2D images which were generated using finite element analysis. The study found that using an AlexNet architecture resulted in the highest performance and accuracy for the prediction of elastic material properties with the network achieving R-values of 0.99 and 0.97 for the prediction of Young's Modulus and Poisson's Ratio respectively [17]. This work highlights the wide range of applications for which materials developed using computational materials science can be used, with composite hydrogels finding use in the medical field. The data set developed as part of this study showed good diversity with the creation of four classes of samples. These classes included uniform-shapes, non-uniform shapes, irregular-shapes, and a final class containing a combination of the other three classes. The quality of this data ensures strong applicability to real-world scenarios. While this study achieved its aim of accurately predicting some elastic material properties, the scope of the study was limited and did not cover more complex behaviours such as predicting stress distributions. Further development of this model would therefore be required to investigate this.

The Fourier Neural Operator was proposed by Li et al. [2] for leaning the resolution-invariant solution operator for the family of Navier-Stokes equations. The method proposed in this study was applied to Burger's Equation, Darcy Flow, and Navier-Stokes, achieving lower error rates than all existing deep learning methods it was compared to. The average inference time per prediction also showed significant improvements compared to numerical methods with inference times reducing from 2.2×10^{1}s to 5.0×10^{-3}s for Navier-Stokes predictions. The model development and analysis carried out as part of this study is thorough and covers a diverse range of complex fluid mechanics problems. While not extending into the field of solid mechanics, this study successfully demonstrated that FNO-based surrogate models can approximate highly non-linear, complex systems of PDEs. The performance of the FNO in this area therefore highlights it for future research to extend its fields of potential application.

In conclusion, the development of a wide range of modelling techniques in the realm of computational material science has extended the tools at researchers' disposal from traditional physics-based solvers to novel deep learning approaches. While each method has its unique advantages, they also come with challenges. Deep learning models demonstrate excellent potential in generating highly accurate full-field predictions within composites. However, the performance of these models is closely related to the quality and diversity of the data sets on which they are trained and evaluated. As well as this, to be applicable to real-world scenarios it is necessary for these training sets to be comprehensive and capture the real-world complexity of the microstructural geometries and behaviours exhibited by composite materials. Despite these challenges, recent advancements in full-field predictions, as discussed so far in this report, show significant promise for providing ever greater levels of insight into material properties and behaviours. Continued research in this area is key for future development of high-performance advanced engineered materials for a diverse range of applications, from structural engineering to biomedical sciences.

3 METHODS

3.1 RVE GENERATION

This section briefly discusses the generation of the RVEs which were used for the development and testing of the surrogate models. A large high-resolution dataset was produced as part of earlier work for the development of a CNN with self-attention for the prediction of stress and damage fields in composites which has been utilised for this study [8]. This RVE generation procedure built on earlier work investigating mechanical properties of composites at a microstructural level [18].

First, a fibre volume fraction is defined and the number of required fibres for this value is calculated. The fibre diameters have a gaussian distribution with a mean of 7µm and a standard deviation of 0.4µm. When the RVE is initialised, these fibres are randomly placed within the 2D plane, and their positions are iteratively updated so that the fibres no longer overlap. The generated RVEs have pixel size 0.2µm therefore the iterative position updating is continued until a fibre separation of 0.28µm is achieved. This ensures there is no overlap in the discretised RVE. A range of fibre volume fractions were selected for this study as shown in **Table 1** below. Each fibre volume fraction was created for six different porosities with 150 samples created per porosity.

Table 1 - *RVE distribution.*

Fibre Volume Fraction	Total Number of Samples	Training Samples
0.25	900	Training and Validation
0.30	900	Training
0.35	900	Validation
0.45	900	Validation
0.50	900	Training
0.55	900	Validation
0.60	900	Training
0.65	900	Validation

The pores are placed randomly within the matrix and their size is increased according to a function related to the centre of the pore and the distance to the nearest fibres, $r = s \,.\, d_{f,min}$. Pores are then iteratively placed until the desired porosity is achieved. Examples of the generated RVEs are shown below in **Figure 1**. Each RVE has dimensions 50µm x 50µm.

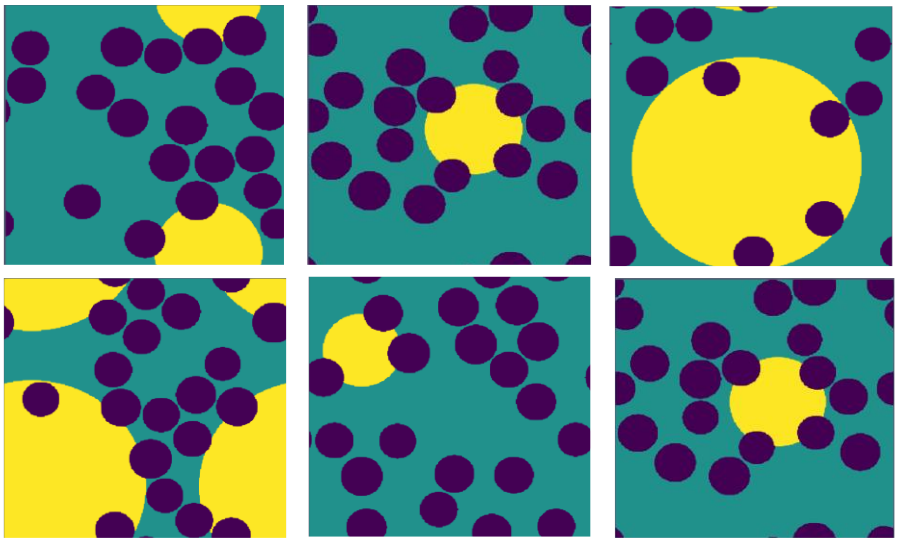

Figure 1 - *Examples of Generated RVEs displaying a range of geometric variability including varying fibre volume fractions and porosities. The three phases are shown as green for matrix, purple for fibre, and yellow for pore.*

3.2 PHASE-FIELD FRACTURE MODELLING AND FFT SIMULATIONS

The elastic response and damage within the microstructure under transverse tension was predicted by implementing the variational phase-field fracture model of Miehe et al [19]. This model was selected as it has been shown to be capable of accurately predicting damage initiation and propagation for complex cracking scenarios. This model describes the damage problem with two coupled PDEs. An FFT method was used to solve this system of PDEs. This method was selected for its efficiency and ease of implementation for parallel computing with high performance computing (HPC) systems [8]. The transverse properties of the carbon fibres and epoxy matrix selected for the study are shown below in **Table 2**.

Table 2 - *Material transverse properties.*

	Elastic Moduli (*GPa*)		**Fracture Parameters (*N/mm*)**
Material	**λ**	**μ**	**g_c**
Carbon Fibres	4.167	6.250	1.000
Epoxy Matrix	4.795	1.353	0.011

The simulations were carried out on the HPC cluster BALENA at the University of Bath and each simulation was carried out until final fracture. Final fracture is identified with a significant and abrupt decrease in overall stress. The computation time was dependent on the complexity of the individual RVE, and stress and damage fields. Therefore, computation time covered a range of values between approximately two and five minutes.

The simulation calculated stress, strain, and damage maps at every load step, examples of which can be seen in **Figure 2**. It is worth noting that the maps are influenced by both local and global microstructure geometries which means that the surrogate models need to be able to successfully capture local and global features to make accurate predictions.

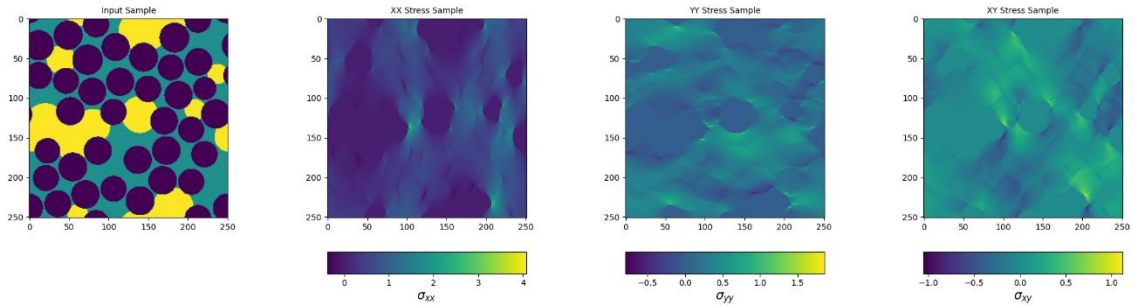

Figure 2 - *Example input mesh and stress maps for each axis at initial load step.*

The FFT produces two damage maps, one for fibre damage and one for matrix damage. For training and evaluation of the damage surrogate, the two maps are summed together to produce a single damage map. This process is shown below in **Figure 3**, alongside the respective input meshes.

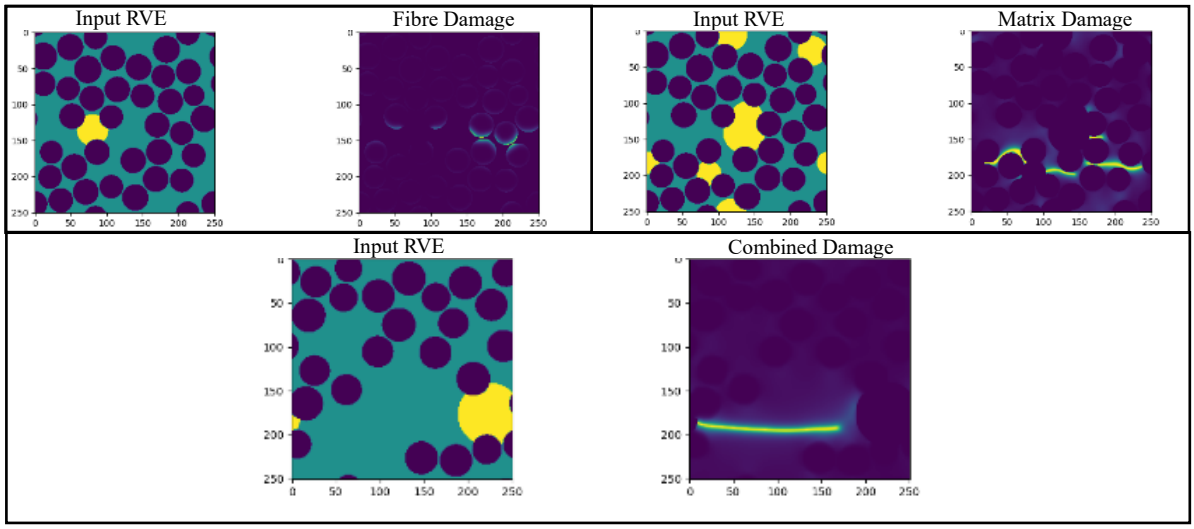

Figure 3 - *Fibre damage (top left), Matrix damage (top right), Combined damage (bottom)*

3.3 Fourier Neural Operator

Traditional solvers attempt to find the solution to the PDEs describing the system, whereas neural operators attempt to directly approximate the operator to find an accurate prediction. This has multiple benefits over traditional solvers which are computationally expensive and only solve for a single instance of the problem. The FNO has been demonstrated to be a powerful method for learning operators normally described by complex systems of PDEs and therefore its application to stress and damage predictions in composite microstructures is explored here.

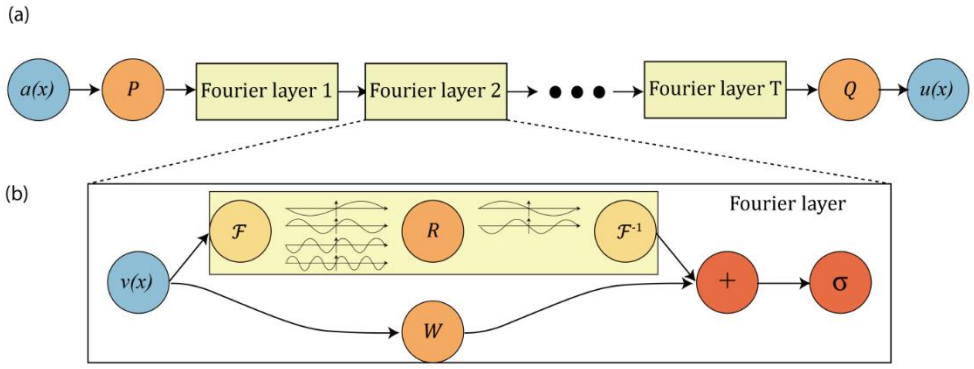

Figure 4 - *(a) Full Neural Operator Architecture (b) Fourier Layer Architecture*

Figure 4 above shows the architecture of the neural operator and the structure of the Fourier layer. The neural operator takes an input and first raises it to a higher dimensional channel space as in a traditional CNN. The Fourier layers are then applied before being reduced back to the required dimensions, by a neural network, for output. The Fourier layers operate by performing a Fourier transform to convert the input to Fourier space. Once in Fourier space, a linear transform is applied to the lower Fourier modes and the higher modes are filtered out. In parallel to this a local linear transform is also applied.

The Fourier layers are resolution invariant as the functions which they learn and are evaluated on, are discretised in an arbitrary manner. Through earlier work, it is shown that FNO-based models display consistent errors at any input and output resolution, a characteristic which is not shared by traditional CNNs where error increases as resolution increases [2].

The architecture selected in this research, takes inputs of dimensions 251x251 pixels and uses two convolutional layers to lift the input data into higher-dimensional space. The data is then passed through the Fourier layers before being projected back to the original space of 251x251 pixels by two further convolutional layers. This architecture has been applied to produce two surrogate models for the prediction of stress fields and damage maps respectively. Both surrogate models take the same three channel inputs of initial geometry map, x coordinate map, and y coordinate map. The stress surrogate produces a three-channel output of stress in the xx axis (σ_{xx}), the yy axis (σ_{yy}), and the xy axis(σ_{xy}). The damage surrogate instead produces a single-channel output of the damage map. This map contains fibre damage and matrix damage; however, due to the greater magnitude of matrix damage, fibre damage is more subtle. For training and evaluation, the fibre and matrix damages are summed during data-loading and preprocessing as discussed earlier.

3.4 Training

The dataset contained 7200 samples which were each complete simulations until final fracture. 240 of these samples were randomly selected as unseen samples for evaluation and contained a diverse range of geometric variabilities. For the stress surrogate, of the remaining samples, 2000 and 200 samples were then randomly selected for training and testing at each epoch. Using Pytorch and the CUDA toolkit, training was carried out for 500 epochs with batch sizes of 4. Mean squared error (L2) was selected as the loss function, and adaptive moment estimation (ADAM) was used as the optimiser. Initial learning rate was 1e-4 which was halved every 100 epochs, and weight decay was set to 1e-5. The damage surrogate instead used 4000 and 200 training and testing samples respectively. Initial learning rate was set to 5e-4 and training was carried out for 50 epochs, with learning rate halving at 25 epochs. All other training parameters remained the same as the stress surrogate. Computation was carried out using an Nvidia RTX3080 GPU with 32GB of 2666MHz DDr4 memory. Training took 424 minutes for the stress surrogate and 75 minutes for the damage surrogate.

3.5 Training Monitoring

To effectively train and evaluate the performance of the models, it was essential to implement robust monitoring mechanisms. Therefore, Weights and Biases logging was implemented [20]. This enabled the training process to be closely monitored to better understand the model training process in real-time. Identifying training issues such as overfitting, and hyperparameter tuning was able to be efficiently conducted. Logging was integrated into the training process using the *wandb* python library. Vital metrics including loss functions, learning curves and gradients, as well as model architectures, hyperparameters, and evaluation metrics were logged for each training run allowing quantitative comparison between runs. By logging this information, the overall performance of the model was able to be quickly developed, guided by the insights provided by analysing and comparing these metrics. This iterative process, supported by the detailed data logging, was key to the development of the models.

During initial development, training sweeps were also integrated using the *wandb* library. Bayesian sweeps [21] iterated through a range of hyperparameter values, including training set size, learning rate, and number of layers. After each run, the model was evaluated, and new hyperparameters were automatically selected to improve the performance of the model. This automation allowed rapid development of the model and quickly identified correlations between hyperparameter values and performance, giving direction for future development. An example of one of these sweeps is shown below in **Figure 5**. With each iteration, the sweep calculates correlations between hyperparameter values and resulting F1 scores which increases the likelihood of the sweep selecting more optimal values. It is possible to select a wide range of hyperparameters and a large set of values for testing. However, such a sweep would be inefficient and time consuming therefore sweep results must still be closely analysed and hyperparameters and values must be intelligently selected.

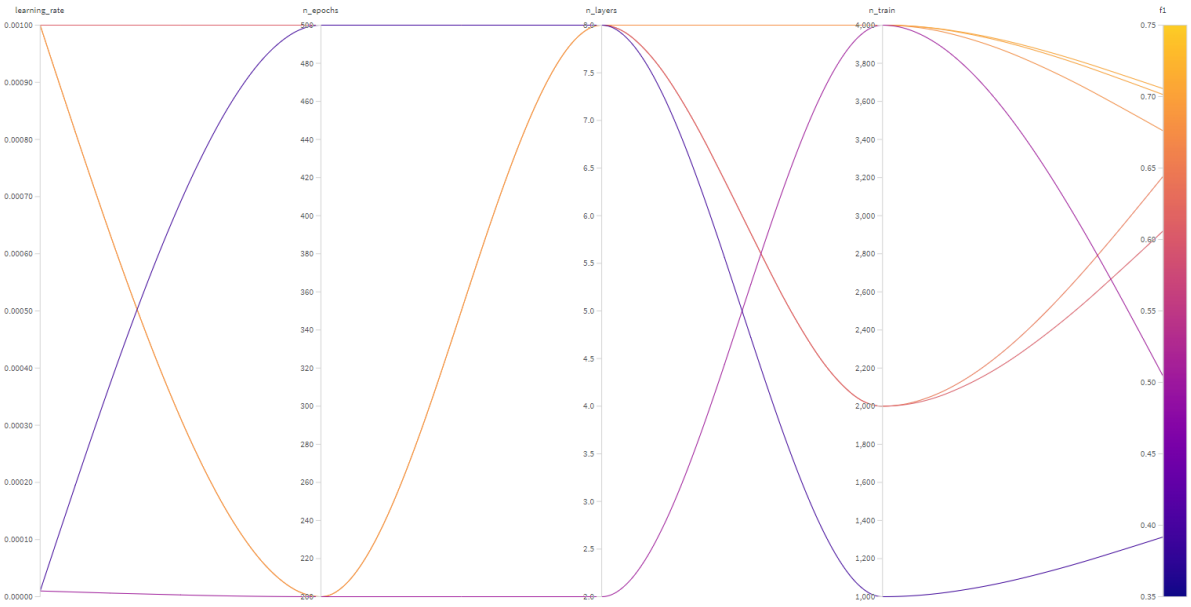

Figure 5 - *Bayesian sweep for damage surrogate using Weights and Biases.*

4 RESULTS

The two surrogate models were assessed against the results of the FFT simulation which is considered as 'ground truth'. F1 score was calculated to provide a quantitative assessment of the accuracy of the models. The capability of the models to perform zero-shot super-resolution was then assessed. The computational efficiency of the surrogate models was also compared against the FFT simulations to demonstrate the strength of the FNO in rapidly producing accurate predictions.

4.1 STRESS SURROGATE MODEL

Figure 6 below shows some example results from the stress surrogate model with RVEs with a range of geometric variabilities. Comparing the results with the 'ground truth' of the FFT simulations shows the FNO model produces results which visually strongly agree with the 'ground-truth'.

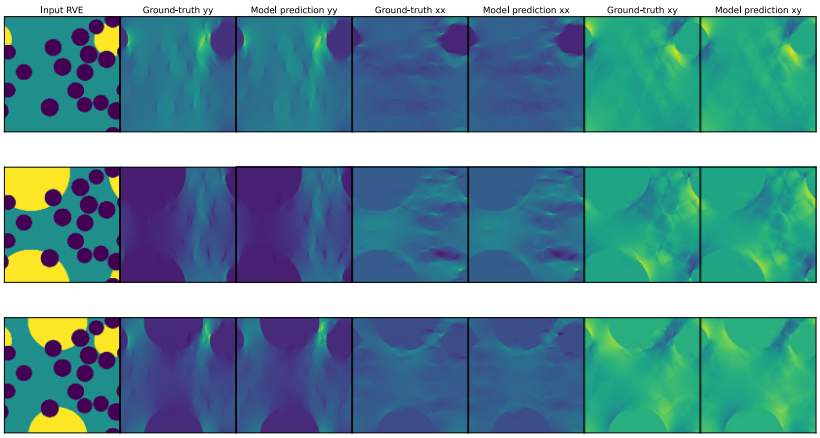

Figure 6 - *Inputs, 'ground-truths', and model predictions for stress surrogate.*

For a quantitative analysis of the accuracy of the stress surrogate, the first invariant of the stress tensor was calculated at each pixel and this map was then segmented. The segmentation was performed by selecting pixels corresponding to the 99th percentile of the map histogram, while excluding pore pixels. This highlights the location and geometries of areas of stress concentration. Segmented maps were produced of the FNO outputs and the 'ground-truth' and then an F1 score was calculated to compare the two.

F1 scores were calculated for all the unseen samples with the histogram of results shown in **Figure 7** below. The histogram shows that the surrogate model is able to produce stress field predictions with an average F1 score of 0.7.

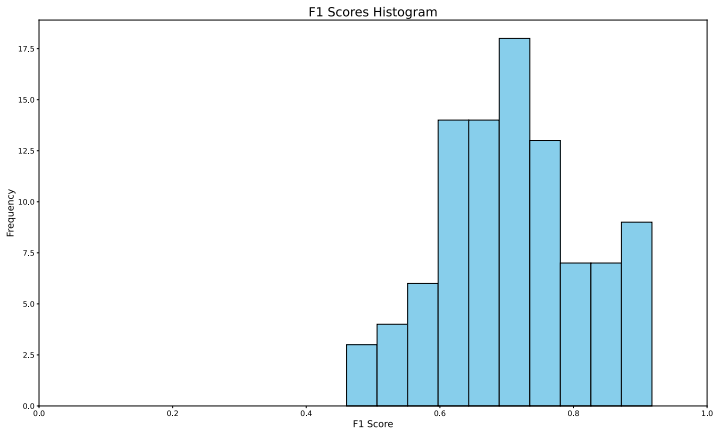

Figure 7 - *Histogram of F1 scores for stress surrogate on unseen samples.*

The value of using F1 score as a quantitative evaluation matrix is shown in **Figure 8**. Here a range of segmented maps for stress prediction are shown alongside their associated segmented 'ground-truth' stress maps and F1 scores. A higher F1 score correlates to a greater similarity between the two maps. As can be seen from the histogram of F1 scores, the model produces predictions which have a narrow distribution of F1 scores between roughly 0.5 and 0.9, centred around 0.7. This means that visually discerning the accuracy between predictions poses a significant challenge. This highlights the benefits of using a quantitative metric such as F1 score to provide greater insight into the performance of the model.

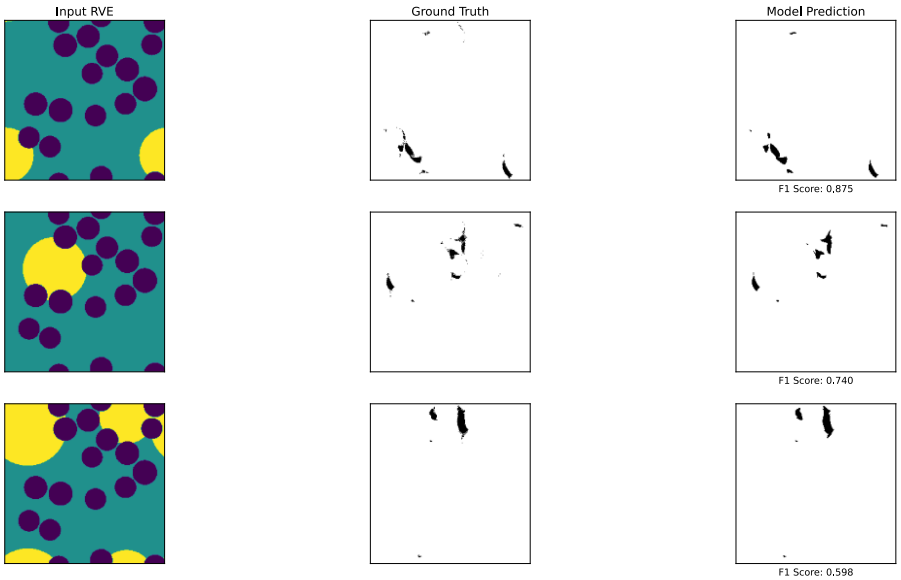

Figure 8 - *Inputs, segmented 'ground-truths', and segmented surrogate model predictions for a range of F1 scores.*

The power of the FNO is further demonstrated when smaller training samples and reduced epochs are used. With a learning rate of 0.005, 200 training samples, and training for 100 epochs, the surrogate model was able to make predictions with a mean F1 score of 0.62. While this score is lower than that of the full run, this is sufficient for visual analysis of the material behaviours with improvements in accuracy above this level being challenging to identify without quantitative analysis as discussed previously. To support this, representative example predictions are shown alongside the 'ground-truth' of the FFT simulation, in **Figure 9** below. The predictions shown in the top two rows have accuracies of 0.55 and 0.59 which is representative of the average accuracy. The bottom row represents the higher end of accuracy with an F1 score of 0.81. Visually comparing the three rows shows negligible difference in accuracy. This training run was completed in just 21 minutes which means that runs such as this could be hugely beneficial for rapid model development if the higher levels of accuracy displayed for a full training run are not required.

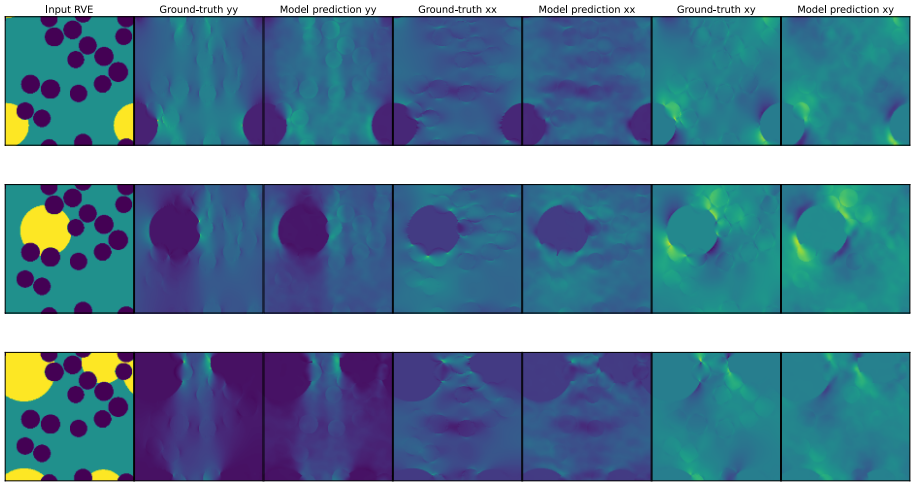

Figure 9 - *Inputs, 'ground-truths', and model predictions for reduced training stress surrogate.*

4.2 DAMAGE SURROGATE MODEL

Figure 10 shows some example predictions from the damage surrogate model alongside the respective input RVEs and FFT 'ground-truths'. Like the stress surrogate model, this model has also been applied to a range of unseen RVEs with diverse geometric variabilities.

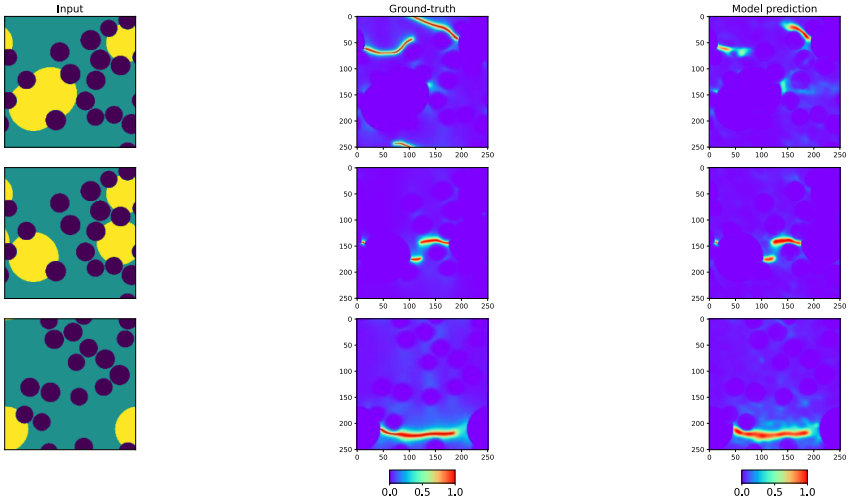

Figure 10 - *Inputs, 'ground-truths', and model predictions for damage surrogate.*

The accuracy of the FNO model was evaluated, again calculating F1 scores for the output of the FNO model against the output of the FFT. A threshold of 0.5 was used to create segmented maps from the damage maps. The segmented maps of the results used for calculating F1 score for the unseen samples is shown below in **Figure 11** which shows that the damage surrogate model is normally able to produce damage map predictions with F1 scores of 0.41. By again comparing low scoring and high scoring maps, the value of calculating F1 scores as a quantitative metric is shown.

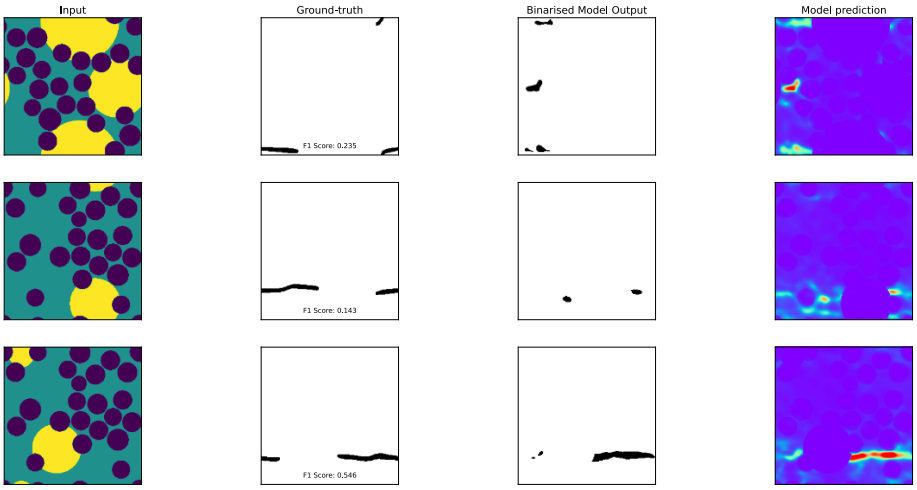

Figure 11- *Inputs, segmented 'ground-truths', and segmented damage surrogate model predictions for a range of F1 scores.*

The accuracies of the damage surrogate predictions are noticeably lower than the stress surrogate predictions; this could be for a range of reasons. Firstly, predicting damage is a more complex problem as the model needs to predict the location of the initial crack and then make an accurate prediction of its propagation through a complex microstructure. Later in this report, the computational efficiency of the surrogates is compared against the FFT simulations. This analysis reveals that the most computationally expensive aspect of the FFT simulation is the crack modelling, with average inference time per sample being roughly two orders of magnitude larger than inference time for stress modelling. This effectively highlights the additional complexity of the problem. Another potential reason is that when performing hyperparameter tuning, the range of values initially selected were informed by the tuning process of the stress surrogate. It was quickly revealed that the damage surrogate performed poorly using similar hyperparameters to the stress surrogate and therefore required significantly different parameters. Due to time restrictions associated with this project this second hyperparameter tuning process was limited; therefore, future development of this model would likely find better performance and accuracy through further tuning.

One of the most significant challenges with the development of the damage surrogate modelling was balancing underfitting and overfitting. Due to overfitting, early iterations of the model quickly learnt the training set and then was unable to make predictions on the unseen data. This continued to be a problem throughout the development process and therefore should be an area of future development. Overfitting can be reduced by exploring different model architectures such as modifying model depth and width. Another area to explore is the effect of training data size on overfitting as larger datasets can reduce this.

4.3 Zero-Shot Super-Resolution

The neural operator is mesh invariant which means it can be trained on lower resolution data and then be evaluated on higher resolution data. It can do this without seeing any higher resolution data before evaluation; hence the name, zero-shot super-resolution [22]. **Figure 12** shows some examples of low-resolution samples, 'ground-truths', and model predictions produced at a resolution of 64x64 pixels. This lower resolution training data was produced by downscaling the existing high-resolution data as part of the data-loading and preprocessing steps. It is worth noting that as the high-resolution data was already generated, the downscaling does not relieve the high-resolution data requirements for this study. However, in future implementations the ability to perform zero-shot super-resolution would mean a low-resolution training and testing data set could be used; only a high-resolution data set would then be required for input RVEs to produce high-resolution predictions. This would significantly reduce the computational cost of producing the required data.

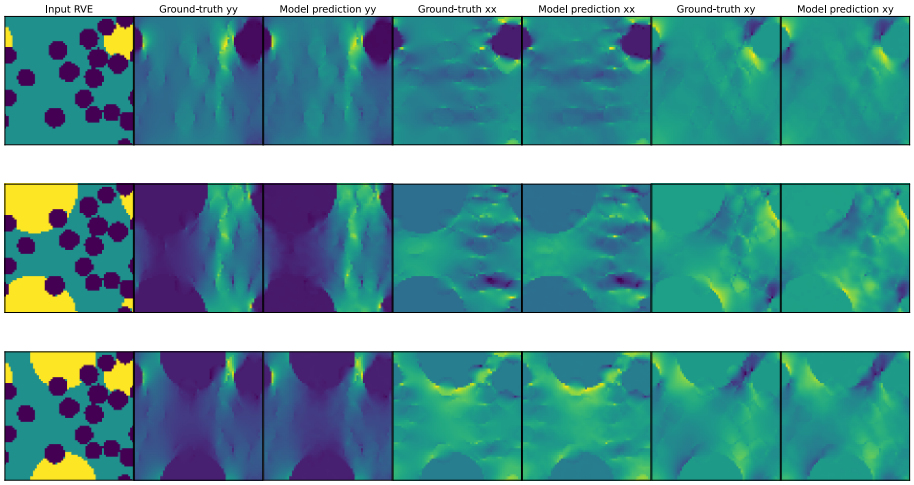

Figure 12 – *Low-resolution inputs, 'ground-truths', and model predictions for stress surrogate.*

Visually, the stress surrogate model continues to perform well at lower resolutions with strong agreement between the 'ground-truths' and the model predictions. **Figure 13** on the next page shows the segmented stress maps for 'ground-truths' and model predictions, which allows for quantitative analysis of model performance at lower resolutions. It is worth noting that the same hyperparameters were selected for the training of the low-resolution model as the full-resolution model. The reduction of parameters associated with lower-resolution training likely means updated hyperparameters could lead to improved performance. Despite this, the lower-resolution model demonstrates excellent accuracy as supported by both visual and quantitative analysis.

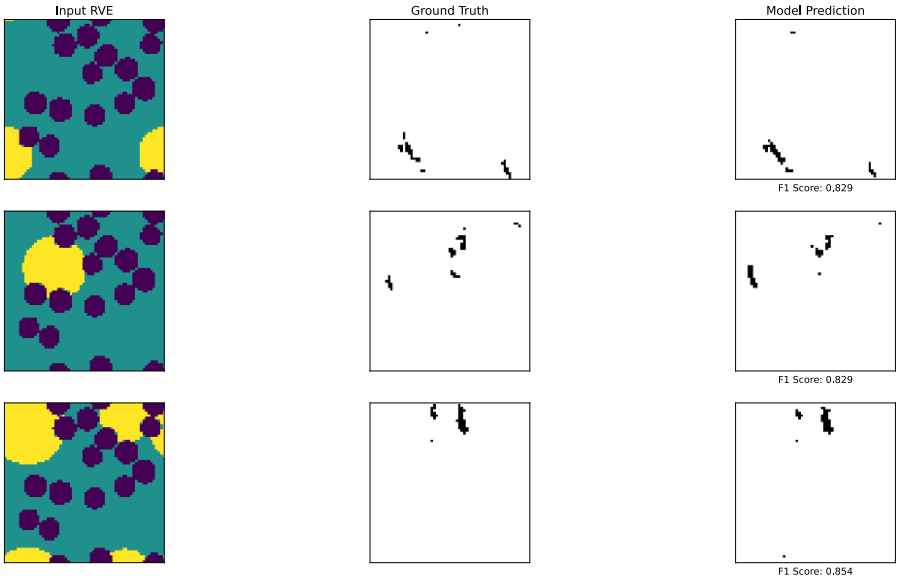

Figure 13 – *Low-resolution inputs, segmented 'ground-truths', and segmented predictions for stress surrogate model.*

Figure 14 below shows some example inputs, 'ground-truths', and zero-shot predictions from the stress surrogate which was trained using low resolution (64x64 pixel) training data and then evaluated using unseen full resolution (251x251 pixel) data. As with the full-resolution surrogate model, training was conducted for 500 epochs using 2000 training samples. Using this process, training time was reduced to just 61 minutes while an F1 score of 0.58 was achieved. These results show that the FNO can successfully perform zero-shot super-resolution and therefore, demonstrates that the model can be used for the prediction of high-resolution data without requiring any high-resolution data for training. This significantly improves overall computational efficiency for only a slight reduction in accuracy.

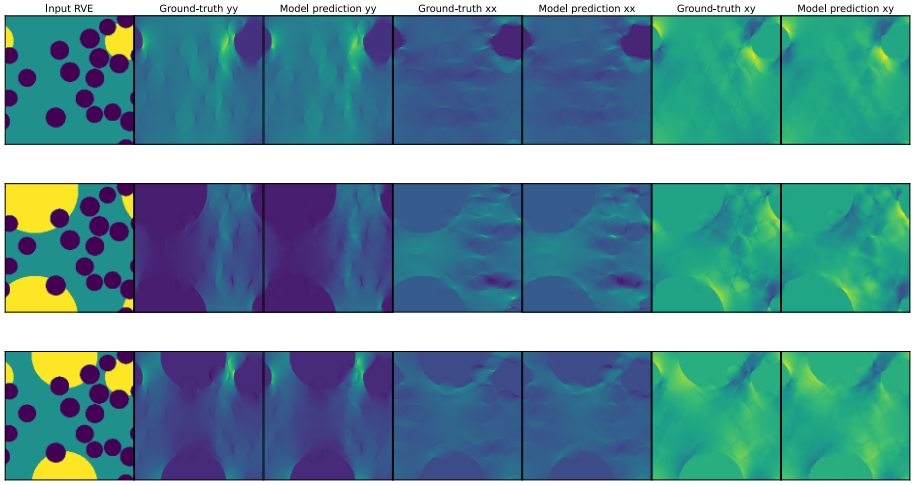

Figure 14 - *Inputs, 'ground-truths', and zero-shot model predictions for stress surrogate.*

Figure 15 shows the segmented stress maps produced by the stress surrogate model when performing zero-shot predictions. As with the standard implementation of the model, these segmented maps are used for the calculation of F1 score for each sample. While visually the model predictions strongly agree with the 'ground-truths' of the FFT simulations, this quantitative analysis further supports the accuracy of the model and confirms its ability to perform zero-shot super-resolution in the spatial domain.

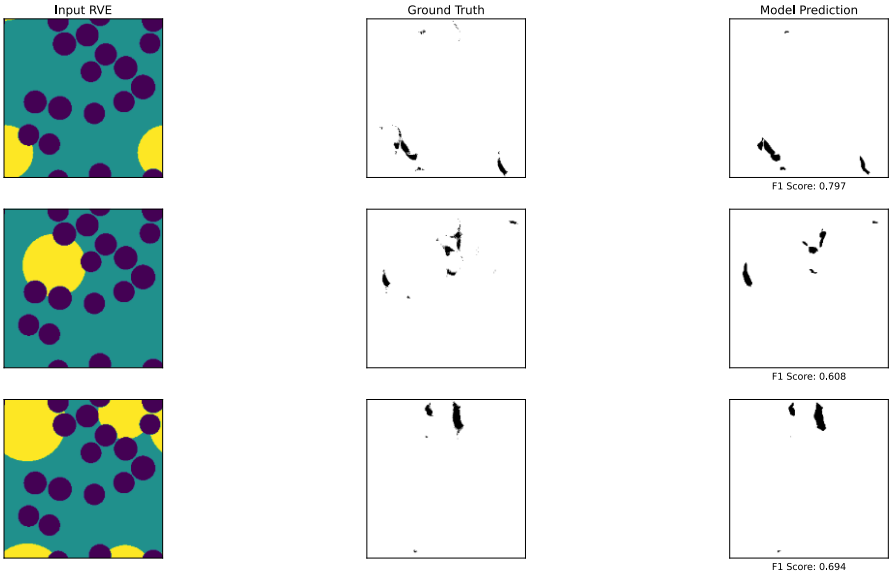

Figure 15 - *Inputs, segmented 'ground-truths', and segmented predictions for zero-shot stress surrogate model.*

The same process was implemented for the damage surrogate model as the stress surrogate model to evaluate its capability at performing zero-shot super-resolution. The training and testing data was downscaled, during the data loading and pre-processing stage, again to 64x64 pixels, and this was used to train a low-resolution model. A selection of low-resolution input RVEs, 'ground-truths', and model predictions are shown below in **Figure 16**. It is worth noting that for the evaluation of the low-resolution model, and later the zero-shot model, the performance should be compared to the full-resolution model for appropriate and accurate analysis. This is because apart from resolution, all other training parameters remain the same as the full-resolution model.

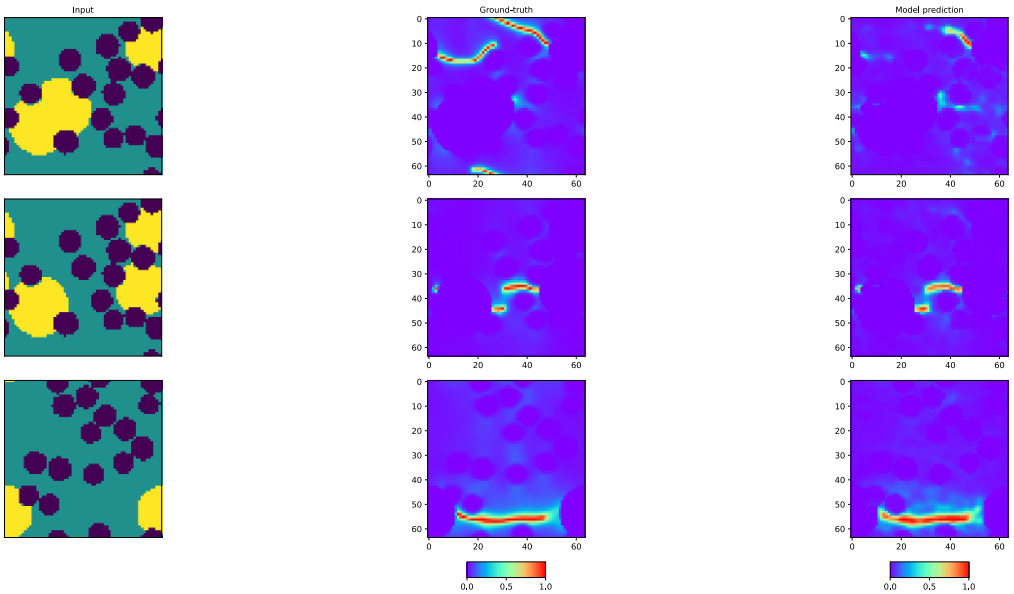

Figure 16 - *Low-resolution inputs, 'ground-truths', and model predictions for damage surrogate.*

The low-resolution damage surrogate model achieved similar performance to the full-resolution model which can be seen in the model predictions above in **Figure 16** and is supported by the segmented damage maps shown in **Figure 17** on the next page. Similarly to the low-resolution stress surrogate model, updating the hyperparameters for model training would likely result in improved performance and this should be explored in future research efforts.

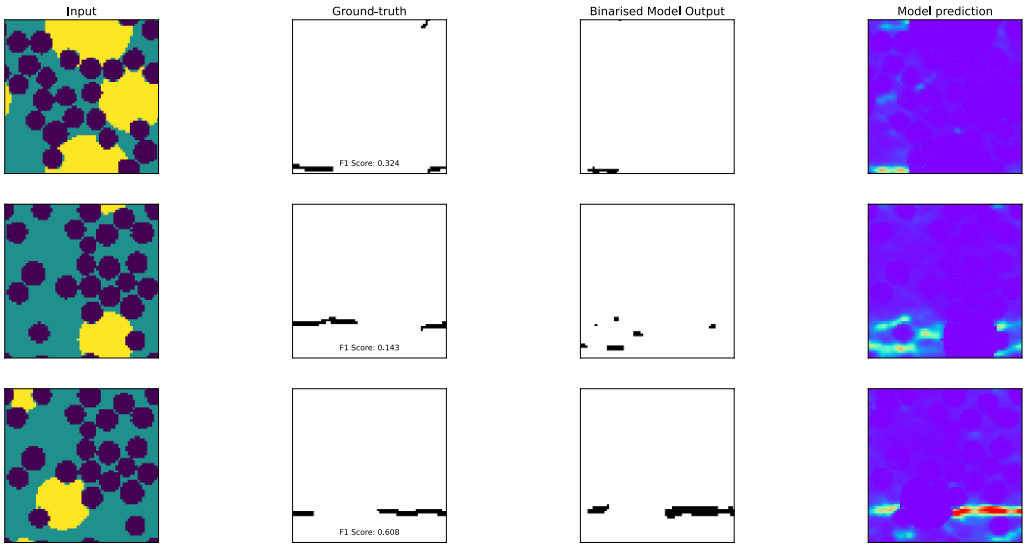

Figure 17 - *Low-resolution inputs, segmented 'ground-truths', and segmented predictions for damage surrogate model.*

The model was then given high-resolution inputs with the goal of producing accurate high-resolution predictions. As can be seen in **Figure 18** and considering the performance of the full-resolution model, this was successfully demonstrated. Comparing the low-resolution predictions shown in **Figure 16** and the zero-shot predictions shown in **Figure 18**, it is clear to see that the model maintains a similar level of accuracy while achieving a far higher resolution prediction.

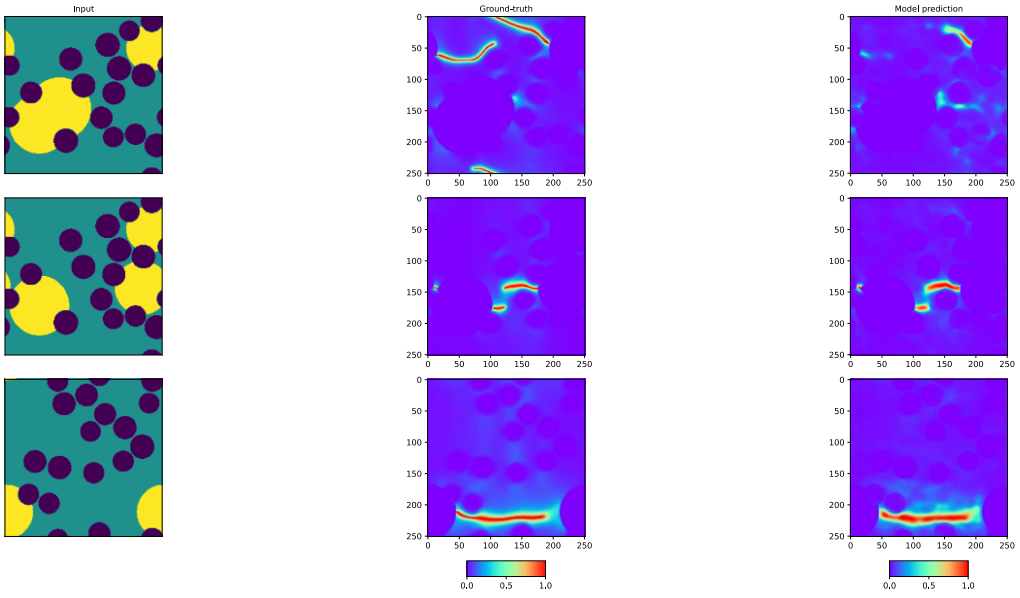

Figure 18 - *Inputs, 'ground-truths', and zero-shot model predictions for damage surrogate.*

The ability of the damage surrogate model to perform zero-shot super-resolution is supported by the segmented damage maps shown in **Figure 19** which are used to calculate the F1 score for each prediction. The zero-shot damage surrogate made predictions with an average F1 score of 0.39. This reduction in accuracy compared to the full-resolution model is similar to the reduction seen in the zero-shot stress surrogate. The benefit of the FNO being capable of performing zero-shot super-resolution is clear, considering the training for the zero-shot damage model took only ten minutes and required no high-resolution training data.

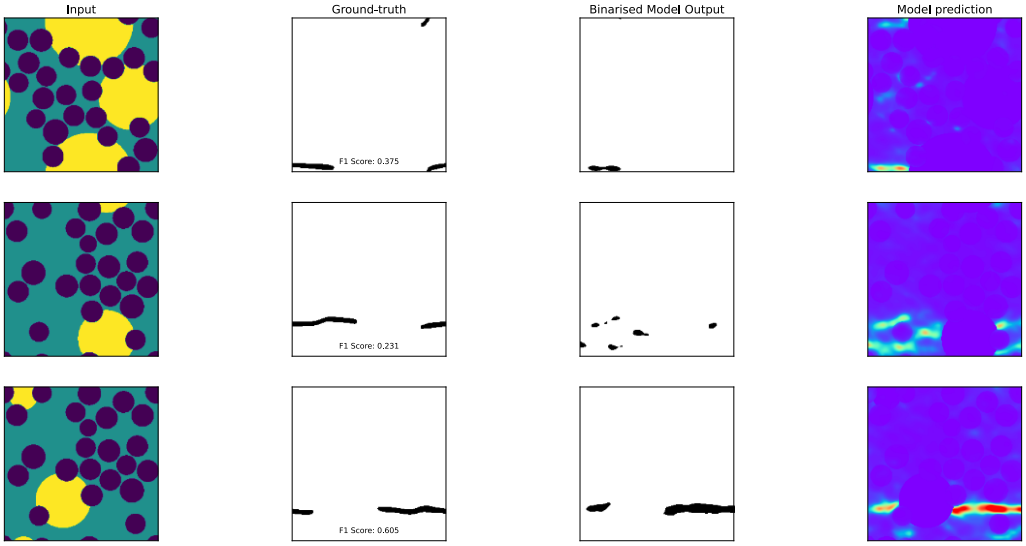

Figure 19 - *Inputs, segmented 'ground-truths', and segmented predictions for zero-shot damage surrogate model.*

Future development of the damage surrogate model is required to fully explore the potential of the FNO in performing zero-shot super-resolution for the prediction of damage maps within composite microstructures. A wider range of model architectures and hyperparameters should be investigated and evaluated to optimise model performance. As well as this, research should be carried out into the generation of the low-resolution samples. This research should include investigating alternative downscaling methods and the potential of generating the low-resolution data directly from FFT simulation. Investigating these alternative methods has the potential of lessening the slight reduction in model accuracy when performing zero-shot predictions. Due to the success of FNO-based surrogate models performing zero-shot super-resolution, it is likely that following these suggestions, the damage surrogate could reach its true performance potential.

4.4 Computational Efficiency

One of the main strengths of deep learning is improved computational efficiency over traditional physics-based solvers. Inference times were calculated for the two FNO based surrogate models as an average time across all unseen samples. While GPU-based inference is faster than CPU-based inference, both were calculated for fair comparison against the FFT simulations which were carried out using a CPU. The average inference times per sample are shown below in **Table 3** with the surrogate models clearly demonstrating superior computational efficiency compared to the FFT simulations. The stress surrogate reduces inference times by an order of magnitude while the damage surrogate shows reductions of three to four orders of magnitude. The final damage prediction of the FFT has a large computational cost, therefore this improvement in computational efficiency is extremely significant.

Table 3 - *Average inference time per sample.*

	Elastic Prediction (s)	Damage Prediction (s)
FNO Surrogate Model (CPU)	2.62×10^{-1}	4.65×10^{-1}
FNO Surrogate Model (GPU)	1.29×10^{-2}	1.18×10^{-2}
FFT Simulation (CPU)	1.82×10^{0}	1.90×10^{2}

5 Discussion

As with all data-driven deep learning methods, there is a significant requirement for large, high-quality, and diverse datasets for training and evaluation. While already created for a previous project, 7200 training and testing samples were produced through FFT simulations. With complex systems of PDEs creating training samples can be computationally expensive and therefore may not always be feasible. Therefore, finding ways to reduce model size and the quantity of required data is important. The ability of the FNO to perform zero-shot super-resolution goes some way towards this by allowing the model to be trained using lower resolution data; however, further reduction could be achieved by combining the FNO with numerical solvers. By taking advantage of the strengths of different modelling techniques, the development of hybrid modelling frameworks offers the potential of enhanced prediction capabilities and further improved computational efficiency.

Future research efforts should explore the potential scope of the FNO in addressing a range of additional challenges in the field of computational materials science. By continuing to develop FNO based models, researchers have the potential to unlock new insights into material behaviours for the development of advanced engineered materials.

The inclusion of Weights and Biases to the training process was extremely effective at providing a robust data logging mechanism. The use of sweeps efficiently automated aspects of the training process; however, the implementation used as part of this study has room for future development. The F1 score calculated at the end of each run was the result of the training and testing data. This meant that monitoring model overfitting was challenging as high scoring runs could be selected as the best model, despite later performing poorly on unseen data. In future, the inclusion of unseen data to the sweep would provide a more accurate representation of model performance, and therefore further increase the efficiency and effectiveness of the hyperparameter tuning process.

It is worth noting that currently two surrogate models are required to predict both stress and damage fields. However, both surrogates use the same three-channel input to produce three-channel and single-channel output for stress and damage respectively. It would therefore be possible to combine the two models to produce a single model with a three-channel input and a four-channel output. This was experimented with during the development of the two surrogate models, however training and hyperparameter tuning became increasingly complex. While beyond the scope of this report, further development of this combined stress and damage model, if successful, would further increase the computational efficiency of the predictions while demonstrating the power of the FNO when applied to stress field and damage predictions within composites.

6 Conclusions

This study has demonstrated the power of FNO based surrogate models for the prediction of stress and damage fields in composite materials. The surrogate models offer accurate and computationally efficient alternatives to traditional physics-based solvers. Through comprehensive testing and analysis, the surrogate models have shown their ability to accurately capture complex microstructure geometries, at both a local and global scale, to accurately predict material behaviours. The models successfully predicted stress fields and damage maps for a diverse range of composite microstructures and their impressive performance highlights their future potential for a range of engineering applications.

The exploration of performing zero-shot super-resolution has shown the potential of the FNO in performing such a feat. Further development of the training parameters of the stress surrogate will hopefully improve on the already impressive accuracy demonstrated when making high-resolution predictions having only ever seen low-resolution training data. While the results of the damage surrogate in this area were less impressive, the demonstration that the FNO can perform zero-shot super-resolution for the prediction of stress fields and damage maps within complex composite microstructures expands its areas of use to include predicting behaviours in the field of solid mechanics, having previously only been demonstrated in the field of fluid mechanics. Therefore, there remains significant potential for FNO-based models to further demonstrate their performance by continuing to investigate and develop complex models such as the damage surrogate model.

Further to this, the study highlights the wider implications of adopting data-driven modelling techniques in computational materials science. Continuing to develop methods such as the FNO will allow researchers to accelerate the prediction process to develop and optimise advanced materials and explore new ideas and innovations more quickly.

However, it is still important to recognise some of the current limitations associated with the FNO and other data driven modelling techniques. These methods require large, high quality data sets to be created for their training and evaluation process. As well as this, there remains the balance between training speed and prediction accuracy. To effectively address these limitations, further research and development is required to optimise their architectures, including the exploration of hybrid modelling techniques.

To conclude, the research carried out as part of this study, and the findings discussed within this paper show the significant potential of the FNO and effectively lays the foundation for future research. This research can look to further push the capabilities of data-driven modelling techniques and continue to unlock insight into complex material behaviours. Continued innovation in this area will push the boundaries of material engineering and lead the way to a more technologically advanced future.

7 Acknowledgements

I acknowledge that this work is my own, and I used ChatGPT 3.5 [23] to support the development of the surrogate models by assisting with bug checking to ensure code robustness.

There are many people who deserve thanks for their assistance and support, not only through this work but also throughout my four years at the University of Bath. Firstly, I would like to thank my supervisor Yang Chen for his support and expertise during this project. His earlier research laid the foundations for this work and the use of his data accelerated my progress. Thanks also to my course peers who I have had the privilege to work alongside over the course of my degree. Finally, I would like to thank my family for their constant care and support; always available for advice and assistance, their impact on my work cannot be overstated.

8 References

[1] S. Niezgoda, A. Kanjarla and S. Kalidindi, “Novel microstructure quantification framework for databasing, visualization, and analysis of microstructure data,” *Integr Mater Manuf Innov 2,* p. 54–80, 2013.

[2] Z. Li, . N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart and A. Anandkumar, “Fourier Neural Operator for Parametric Partial Differential Equations,” Cornell University, 2021.

[3] J. D. Kelleher, Deep Learning, Cambridge: The MIT Press, 2019.

[4] A. Botchkarev, “Performance Metrics (Error Measures) in Machine Learning Regression, Forecasting and Prognostics: Properties and Typology,” *ArXiv,* vol. abs/1809.03006, 2018.

[5] R. Dosselmann and X. Yang, “A comprehensive assessment of the structural similarity index,” *SIViP,* vol. 5, pp. 81-91, 2011.

[6] J. Opitz and S. Burst, “Macro F1 and Macro F1,” *arXiv,* 2021.

[7] R. Sepasdar, A. Karpatne and M. Shakiba, “A data-driven approach to full-field nonlinear stress distribution and failure pattern prediction in composites using deep learning,” *Computer Methods in Applied Mechanics and Engineering,* vol. 397, no. 0045-7825, 2022.

[8] Y. Chen, T. Dodwell, T. Chuaqui and R. Butler, “Full-field prediction of stress and fracture patterns in composites using deep learning and self-attention,” *Engineering Fracture Mechanics,* vol. 286, no. 0013-7944, 2023.

[9] O. Ronneberger, P. Fischer and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” *CoRR,* vol. abs/1505.04597, 2015.

[10] A. Bhaduri, A. Gupta and L. Graham-Brady, “Stress field prediction in fiber-reinforced composite materials using a deep learning approach,” *Composites Part B: Engineering,* vol. 238, no. 1359-8368, 2022.

[11] Y. Shokrollahi, M. Nikahd, K. Gholami and G. Azamirad, “Deep Learning Techniques for Predicting Stress Fields in Composite Materials: A Superior Alternative to Finite Element Analysis,” *Journal of Composites Science,* 2023.

[12] S. Mohammadzadeh and E. Lejeune, “Predicting mechanically driven full-field quantities of interest with deep learning-based metamodels,” *Extreme Mechanics Letters,* vol. 50, no. 2352-4316, 2022.

[13] N. Ibtehaz and M. S. Rahman, "MultiResUNet : Rethinking the U-Net Architecture," 2019.

[14] Z. Yang, C. Yu and M. Buehler, "Deep learning model to predict complex stress and strain fields in hierarchical composites," *Science advances,* vol. 7, 2021.

[15] K. Gholami, F. Ege and R. Barzegar, "Prediction of Composite Mechanical Properties: Integration of Deep Neural Network Methods and Finite Element Analysis," *Journal of Composites Science,* vol. 7, no. 2, 2023.

[16] T. Hoare and D. Kohane, "Hydrogels in drug delivery: Progress and challenges," *Polymer,* vol. 49, no. 8, pp. 1993-2007, 2008.

[17] M. Alom, T. Taha, C. Yokopcic, S. Westberg, P. Sidike, M. Nasrin, B. Esesn, A. Awwal and V. Asari, "The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches," *CoRR,* vol. abs/1803.01164, 2018.

[18] Y. Chen and J. Marrow, "Effect of irradiation swelling on the mechanical properties of unidirectional SiC/SiC composites: A numerical investigation at microstructural level," *Journal of Nuclear Materials,* vol. 569, no. 0022-3115, 2022.

[19] C. Miehe, M. Hofacker and F. Welschinger, "A phase field model for rate-independent crack propagation: Robust algorithmic implementation based on operator splits," *Computer Methods in Applied Mechanics and Engineering,* vol. 199, no. 45–48, pp. 2765-2778, 2010.

[20] L. Biewald, *Experiment Tracking with Weights and Biases,* 2020.

[21] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams and N. d. Freitas, "Taking the Human Out of the Loop: A Review of Bayesian Optimization," *Proceedings of the IEEE,* vol. 104, pp. 148-175, 2016.

[22] X. Chen, Z. Fu and J. Yang, "Zero-Shot Image Super-Resolution with Depth Guided Internal Degradation Learning," *Computer Vision – ECCV 2020. ECCV 2020. Lecture Notes in Computer Science,* vol. 12362, 2020.

[23] OpenAI, "ChatGPT," [Online]. Available: https://chat.openai.com/. [Accessed 2024].

9 APPENDICES

Appendix 1: Final Costing

Due to the nature of the work carried out as part of this project, no costs were incurred. All resources and facilities were either free to use or available through the University of Bath.

Appendix 2: Code

All code developed as part of this project has been uploaded to GitHub and can be found at the link below. This repository also contains some of the best performing models as well as their associated config files.

https://github.com/Custom-Url/StressDamageFNO