Theoretical and Experimental Forecasting of Light-Matter Interaction Dynamics Through Different Architectures of Physics-Informed Neural Networks

Hakob Janesian

Vahan Yeranosyan

Supervisor: Aleksandr Hayrapetyan

Submitted for the
degree of

BS Data Science
College of Science and Engineering
American University of Armenia

Yerevan 2024

Abstract

This paper explores the application of Physics-Informed Neural Networks (PINNs) in both experimental and theoretical physics, as evidenced by a comprehensive review of the relevant literature. It includes numerical simulations and real-world experiments conducted on synthetic and empirical datasets. The synthetic dataset was generated using a numerical method with a fixed seed to ensure reproducibility. Specifically, the focus is on light-matter interaction dynamics under varying external conditions, such as different electric fields. Various neural network architectures were employed to construct PINNs, which were then used to predict the complex dynamics of physical phenomena, such as the liquid crystal behavior of molecules during reorientation. The architectures utilized range from simple deep neural networks to more computationally intensive dense networks, with the latter providing more precise results despite being heavy for training. The paper presents the outcomes associated with different initial parameters of the used architectures, highlighting the advantages and drawbacks of each approach. Due to the extensive data involved in the experimental part of the research, the experimental data, being heavy, can be found following this link: MEGA link. Additionally, all capstone project-related files, including Jupyter notebooks, Python files, result images, and videos, are available in this GitHub repository: GitHub Repository.

Contents

Chapter 1

Introduction

As modern science progresses, it reveals solutions to complex problems such as big data analysis, prediction of dynamic systems, and the application of these results across various industries, particularly in science. Experience has demonstrated that a scarcity of data leads to inadequate predictions, thereby restricting the use of various models in several scientific fields, including physics, dynamic systems predictions, biological systems behavior, and engineering systems modeling (Raissi et al., 2019). Notably, in experimental physics, a data deficiency creates barriers that limit the deployment of contemporary machine learning models, slowing the field's advancement. However, recent advancements in spatiotemporal deep learning have enabled scientists to develop models with high predictive power using a limited amount of training data. The primary strength of these models lies in their incorporation of the ordinary and partial differential equations observed in phenomena. These equations, often called state equations, describe numerous physical systems and enhance the robustness and predictive accuracy of the models (Raissi et al., 2017). Known as physics-informed neural networks (PINNs), these models have gained widespread acceptance and usage in scientific research for their precise dynamic predictions and potential to substitute some numerical methods. The distinct nature of these models stems from their integration of the aforementioned differential equations into the model's loss function, which provides flexibility and insights into the underlying physical dynamics.

In the scope of this research, the architecture of the models previously mentioned was utilized to predict the dynamics of complex physical systems involving

light-matter interactions. The experiment involved observing the dynamics of a homeotropic-aligned liquid crystal cell covered with bismuth silicate oxide (BSO) crystal under the influence of an external field. This setup included a laser directly illuminating a digital micromirror device (DMD) screen, which in turn reflected the laser beam onto the liquid crystal (Barboza et al., 2012). BSO is characterized as a yellowish, high-efficiency photoelectric photoconductor with robust performance. Homeotropic alignment in liquid crystals occurs when rod-like molecules orient perpendicularly to the substrate surfaces. A digital micromirror device (DMD) is a semiconductor-based optical switch utilizing thousands of tiny mirrors to precisely control light patterns in applications such as projectors and optical imaging systems (Carvalho et al., 2020). The components described form a nonlinear optical system, supplemented by other optical elements and a special camera to record the internal dynamics of the liquid crystal, providing an experimental dataset (Fu et al., 2022). This dataset was employed to train the PINN models. Due to the computationally intensive nature of working with experimental data compared to synthetic data, the precision of these models was also tested using synthetic data. To create a synthetic dataset, numerical methods were employed to analytically solve the well-known partial differential equation, the complex Ginzburg–Landau (CGL) equation (Aranson and Kramer, 2002). This equation describes the behavior of dynamic systems undergoing nonequilibrium transitions, capturing the dynamics of amplitude and phase of waves in various physical contexts. The following chapters will delve deeper into the content above and further explore the results.

Chapter 2

Literature Review

2.1 Introduction to Physics-Informed Neural Networks

As previously outlined, the core of this research involves Physics-Informed Neural Networks (PINNs), which merge deep learning with physics through the utilization of deep neural networks and state equations (Raissi et al., 2019). The essential concept focuses on embedding these state equations directly into the neural network's loss function. This method is effectively supported by advanced computational frameworks, notably Pytorch, using its Autograd library to model these equations. The specific problem addressed exhibits complex spatiotemporal behavior and has the following form:

$$u_t + \mathcal{N}[u] = 0,\ x \in \Omega,\ t \in [0, T] \tag{2.1}$$

Where $\mathcal{N}[\cdot]$ is a nonlinear differential operator, $u(t, x, y)$ denotes the hidden solution, and Ω is a subset of $\mathcal{R}^D$. The hidden solutions are generally derived with the help of various numerical methods. In this research case, the exponential time differencing method was utilized in conjunction with the fourth-order Runge-Kutta method (García-Morales and Krischer, 2012). So, the described methods are nothing else but discrete time models. Finally, the derived state equations should be integrated into the PINNs' loss function, which has the general form below.

$$\mathcal{L} = \mathcal{L}_{\mathcal{NN}} + \mathcal{L}_{physics} \tag{2.2}$$

Where $\mathcal{L}_{\mathcal{NN}}$ represents the MSE measuring the difference between the actual and predicted states, and $\mathcal{L}_{physics}$ represents the loss associated with the actual physical system. The physical system loss has the following form

$$\mathcal{L}_{physics} = u_t + \mathcal{N}[u] \tag{2.3}$$

The NN's loss function has the following form

$$\mathcal{L}_{NN} = \frac{1}{N}\sum_{i=1}^{N} |u(t_i, x_i, y_i) - u_i|^2 \tag{2.4}$$

2.2 PINNs different architecture

During the research, various architectures were explored, including PINNs with underlying fully connected and dense networks. Below, the fully connected neural network embedded into the PINN is presented.

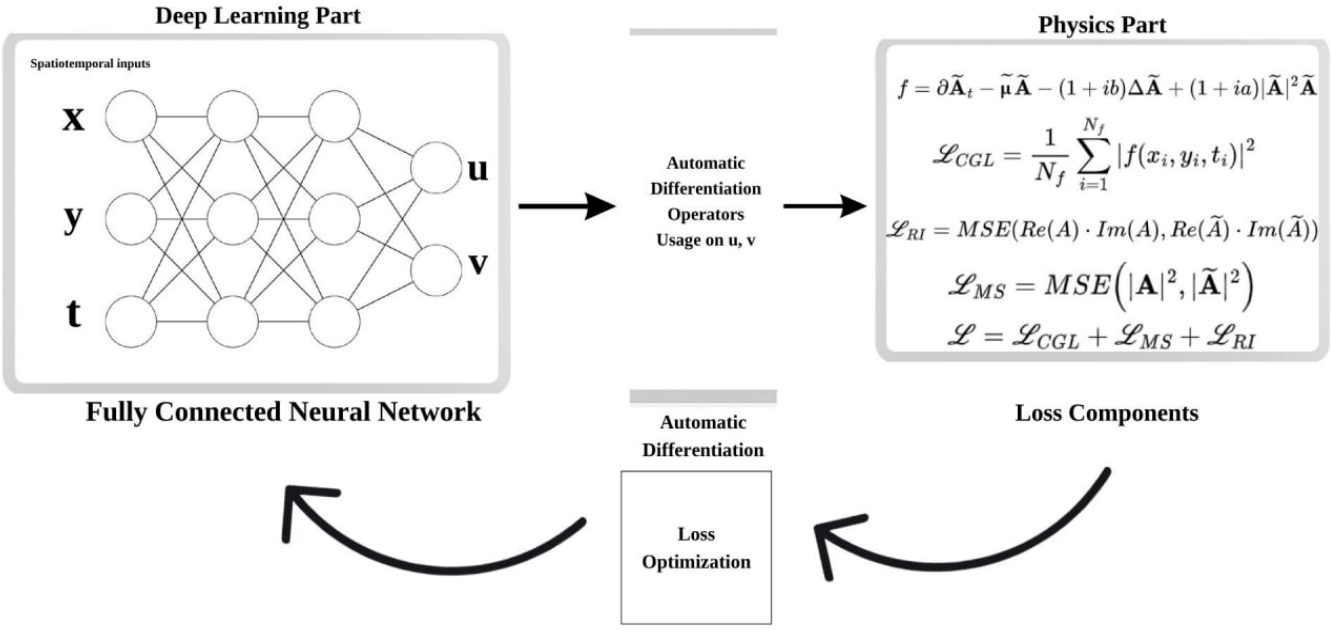

Figure 1: PINN Training Cycle

The schema illustrates the learning flow of the constructed PINN model, highlighting key components: the neural network, the application of automatic differentiation operators on network outputs, the inclusion of physics-informed loss components, and the mini-batch gradient descent process.

The DenseNet (Zhu and Newsam, 2017) architecture that is utilized in our experiments is presented in the provided schema below.

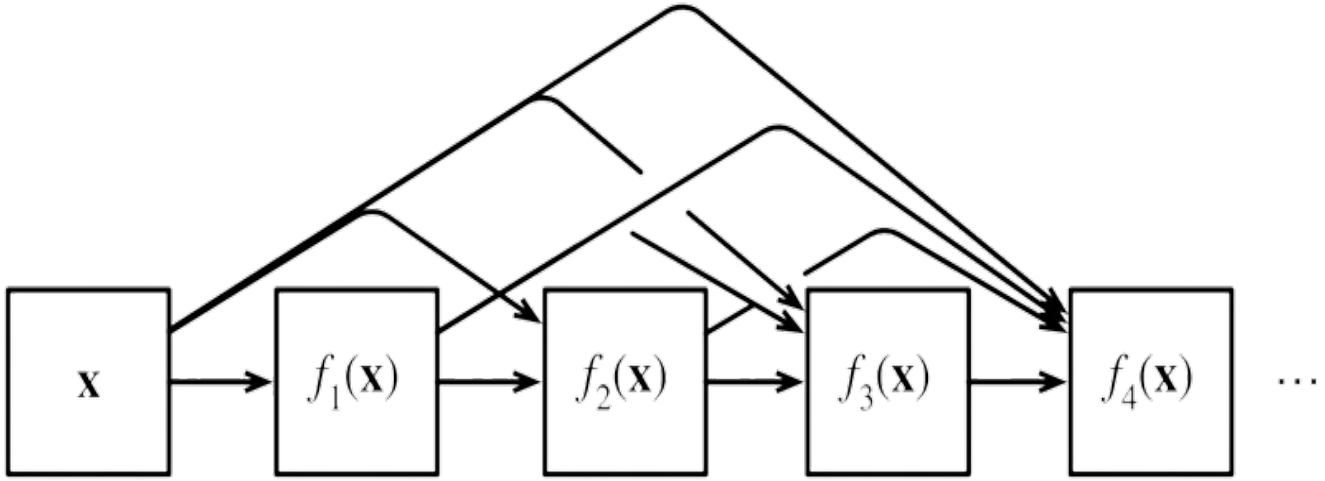

Figure 2: DenseNet Architecture

The primary benefit of DenseNet is its capability to predict complex dynamical systems, as each layer connects with all preceding ones. This architecture clearly outperforms the Fully Connected Neural Network, as detailed in the results chapter.

2.3 Light-Matter Interaction: Theoretical & Experimental Interpretations

Light-matter interaction describes how electromagnetic radiation affects matter, specifically influencing the partial reorientation of liquid crystal molecules (Goldstein, 2017). This interaction is crucial for numerous modern technologies and natural phenomena, including photosynthesis and optical communications. The study explores this interaction by simultaneously addressing both theoretical modeling and experimental aspects. Theoretical investigations involved modeling

the interaction between light and liquid crystals, as detailed in Section 2.2, and numerically solving the state equation to meet initial and boundary conditions. This approach enhances the understanding and prediction of experimental outcomes without their physical execution. Additionally, it facilitates the collection of synthetic datasets for training future models.

On the other hand, liquid crystal and light interactions can be examined experimentally using nonlinear optical systems and laser experiments. These are crucial for validating model predictions and generating new empirical datasets, which serve as complex training sets for models.

Numerous state equations describing light-matter interactions have been discovered over the past decades. This study focuses on the CGL equation (Aranson and Kramer, 2002), pivotal in providing a mathematical framework for describing phase transitions in light and liquid crystal interactions within optical nonlinear systems. It has the following form:

$$\partial_t A = \mu A + (1 + ib)\Delta A - (1 + ic)|A|^2 A \tag{2.5}$$

Where A is a complex function of time and space, b and c are the real parameters characterizing linear and nonlinear dispersion, and μ is a bifurcation parameter. CGL equation models the dynamics of complex fields, including but not limited to the effects of nonlinearity, coherence, and dispersion, based on the macroscopic properties of materials.

Chapter 3

Methodology

3.1 Synthetic & Experimental Datasets Representation

Before discussing the models, an explanation of the datasets utilized is necessary. As indicated in section 2.3, synthetic and experimental datasets were created. This section delves deeper into the synthetic dataset's creation and usage. The synthetic dataset was generated using various numerical methods, including exponential time differencing and fourth-order Runge-Kutta, which aided the analytical solution of the CGL equation. In the Fourier space, the equation has the following form.

$$\partial_t \hat{A} = (1 - k^2(1 + bi))\hat{A} - \mathcal{F}[(1 + ic)|A|^2 A] \tag{3.1}$$

where $\hat{A}$ is the representation of the state in the Fourier space. Let's denote the first part of the right-hand side with q and the second part of the right-hand side with $\mathcal{N}$. After applying the aforementioned numerical schema, the following approximation for the CGL equation will be observed:

$$\hat{A}_{n+1} = \hat{A}_n e^{qh} + \mathcal{N}_n \frac{(1 + hq)e^{qh} - 1 - 2hq}{hq^2} + \mathcal{N}_{n-1} \frac{-e^{qh} + 1 + hq}{hq^2} \tag{3.2}$$

The numerical schema illustrated demonstrates the computational process of the

CGL equation using specified methods. The corresponding code schema inputs simulation-related parameters such as the time domain, spatial coordinates, grid point numbers, and constants specific to the equation, ultimately yielding the analytical solution for the CGL equation within the defined domains. Although some solutions include synthetically created diffusion effects, such as planar flows, adjusting certain parameters to meet well-known criteria can induce spatiotemporal chaos. The generation of synthetic data occurs through the sequential execution of specified numerical methods, ensuring that identical datasets are reproduced when the codes are run with unchanged initial and boundary conditions. The figure below displays generated data points that exhibit planar deterministic flows and non-equilibrium chaos.

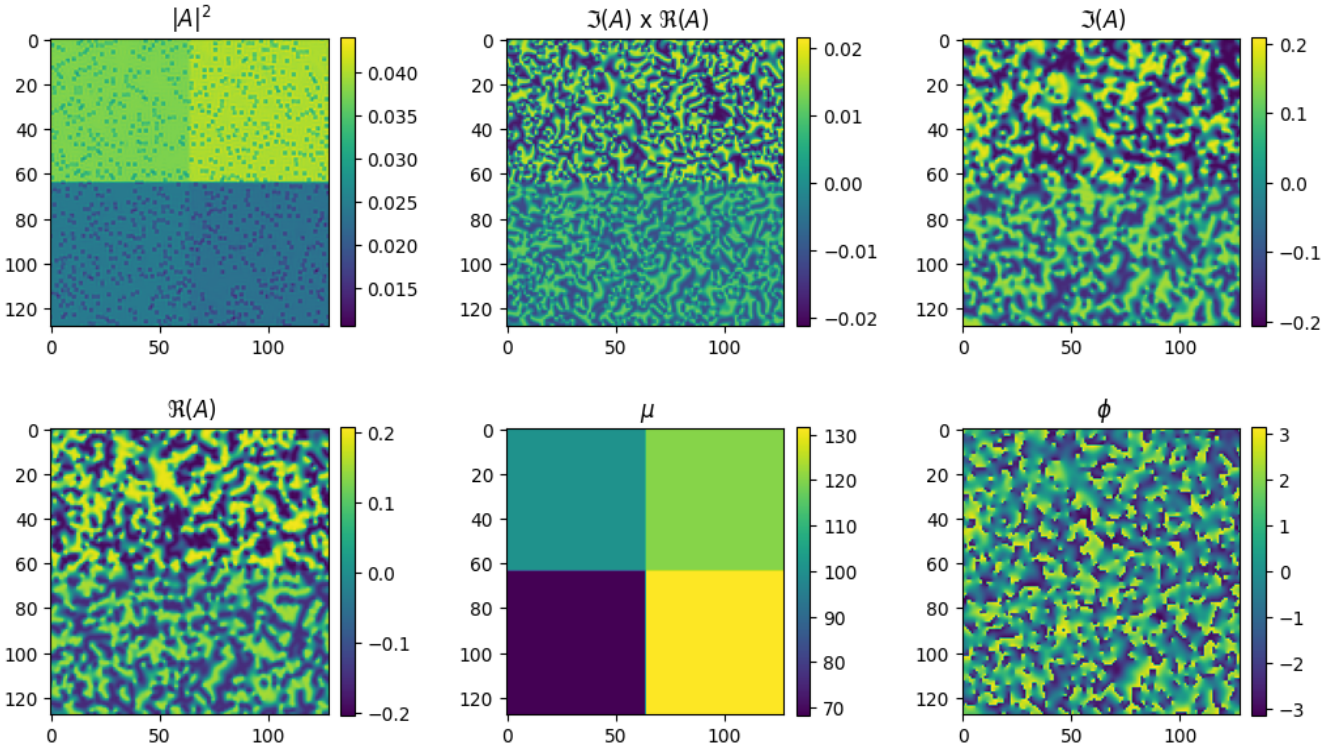

Figure 3: Synthetic Dataset

Experimental datasets were collected during real experiments utilizing a nonlinear optical system with a BSO liquid crystal cell. A series of experiments captured spatio-temporal datasets from interactions between light and the liquid crystal cell under an external field. A specialized polarization camera recorded these interactions, storing the information on a computer as a sequence of frames. Ultimately, all collected datasets were used to train models.

3.2 Experimental Dataset Mapping to Digital States

The experimental analysis yielded intensities under different polarization angles: 0 degrees, 45 degrees, 90 degrees, and 135 degrees. These intensities correlate with the state of the liquid crystal based on established theoretical relationships. A specialized polarization camera was utilized to capture these intensities. This camera includes numerous micro-polarizer filter arrays; each 2x2 pixel array records a specific intensity, with each pixel dedicated to a single intensity reading. The key relationship connecting the intensities to the liquid crystal state involves the azimuthal angle (ψ), and theta, which represents the projection angle of the director on the x-y plane.

$$n = \{\cos\psi \sin\theta, \sin\psi \sin\theta, \sqrt{1 - \cos^2\psi \sin^2\theta - \sin^2\psi \sin^2\theta}\} \tag{3.3}$$

The azimuthal and projection angles have the following forms below.

$$\psi = \arctan\left(\frac{n_y}{n_x}\right) \tag{3.4}$$

$$\theta = \sqrt{\frac{\Delta}{k}} \tag{3.5}$$

where Δ is the phase retardance, and k is the wave number, and n_x and n_y are the projection of the liquid crystal on the x-y plane. In conclusion, the CGL state can be constructed from the previously mentioned relations in the following form.

$$A = n_x + in_y \tag{3.6}$$

3.3 Utilized Loss Functions for PINNs

This section discusses the loss functions utilized within the scope of this research. The general form of the custom loss function was introduced in Section 2.1, with experimentation conducted on various custom loss functions tailored to

specific problem needs. Initially, the following loss function was employed:

$$\mathcal{L} = \mathcal{L}_{NN} + \mathcal{L}_{physics} \tag{3.7}$$

where $\mathcal{L}_{NN}$ is defined as in Equation (2.4), and $\mathcal{L}_{physics}$ is given by:

$$\mathcal{L}_{physics} = \frac{1}{N_f} \sum_{i=1}^{N} |f(t_i, x_i, y_i)|^2 \tag{3.8}$$

The function f is described by:

$$f = \mu A - \Delta A + |A|^2 A \tag{3.9}$$

This equation represents the residuals of the CGL equation with coefficients $b = c = 0$. This is the simplest loss function utilized. However, this form sometimes fails to accommodate all the flow fluctuations that the data might exhibit. The capabilities of the custom loss function have been enhanced through the incorporation of additional components, thus making the final loss function more robust. The augmented loss function is expressed as the following loss function:

$$\mathcal{L} = \mathcal{L}_{RI} + \mathcal{L}_{MS} + \mathcal{L}_{physics} \tag{3.10}$$

Here, $\mathcal{L}_{RI}$ represents the Mean Squared Error (MSE) between the product of the real and imaginary parts of the target dynamics and the predicted state, and $\mathcal{L}_{MS}$ refers to the MSE for the modulus squares of the target and predicted states. Given that the state in the CGL equation is complex, the configuration of these loss functions provides more insights into the underlying dynamics, enhancing the robustness of the models. In the context of physics, computed quantities such as the modulus square of the state convey intensity information, thereby increasing the flexibility of the models. Furthermore, the $\mathcal{L}_{RI}$ component supplies all necessary information about the state dynamics under certain boundary conditions.

3.4 FCNN as an Underlying Neural Network for PINN

This section details the first of two distinct PINNs developed, each with different underlying architectures. The focus here is on the FCNN, which comprises interconnected neurons in successive layers. This structure ensures comprehensive information processing from input to output, with each neuron connection characterized by a weight that adjusts during training to minimize output prediction errors. FCNNs are typically employed in classification, regression, and pattern recognition tasks where input features exhibit significant interdependence.

This research explored various configurations of neurons, hidden layers, and activation functions. The hyperbolic tangent was initially employed as the nonlinear activation function across all network layers. However, due to its instability - which impeded accurate predictions of the complex dynamics between liquid crystals and light - the Softplus function was adopted instead. The Softplus function, defined as

$$SoftPlus(x) = \frac{1}{\beta} \log\left(1 + e^{\beta x}\right) \tag{3.11}$$

with a default $\beta = 1$, serves as a smooth approximation of the widely recognized Rectified Linear Unit (ReLU) function. This function ensures that the output remains positive and conducive to the needs of our model.

Figure 1 illustrates the workflow of our FCNN-based model. The neural network takes spatiotemporal parameters as inputs and outputs the estimated states of the CGL equation. Subsequent stages involve computing necessary partial derivatives from these approximations. These computed values are then utilized in the custom loss computation phase, which incorporates residuals of the CGL equation to form a comprehensive loss function. This function encapsulates both the neural network's functionality and the underlying physical processes.

Model training is bifurcated into two stages. The initial phase seeks to identify states that satisfy the CGL equation without incorporating external field information, focusing solely on the internal dynamics. The second training phase integrates bifurcation parameter data to facilitate data-driven discovery. This pa-

rameter, dependent on spatiotemporal coordinates, plays a critical role in defining the system's behavior under varying conditions. Given the extensive datasets available, all training sessions were conducted in batches, each comprising 10000 data points, optimizing the custom loss function to identify bifurcation parameters that satisfy the CGL equation. This training methodology is consistently applied across all models, with outcomes presented in Section 4.1.

3.5 DenseNet as an Underlying Neural Network for PINN

This architecture is particularly advantageous for uncovering the latent dynamics within the data. Due to its unique design, which features connections from each layer to all subsequent layers, DenseNet facilitates comprehensive information flow through the network. This type of connectivity enhances the robustness and flexibility of the model.

The inputs to the DenseNet are identical to those of the FCNN described earlier. Similarly, the output remains consistent with the FCNN, generating approximated states for the CGL equation. Notably, the states produced by DenseNet architecture demonstrate a closer alignment with the empirical ground truths, enhancing the model's predictive accuracy.

An instance of this network can have the following configuration: a single input layer, three hidden layers, and an output layer. The input layer processes spatiotemporal coordinates comprising spatial (x and y) and temporal (t) components by concatenating them into a unified 3D tensor.

The network's output is similarly structured as a 3D tensor, maintaining the same spatiotemporal characteristics as the input. This architecture shares the data flow process with the FCNN, involving the computation of derivatives and initiating a data-driven discovery process. The performance and results of this network are detailed in Section 4.2.

Chapter 4

Results & Discussions

4.1 FCNN-PINN Results

This section presents significant outcomes from numerous experiments conducted with the FCNN-based PINN. Both synthetic and experimental datasets served as inputs. The synthetic dataset exhibits specific characteristics, and the architectural details of the model are further elaborated. Figure 4 illustrates the loss curves under different learning rates.

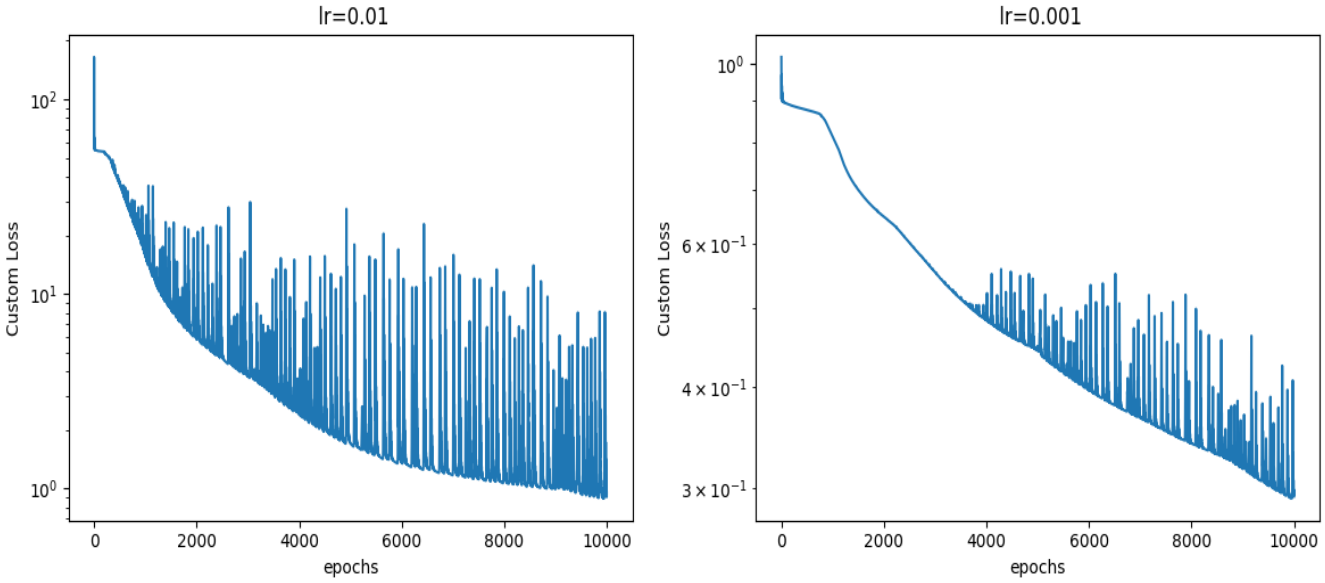

Figure 4: FCNN Losses

Figure 5, presented below, displays comparative visualizations of the original dataset alongside predictions made by the model. This offers a clear depiction of the network's predictive accuracy and its architectural effectiveness in handling

complex datasets.

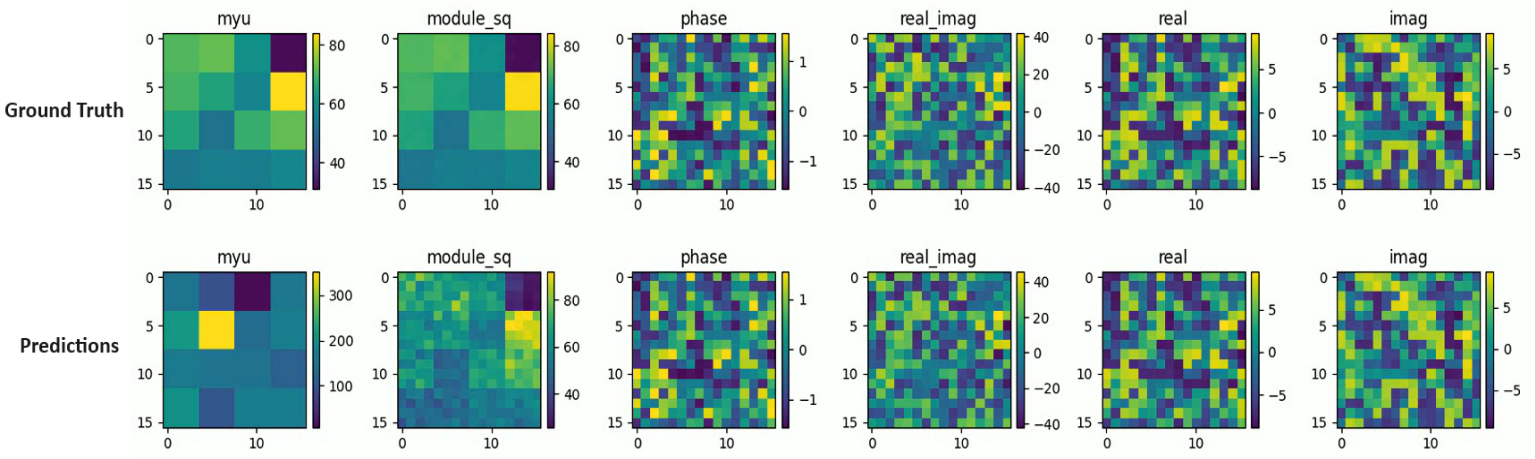

Figure 5: FCNN Results

In the Figure 6 below is shown the learning curve of the μ training

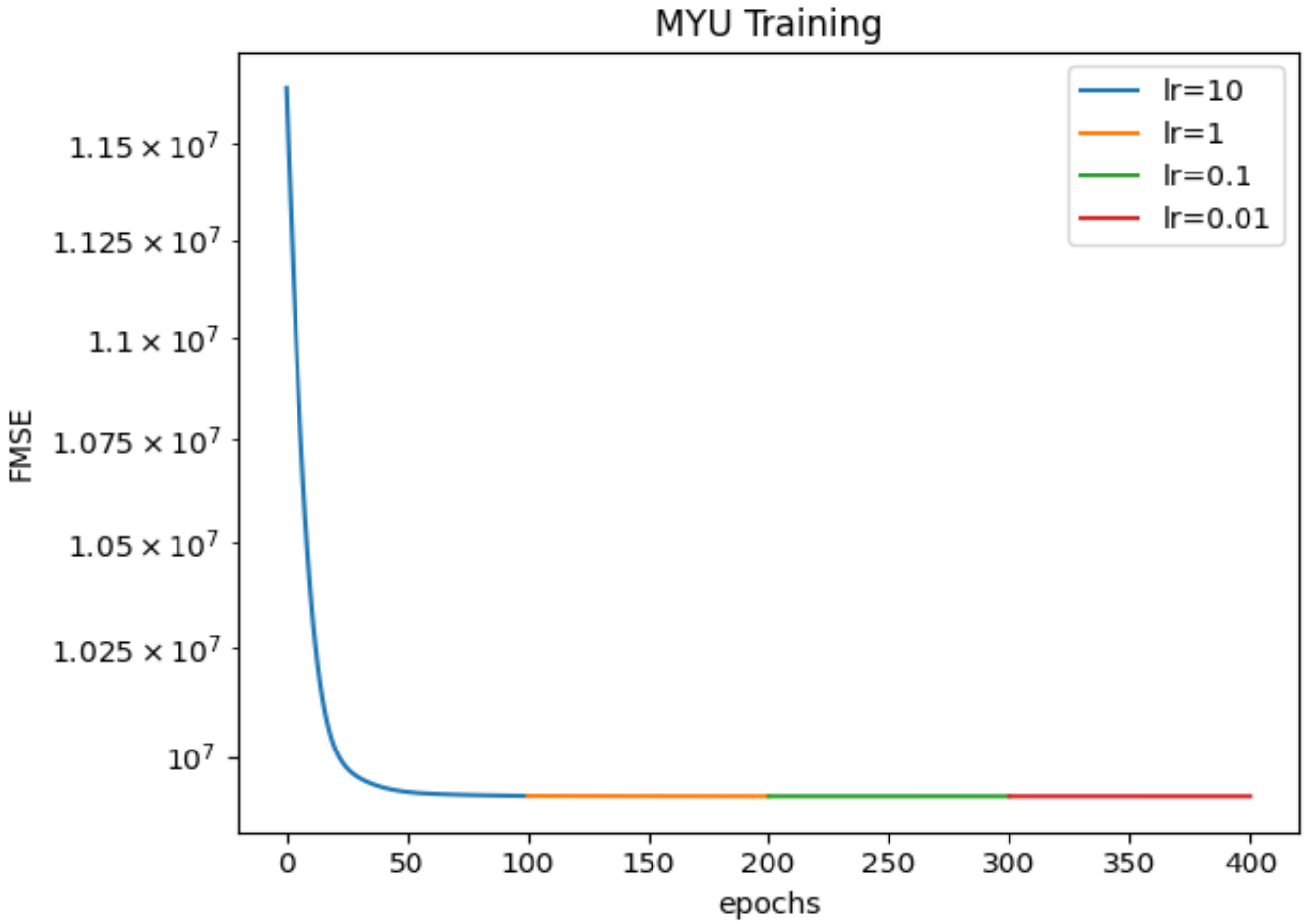

Figure 6: FCNN bifurcation parameter training curve

Different datasets were used, yielding the results shown below in Figure 7.

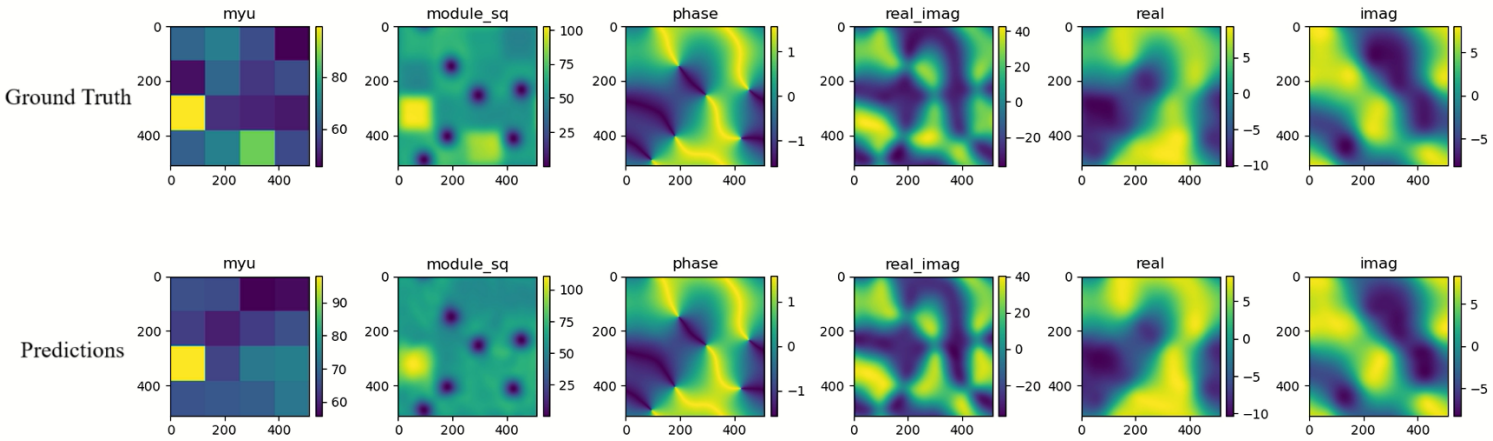

Figure 7: FCNN-PINN Prediction on Different Dataset

Figure 8 shows the learning curves for the above dataset in conjunction with the μ training.

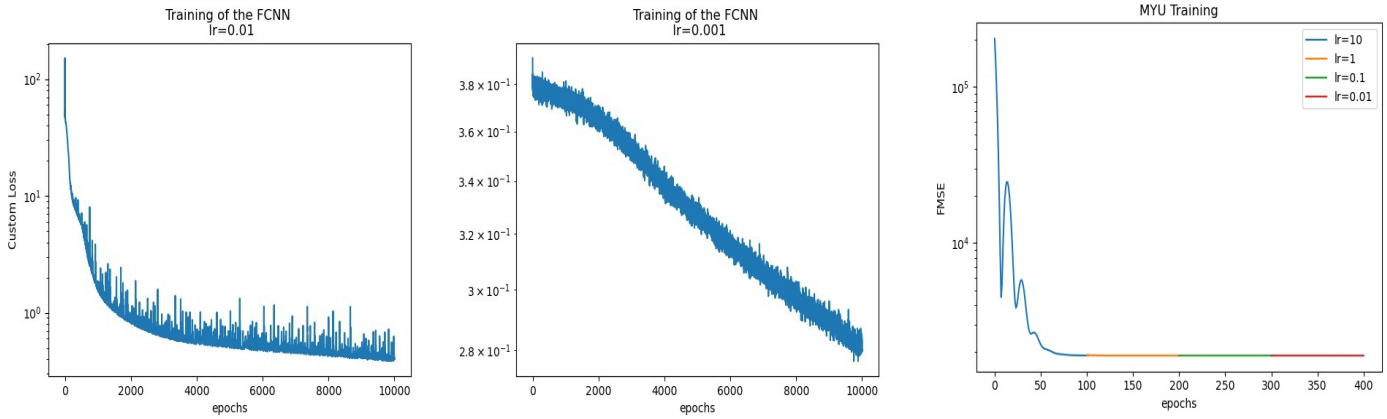

Figure 8: FCNN-PINN Learning Curves with the μ Learning Curves

Figure 9 shows the prediction results on the experimental dataset.

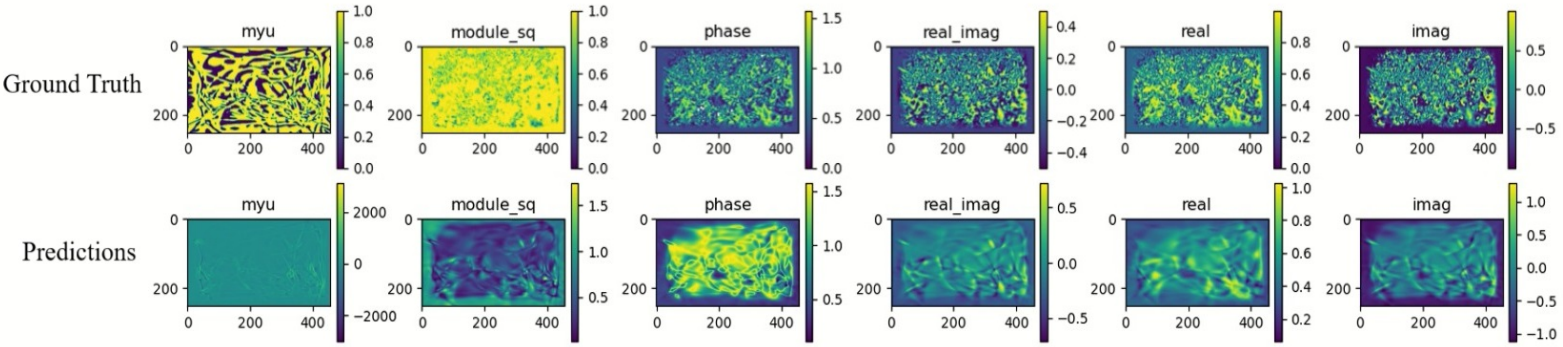

Figure 9: FCNN-PINN Predictions on the Experimental Dataset

As expected, the results on the experimental dataset will not be as good as those of the synthetic data case. The evaluations of these results will be presented in section 4.3.

4.2 DenseNet-PINN Results

Results from multiple experiments conducted using the DenseNet-PINN are presented here, and synthetic and experimental datasets are used. The synthetic dataset exhibits specific characteristics. Below, Figure 10 displays a visualization comparing the original dataset with predictions made by this model.

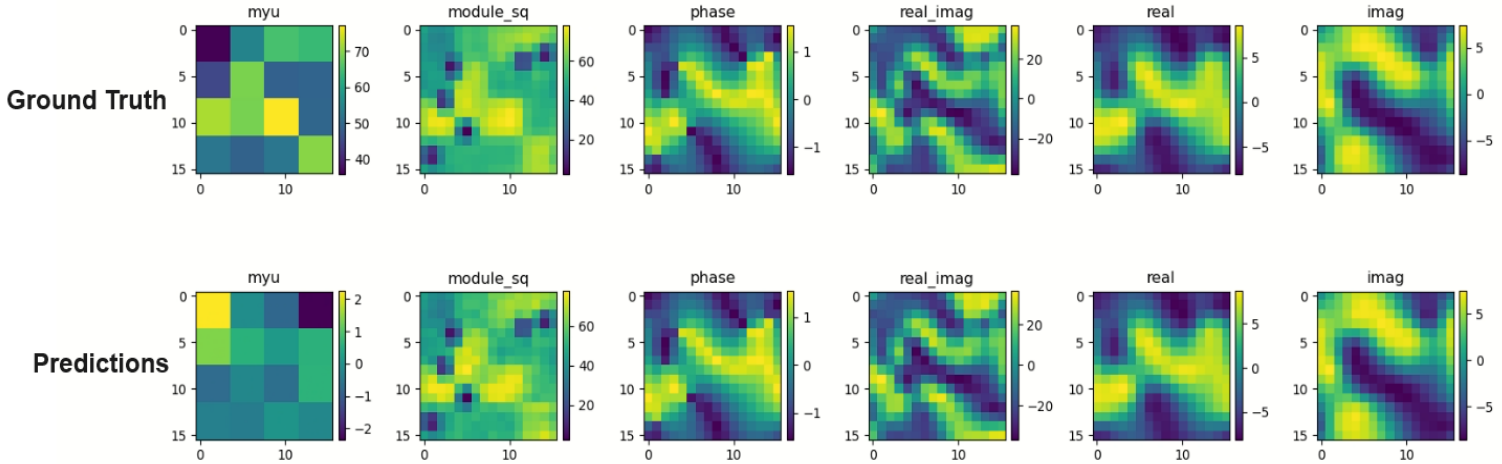

Figure 10: DenseNet-PINN predictions on the synthetic dataset

Figure 11 shows the learning curves of the model with the μ parameter learning curves.

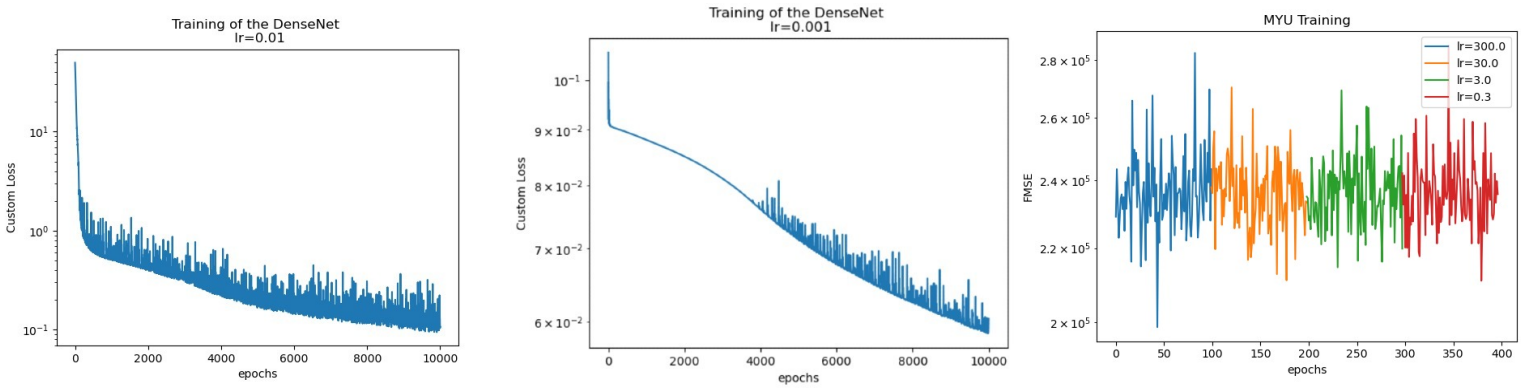

Figure 11: DenseNet-PINN learning curves with the μ learning curves

Results on the synthetic dataset are presented below in Figure 12.

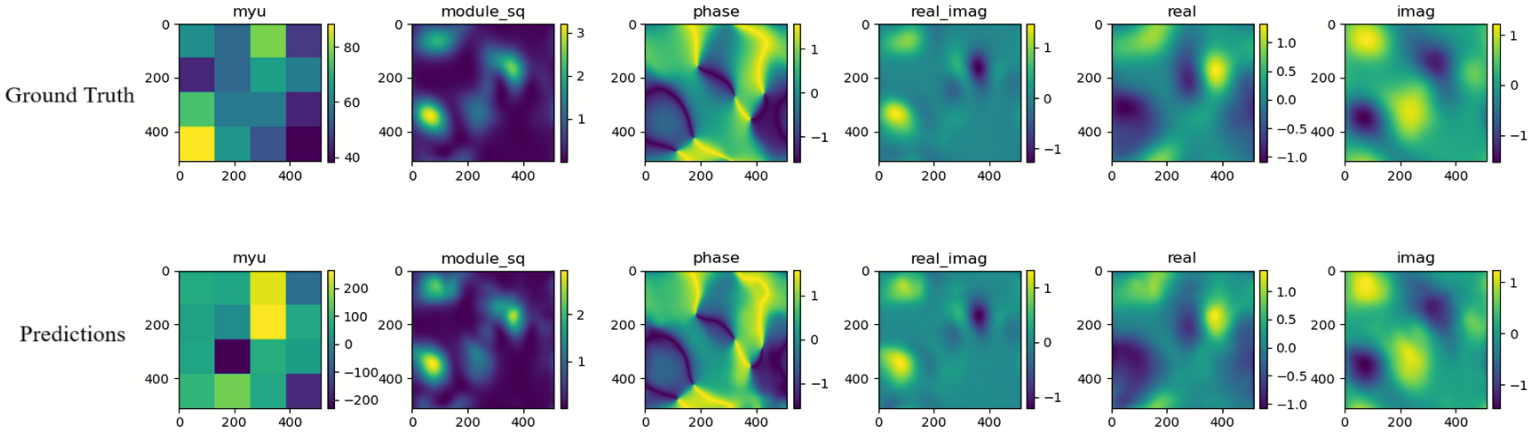

Figure 12: DenseNet-PINN prediction on the bigger dataset

As previously mentioned, the results of the empirical dataset do not match the quality of those from the synthetic dataset. The results are shown in Figure 13.

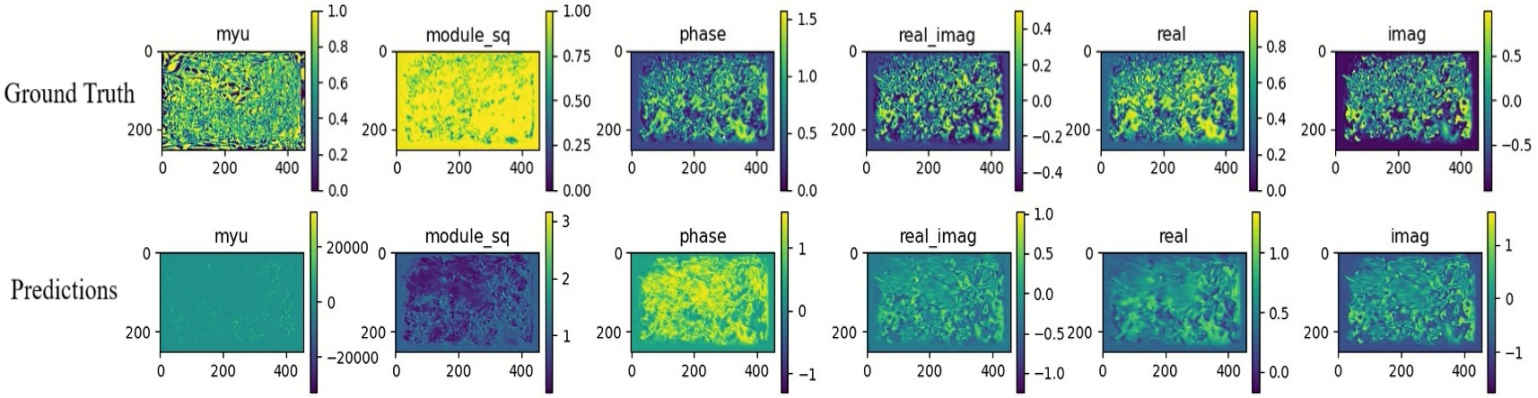

Figure 13: DenseNet-PINN prediction on the experimental dataset

The evaluations of these results are detailed comprehensively in section 4.3 of the document. This section will provide an in-depth presentation and analysis of the findings.

4.3 Results Evaluations

The analyses presented in Sections 4.1 and 4.2 found that the DenseNet-PINN outperformed the FCNN-PINN. The simpler architecture of the FCNN-PINN produces predictions that approximate the ground truth closely, yet some discrepancies remain. This observation is also applicable to the network's learning process. The learning curves indicated a reduction in loss, although it did not converge to the local minimum of the function. In scenarios of lower complexity, the performance of this network was deemed satisfactory. However, in more intricate scenarios, its efficacy was inferior to the subsequently introduced DenseNet-PINN.

In contrast, the DenseNet-PINN consistently outperforms the FCNN-PINN across all aspects and datasets. This indicates that the predictions generated by this model more closely resemble the original ground truth data. Specifically, this neural network predominantly captures the complex dynamical fluctuations observed in both synthetic and experimental datasets. The intricate structure of the DenseNet-PINN, characterized by its dense connections, facilitates complex predictions. Consequently, this architecture proves to be more suitable for dynamic prediction tasks, as demonstrated within the scope of this research.

Chapter 5

Conclusion & Future Work

Based on the insights derived from the analyses presented in the scope of this capstone section, it is appropriate to conclude that PINNs adeptly encapsulate the intricate dynamics inherent to physics-related data. These models demonstrate a commendable capacity to manage the inherent fluctuations of the data and yield robust predictions. Comparative evaluations indicate that the DenseNet architecture model outperforms the FCNN model. This advantage likely stems from the complex nature of the underlying flow dynamics beyond the representational capacity of simpler, fully connected architectures. The superior architecture benefits from enhanced memory capabilities, facilitating a deeper understanding of complex data patterns.

Both model architectures were evaluated using synthetic and experimental datasets. Consistent with previous observations, model performance was superior on synthetic data, attributed to the relative simplicity and controlled variability of these datasets compared to their experimental counterparts. Experimental datasets typically incorporate numerous nonlinear elements, rendering them more complex and mimetic of real-world conditions. Despite sharing common dynamic features, the flow characteristics in the synthetic datasets are predominantly influenced by specified simulation parameters. In contrast, experimental datasets are contingent upon many factors, including the applied electric fields, laser intensity, and the operational parameters of polarization cameras.

Thus, the dynamism exhibited by the experimental datasets is substantially greater than that of the synthetic datasets, complicating the task of making accurate predictions. The analysis confirms that models are more effective when applied to synthetic data, as opposed to experimental datasets where the prediction of flow dynamics poses significant challenges.

Looking ahead, future research will involve conducting experiments with a diverse array of loss functions to identify those that perform optimally and provide deeper insights into the physics of the systems being modeled. Additionally, exploring various neural network architectures will be crucial to ascertain which configurations most effectively capture the complexities of physical datasets. The outcomes of these experiments should be meticulously compared with the findings discussed in previous sections. This iterative approach is anticipated to yield novel applications for traditional neural networks in the domain of scientific machine learning, particularly in predicting chaotic data flows in real physical systems.

Bibliography

Igor S Aranson and Lorenz Kramer. The world of the complex ginzburg-landau equation. *Reviews of modern physics*, 74(1):99, 2002. doi: https://doi.org/10.1103/RevModPhys.74.99.

R Barboza, U Bortolozzo, G Assanto, E Vidal-Henriquez, MG Clerc, and Stefania Residori. Vortex induction via anisotropy stabilized light-matter interaction. *Physical review letters*, 109(14):143901, 2012.

Francelino Freitas Carvalho, Carlos Augusto de Moraes Cruz, Greicy Costa Marques, and Kayque Martins Cruz Damasceno. Angular light, polarization and stokes parameters information in a hybrid image sensor with division of focal plane. *Sensors (Basel, Switzerland)*, 20(12), 2020.

Qiang Fu, Wei Yang, Haodong Shi, Yingchao Li, Su Zhang, Juntong Zhan, Jianan Liu, Chao Wang, Zhuang Liu, Yong Zhu, et al. Design and experiment of high-resolution multispectral polarization imaging system. *Applied Sciences*, 12(21): 10712, 2022.

Vladimir García-Morales and Katharina Krischer. The complex ginzburg–landau equation: an introduction. *Contemporary Physics*, 53(2):79–95, 2012.

Dennis H Goldstein. *Polarized light*. CRC press, 2017.

Maziar Raissi, Paris Perdikaris, and George Em Karniadakis. Physics informed deep learning (part i): Data-driven solutions of nonlinear partial differential equations. *arXiv preprint arXiv:1711.10561*, 2017.

Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse

problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.

Yi Zhu and Shawn Newsam. Densenet for dense flow. In *2017 IEEE international conference on image processing (ICIP)*, pages 790–794. IEEE, 2017.