Supplementary Material

1 Complex Kolmogorov Arnold Network

1.1 Explanation of the Limitations of KAN in Complex-Valued Fitting

In this section, we will derive the limitations of the classical complex-valued computation process in KAN and introduce how CKAN overcomes these limitations. Assume a KAN with inputs $x \in \mathbb{C}$; the output can be expressed by the following formula:

$$\Phi_n(\Phi_{n-1}(\dots(x))).$$

Among them, Φ_n represents the coupled one-dimensional spline function, whose hyperparameters are learned by backpropagation. The Kolmogorov–Arnold theorem, as the foundational theorem of KAN, states that multivariate continuous functions can be approximated by nested combinations of simpler functions. Therefore, KAN generalizes depth by coupling multiple continuous functions. However, even with any finite depth extension, KAN still exhibits limitations in the complex domain. We present the following theorem:

Theorem: Suppose the real and imaginary parts of x, each continuous, are input into a KAN where both the input and output are one-dimensional. The statement can in fact be generalized to multivariate functions under suitable continuity assumptions. If the KAN fits the real and imaginary parts separately, then the result after the $(n-1)$-th layer of the KAN is denoted as $\Phi_{n-1}(x) = (\Psi_1, \dots, \Psi_m)$, which is continuous and monotonic, and the structure of the final layer is represented as $\Phi_n = (\Phi_n^1, \Phi_n^2)$. When $\Phi_n^1(x)$ perfectly fits the target function, there exists a real polynomial $P_k(x)$, and $\Phi_n^2(x)$ cannot be approximated in the limit sense.

Proof: The idea of the proof is to fix $\Phi_n^1(\Phi_{n-1}(x))$, $\Phi_n^2(\Phi_{n-1}(x))$, while using $\Phi_{n-1}(x)$ as the input, which restricts the network's fitting capability. The proof process is established after the network training is completed. Φ_n^1 has already perfectly fit the target function, which can be arbitrary, and at this point, $(\Psi_1, \dots, \Psi_m)$ is fixed.

Given the monotonic continuous output $(\Psi_1, \dots, \Psi_m)$, construct a convolution form in the function space:

$$P_m^j(x) = \int Q_m(x-z)\Psi_j(z)\,dz,$$

where

$$Q_m = \int (1-x)^m dx,$$

and $P_m^j(x)$ is a polynomial. According to the approximation properties of polynomials and the monotonic continuity of $\Phi_{n-1}(x)$, $P_m^j(x)$ will uniformly converge to $\Psi_j(x)$.

Examining the function fitting space of $\Phi_n^2(\Phi_{n-1}(x))$, due to the uniform continuity of the spline function, $\Phi_n^2(\Psi_j(x))$ uniformly converges to $\Phi_n^2(P_m^j(x))$, and since the spline function is a polynomial function, $\Phi_n^2(P_m^j(x))$ is also a polynomial function.

Due to the property of uniform convergence, when $\exists M_j \in \mathbb{N}$ such that for $\forall m > M_j$, the convergence of $P_m^j(x)$ holds. However, examining all superscripts j, there exists a maximum polynomial order M such that for all $\forall j$, the order of $P_m^j(x)$ is less than M.

Furthermore, since Φ_n^2 is also a polynomial function and its order M_Φ is fixed once the network is set, we can construct an arbitrary polynomial of order $M_\Phi M+1$ that is not within the fitting space $\Phi_n^2(\Phi_{n-1}(x))$, meaning $\Phi_n^2(x)$ cannot be approximated. This is determined by the fundamental properties of polynomial bases.

1.2 Modification of KAN and Proposal of CKAN

It can be observed that the limitation of KAN regarding complex-valued fitting stems from overfitting in either the real or imaginary part. When one part is fitted too completely, it induces fitting limitations in the other part. From a network structure perspective, this occurs because both the real and imaginary parts are computed using the same portion of the network. Therefore, CKAN separates the real and imaginary parts into two pathways, thereby lifting the limitations of KAN. The specific calculation formula is as follows:

$$\phi_n(x) = \big(\phi_n^r(\mathrm{Re}\,\Phi_{n-1}^r(x)), \phi_n^i(\mathrm{Im}\,\Phi_{n-1}^i(x))\big).$$

Here, ϕ_n^r represents the pathway for the real part in one layer of the model, and ϕ_n^i represents the pathway for the imaginary part in one layer of the model. In practice, the output of each layer is separated into phase and amplitude, primarily to simulate the form of pulses.

1.3 Performance Test Results of CKAN

We designed the following experiment, using four models: CKAN, KAN 2.0, EFFKAN [1], and WavKAN [2]. EFFKAN removes the adaptive adjustment of the activation function in the original KAN, while WavKAN implements the KAN structure using wavelet functions. The four models will respectively distill a pre-trained standard PINN to evaluate the fitting capability of each model for pulses. The PINN solved for the standard single soliton, double soliton, and dark soliton of the NLSe (SFig 1a) and two single solitons of the GP equation [3]. The forms of NLSe and GPe are as follows:

$$iu_t - u_{xx} + 2|u|^2 u = 0;$$

$$\begin{cases} iu_t + \frac{1}{2}u_{xx} + 2f(t)(|u_1|^2 + |u_2|^2)u_1 + V_{\text{ext}}(x,t)u_1 - \sigma u_2 = 0, \\ iu_t + \frac{1}{2}u_{xx} + 2f(t)(|u_1|^2 + |u_2|^2)u_2 + V_{\text{ext}}(x,t)u_2 - \sigma u_1 = 0. \end{cases}$$

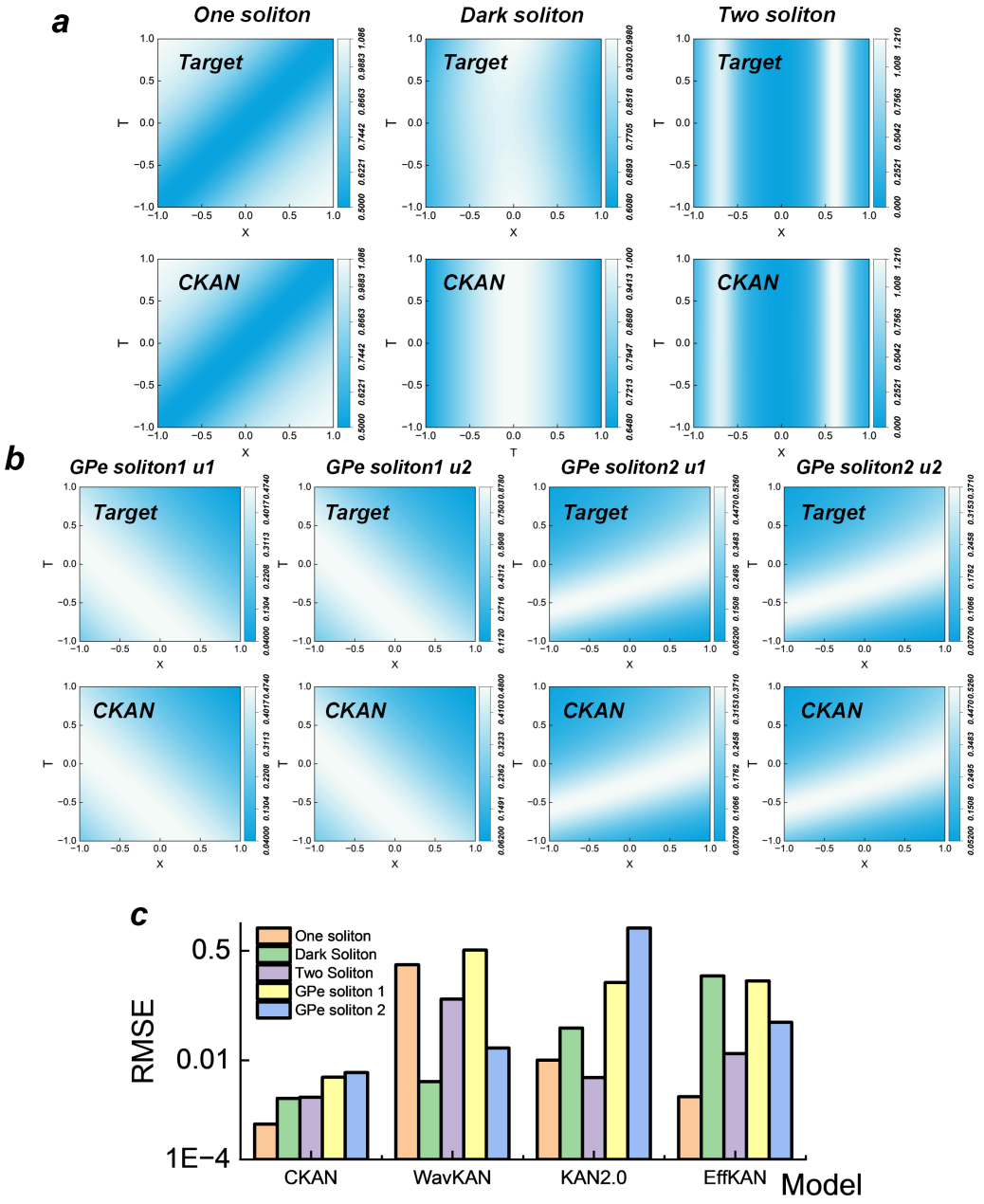

Figure 1: CKAN performance validation and error summary. (a) Standard NLSe single-soliton, double-soliton and dark-soliton solutions, together with the corresponding CKAN distillation results. (b) Two single-soliton solutions of the GPe and the corresponding CKAN distillation results. (c) Relative mean squared errors of the distillation results obtained by different models. (d) Schematic diagram of the CKAN architecture. (SFig. 1).

Here, $f(t)$ and $V_{\text{ext}}(x,t) = -\frac{1}{2}\rho x^2 - 2\alpha(t)$ denote the nonlinear coefficient and the external potential, respectively, and

$$\rho = \frac{d^2}{dt^2}\ln(2f(t)) - (\frac{d}{dt}\ln(2f(t)))^2.$$

We fix $\sigma = 0$ and consider two cases: (i) $V_{\text{ext}}(x,t) = 0$ and $f(t) = 0.5$; (ii) $V_{\text{ext}}(x,t) = -2x\cos(t)$ and $f(t) = 0.5e^{0.5}$. In both cases, the corresponding soliton solutions shown in SFig 1b can be obtained. The error function used for distillation is the classic MSE error. In all distillation tasks, CKAN performed the best (SFig 1c), demonstrating the superiority of CKAN in handling complex-valued data among KAN networks and their variants.

2 Data Acquisition Process and Network Construction Used for Training and Testing Supplementation

2.1 Laser Configuration

The specific design details of the fiber laser generating QML pulses are as follows (SFig 2a): In the cavity structure design section, a 980 nm laser diode is mainly used as the pump to provide energy for the operation of the resonant cavity; a wavelength division multiplexer is used as the energy coupling device to couple the pump energy into the resonant cavity to achieve excitation amplification; as previously mentioned, a 20 cm high-doping-concentration erbium-doped fiber is selected as the gain medium to achieve pump laser frequency conversion and the generation of Q-switched pulses; a polarization-independent isolator is used to suppress back-reflection inside the resonant cavity and protect some unidirectional passive components from damage caused by high peak-power laser radiation; the coupler is mainly used for outputting the laser energy generated by the resonant cavity for subsequent applications and optical detection; and the polarization controller is used for fine adjustment of the polarization state of pulses inside the resonant cavity. It should be noted that except for the polarization controller and gain fiber, the remaining devices inside the resonant cavity are all fusion-spliced with single-mode fiber (Corning, SMF-28e), with a total single-mode fiber length of about 1.2 m, forming together with the gain fiber an all-fiber resonant cavity with a total length of about 1.45 m, which together determine the oscillation frequency of the mode-locked pulses.

The specific design details of the fiber laser generating MCS pulses are as follows (SFig 2b): In the microcavity structure design part, a Knot-type all-fiber microcavity is mainly adopted to achieve the optical comb effect; it is then incorporated into the fiber laser, where the fiber laser serves as the main cavity, using a 60 cm high-concentration erbium-doped gain fiber to continuously provide gain for the microcavity; a polarization-dependent isolator is used to suppress back-reflection inside the resonant cavity and to polarize the transmitted optical field; the fiber coupler is mainly used for internal optical field measurement of the laser, extracting 10% of the laser from the main cavity for subsequent laser measurement; the polarization controller is used to adjust the intracavity polarization state and weak birefringence effect. It should be noted that the free spectral range of the optical comb generated by the fiber microcavity is approximately 0.064 nm, and the microcavity diameter measured using a microscope corresponds to this value according to the conversion formula. The overall structure adopts an all-fiber configuration, ensuring both stable output states inside the cavity and the ability to switch between different intracavity states at any time by adjusting the pump power.

The specific design details of the fiber laser generating traditional pulses are as follows

]

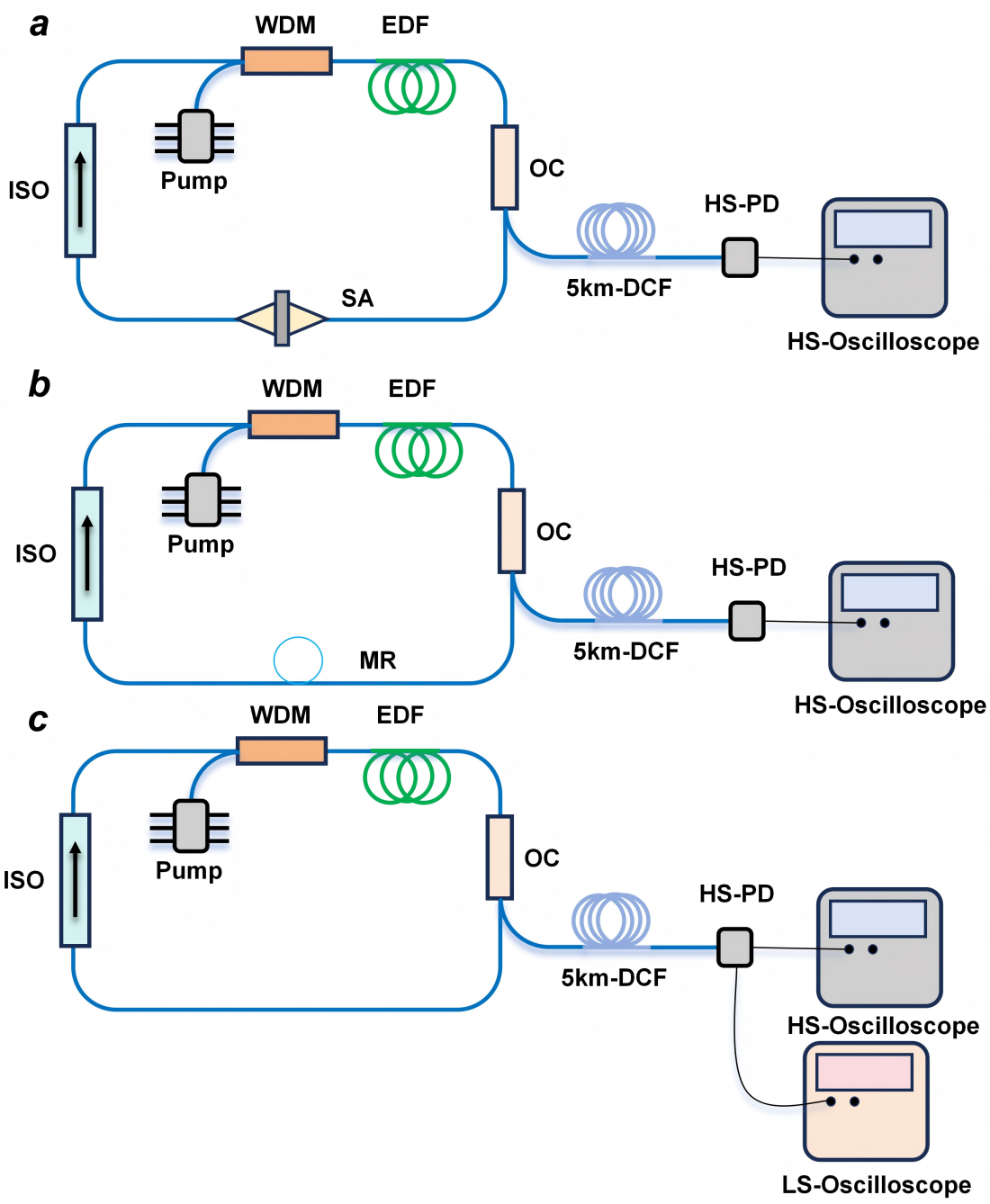

Figure 2: Schematic diagram of the laser configuration used in the experiment. (a) Schematic diagram of the laser configuration used for QML pulse generation, where SA denotes the saturable absorber. (b) The laser used for MCS pulse generation, where MR denotes the microresonator. (c) The traditional cavity laser configuration in the oversampling experiment, where LS-Oscilloscope denotes the low-speed oscilloscope used to sample pulses required for oversampling prediction, and HS-Oscilloscope denotes the high-speed oscilloscope used to sample the target pulses of the oversampling prediction (SFig. 2).

(SFig 2c): The laser generating the output pulses mainly consists of an all-fiber structure and uses conventional nonlinear polarization evolution technology to lock and operate the intracavity mode states. The internal cavity structure includes a 980 nm laser diode as the pump source to continuously provide pump energy for the resonant cavity to achieve excitation amplification; a 20 cm high-doping-concentration erbium-doped fiber is used as the gain medium to achieve wavelength conversion and gain amplification; a polarization-dependent isolator is used to suppress back-reflection inside the resonant cavity and to polarize the transmitted optical field to generate traditional mode-locked pulse output; the fiber coupler is mainly used for internal optical field testing of the laser, extracting 10% of the energy from the main cavity for subsequent laser measurement; the cavity contains two polarization controllers, one for coarse adjustment and one for fine adjustment of the intracavity polarization state. It should be noted that except for the polarization controllers and gain fiber, the other device ends inside the resonant cavity are all fusion-spliced with single-mode fiber (Corning, SMF-28e).

2.2 Introduction of the Nine Comparison Models

Unet is a classic architecture in the field of semantic segmentation, with a core structure of "encoder–decoder + skip connections": the encoder extracts multi-scale features through downsampling (from details to semantics), the decoder restores resolution through upsampling, and the skip connections fuse the high-resolution detailed features from the encoder with the semantic features of the decoder. The following are Unet variants with different backbones. To ensure experimental accuracy, the study directly uses the segmentation_models_pytorch package for model invocation, and the configuration is provided in the GitHub link given in the main text:

Dual Path Unet++ (DPUN): A hybrid architecture combining the residual connections of ResNet and the dense connections of DenseNet. It contains two parallel paths: the Residual Path, which transmits original features through skip connections (similar to ResNet), and the Dense Path, which accumulates historical features through concatenation. The dual-path design enhances feature reuse and fusion capability. The study uses the variant with a depth of 68 layers. When used as a Unet encoder, its dual-path feature extraction capability is stronger, enabling it to capture richer multi-scale contextual information, suitable for segmentation tasks in complex scenes.

EfficientUnet++ (EUN): Based on Compound Scaling design, it balances accuracy and efficiency by uniformly scaling network depth, width, and input resolution. Its core module is the Mobile Inverted Residual Block, which uses depthwise separable convolution to reduce computation and incorporates the SE attention mechanism. The study uses a medium-scale variant. When used as a Unet encoder, it extracts features more efficiently than traditional ResNet under the same computational cost, suitable for resource-limited scenarios while maintaining high accuracy.

MobileOne-S2-Unet++ (MO-S2): Uses MobileOne-S2 as the encoder. This backbone is a high-efficiency network designed specifically for mobile devices, with the core being structural re-parameterization. During training, multi-branch structures are used to enhance feature learning ability, while during inference, the multi-branch structure is merged into a single branch through algebraic operations to eliminate computational redundancy, achieving

efficient inference. The study uses the variant that balances speed and accuracy. When used as a Unet encoder, it can provide sufficient feature representation capability with extremely low latency.

timm-SKResNet-Unet++ (SKUnet): Combines the SKResNet backbone from the timm library. This backbone is an improved version of ResNet, with the core being the Selective Kernel mechanism, which introduces convolution kernels of multiple sizes into convolution layers. Through an attention mechanism, the network dynamically learns the kernel weights and adaptively selects the most suitable kernel size based on the input features, enhancing adaptability to multi-scale targets. When used as a Unet encoder, it can more flexibly capture target features at different scales, suitable for tasks where target size varies greatly.

ResNet-Unet (ResUnet): Uses ResNet as the backbone. This backbone addresses the gradient vanishing problem in deep networks through residual connections. Its core module is the residual block, which allows the network depth to increase significantly while maintaining stable feature extraction ability. It is one of the most classic backbones in computer vision. As the encoder of Unet, it is one of the most widely used combinations, with simple structure and stable performance.

Fourier Neural Operator (FNO): An operator learning model based on Fourier transform, which breaks through the efficiency limitations of traditional CNNs and Transformers in modeling global dependencies in high-dimensional space. Traditional methods relying on local convolution struggle to capture long-range dependencies, and although Transformers can model global dependencies, their computational complexity increases quadratically with input size. The core of FNO is the introduction of Fourier layers, mapping input features to the frequency domain through Fourier transform. In the frequency domain, a small number of low-frequency components dominate global information for feature transformation, and the features are mapped back to the spatial domain through inverse Fourier transform. Its advantages include computational complexity that grows linearly with input size and efficient modeling of global dependencies in high-dimensional space. It is suitable for tasks such as solving partial differential equations, fluid dynamics simulation, and weather forecasting that require modeling global physical laws.

Fourier Neural Operator-big (FNO-big): A parameter-expanded version of the FNO model.

Unet-Fourier Neural Operator (UFNO): A combined model of Unet and FNO, integrating the skip-connection encoder–decoder structure of Unet with the Fourier-domain modeling capability of FNO. Its encoder replaces traditional convolution with Fourier layers from FNO, capturing global dependencies through frequency-domain transformation while extracting multi-scale features; the decoder restores resolution through upsampling and fuses the high-resolution local features from the encoder with the global semantic features generated by the decoder through skip connections. This design preserves Unet's precise local feature capture capability while enhancing the modeling of global physical laws (such as long-range spatial correlations) through FNO, making it especially suitable for tasks that require balancing local detail and global dependencies.

Unet-Fourier Neural Operator-big (UFNO-big): A parameter-expanded version of the UFNO model.

Koopman Operator (KO): Koopman theory is a mathematical tool for studying nonlinear dynamical systems, and the Koopman operator is its data-driven extension. For nonlinear

dynamical systems, by introducing observable functions, the system can be mapped into a linear space, thereby transforming nonlinear evolution into a linear evolution process in the linear space, achieving a linearized description of the system. Its application scenarios include time series prediction, dynamical system control, and physical system modeling.

2.3 SNOO Training Process

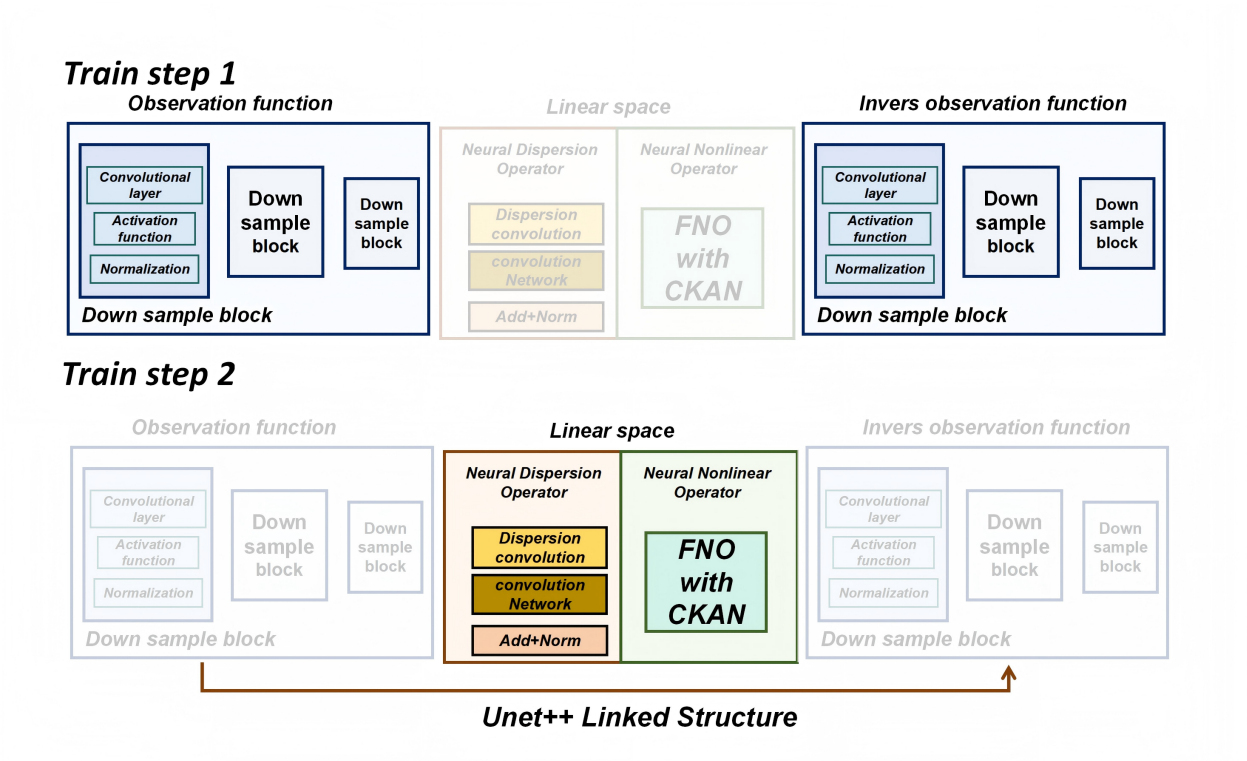

Figure 3: The components of each module of SNOO and the two-stage training method. Train step 1 requires the network to freeze the internal linear operator structure and train itself by reconstructing pulses. Train step 2 requires the network to freeze the external framework and train the internal linear operator in an autoregressive manner. Note that the operators consist of multiple layers and are embedded into the links of Unet++ (SFig. 3).

2.4 Different Training Methods of SNOO

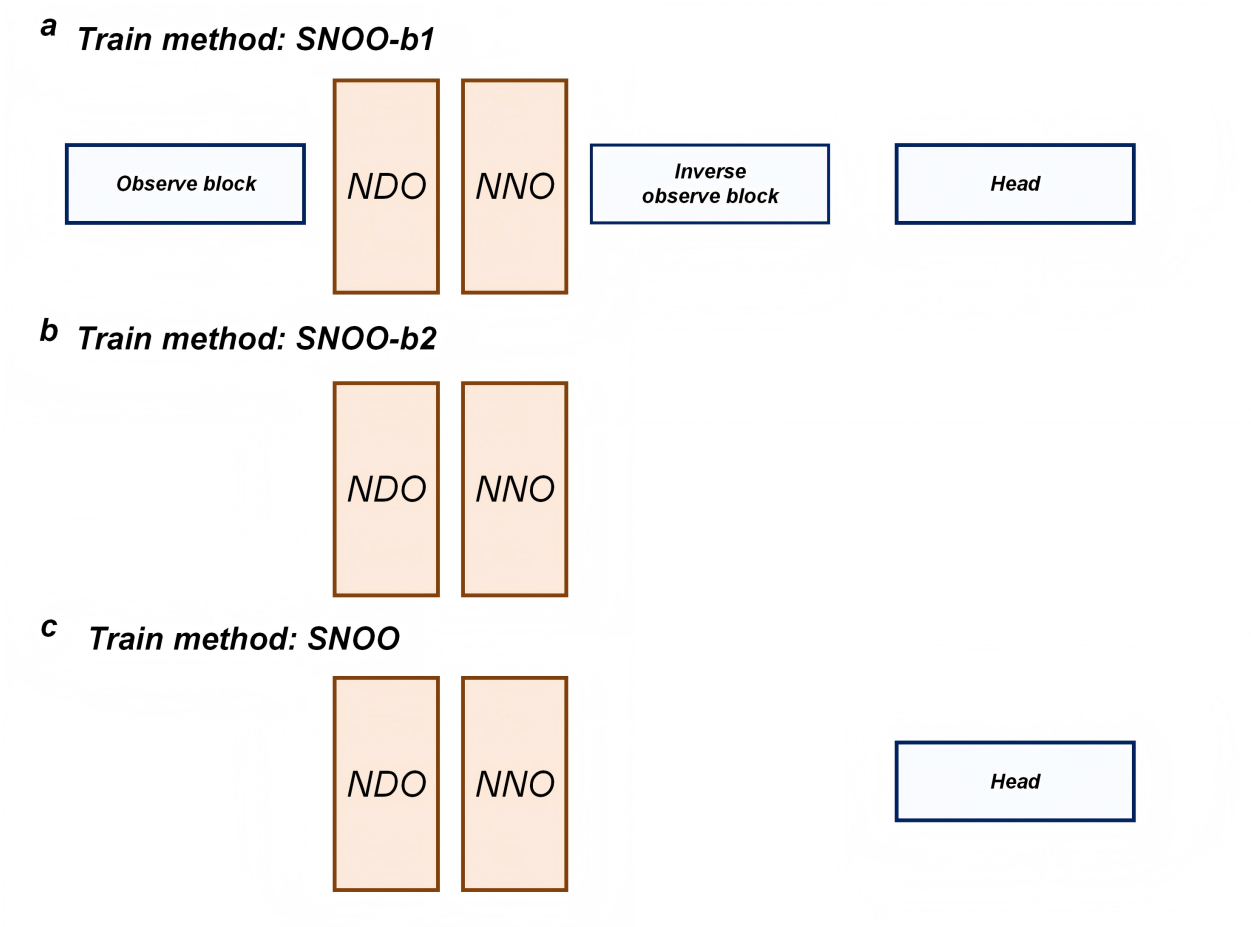

Figure 4: Different training methods of SNOO. (a) When training the internal linear operator, all trainable parameters of the external framework are not frozen. (b) When training the internal linear operator, all trainable parameters of the external framework are frozen. (c) When training the internal linear operator, only the head part of the external framework is open for training (SFig. 4).

References

[1] Huang, X. & Alkhalifah, T. Efficient physics-informed neural networks using hash encoding. *Journal of Computational Physics* 501, 112760 (2024).

[2] Mai, C., Zhang, L., Chao, X., Hu, X. & Behar, O. A multi-scale feature extraction and fusion framework based on wavelet Kolmogorov-Arnold networks and parallel Bi-directional gated recurrent units for electric load forecasting. *Engineering Applications of Artificial Intelligence* 162, 112517 (2025).

[3] Wang, H., Yang, H., Meng, X., Tian, Y. & Liu, W. Dynamics of controllable matter-wave solitons and soliton molecules for a Rabi-coupled Gross-Pitaevskii equation with

temporally and spatially modulated coefficients. *SIAM Journal on Applied Dynamical Systems* 23, 748–778 (2024).