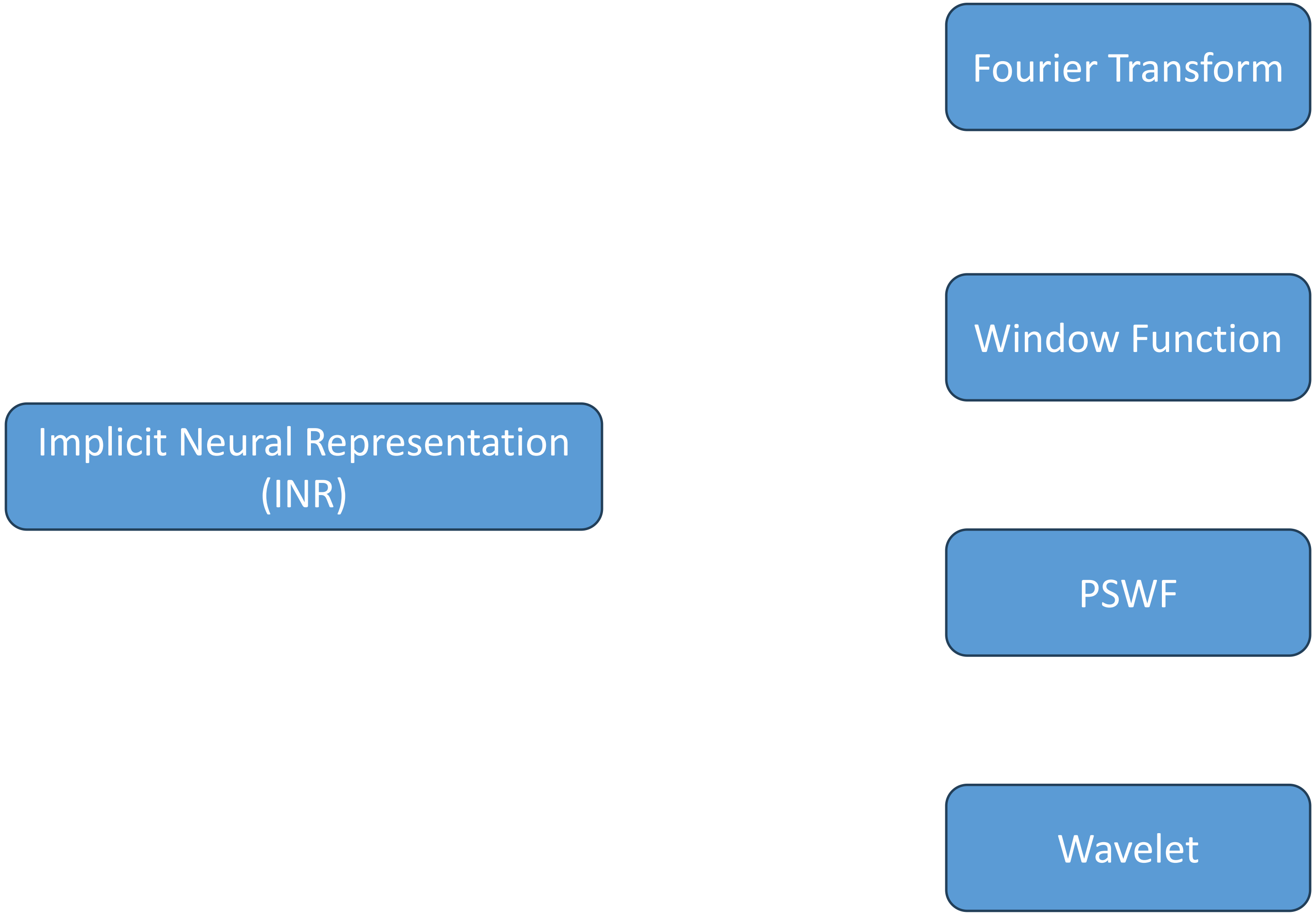
PSWF and Complex Wavelets as Activation Functions in INR


Fourier Transform
Window Function
Implicit Neural Representation (INR)
PSWF
Wavelet

Implicit Neural Representation (INR)

- INR represents data as a continuous function, which is modeled by a neural network such as an MLP.
- For example, a 100×100 image can be represented as a continuous function:

$$F(x, y) = (R, G, B)$$

- where (x,y) are spatial coordinates and F outputs RGB color values.

Mathematical Foundations

- Taylor Series:

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \frac{x^5}{5!} + \cdots$$
$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots$$
$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots$$

- Euler's Formula:

$$\begin{aligned} e^{i\omega t} &= 1 + i\omega t + \frac{(i\omega t)^2}{2!} + \frac{(i\omega t)^3}{3!} + \frac{(i\omega t)^4}{4!} + \cdots \\ &= \left(1 - \frac{(\omega t)^2}{2!} + \frac{(\omega t)^4}{4!} - \cdots\right) + i\left(\omega t - \frac{(\omega t)^3}{3!} + \frac{(\omega t)^5}{5!} - \cdots\right) \\ &= \cos(\omega t) + i\sin(\omega t) \end{aligned}$$

$$\cos(\omega t) = \frac{e^{i\omega t} + e^{-i\omega t}}{2}, \quad \sin(\omega t) = \frac{e^{i\omega t} - e^{-i\omega t}}{2i}$$

Mathematical Foundations

- Fourier Transform: All Functions Are Combinations of Frequencies

$$f(t) = \int_{-\infty}^{\infty} \hat{f}(\omega)\, e^{i\omega t}\, d\omega$$

- Why Frequency?
  - Many real-world signals (images, audio) contain inherent periodic patterns, and Trigonometric functions (sin/cos) can capture these patterns more compactly and expressively.
  - Neural networks with ReLU or GELU tend to learn low-frequency components first (*spectral bias*). Using frequency-based activations (e.g. complex wavelet, PSWF) help the model to learn high-frequency information.

Window Functions

- Why Window Function?
  - We are only interested in certern range of time (and frequency) in a signal. So we apply a window to focus on a local region of the signal to cut into what we want.

$$f_{\text{windowed}}(t) = f(t) \cdot w(t)$$

- The Simplest Example: Rectangular Window

$$w(t) = \begin{cases} 1, & |t| \leq T \\ 0, & \text{otherwise} \end{cases}$$

Prolate Spheroidal Wave Function (PSWF)

- Band-limited to [−W, W], time-limited to [−T, T], solve the following dual equation, and we get the PSWF.

$$\max_f \quad \frac{\int_{-T}^{T} |f(t)|^2 dt}{\int_{-\infty}^{\infty} |f(t)|^2 dt} \quad \text{subject to } \hat{f}(\omega) = 0 \text{ for } |\omega| > W$$

$$\max_f \quad \frac{\int_{-W}^{W} |\hat{f}(\omega)|^2 d\omega}{\int_{-\infty}^{\infty} |\hat{f}(\omega)|^2 d\omega} \quad \text{subject to } f(t) = 0 \text{ for } |t| > T$$

- Solutions to both problems are the same: the Prolate Spheroidal Wave Functions (PSWFs) — optimally concentrated in both time and frequency.

Motivation

- Activation Functions ≈ Basis Functions
  - Activation functions play the role of basis functions in neural networks.
- Window Functions ≈ Good Basis Functions
  - In signal processing, window functions are designed to achieve localization in both time and frequency, making them effective for signal representation.
- → Use Window Functions as Activation Functions
  - Both papers use popular window functions (complex wavelets/PSWF) as activations.

What does PIN do?

- PSWF Activation Replacement

$$\Phi_\theta(x) = \sum_{k=0}^{K} a_k \cdot \mathrm{pswf}_k(Wx + b)$$

- Learnable Time-Frequency Localization (parameter in PSWF)

$$\hat{\psi}(x) = T \cdot \psi(wx) + b$$

Experiment

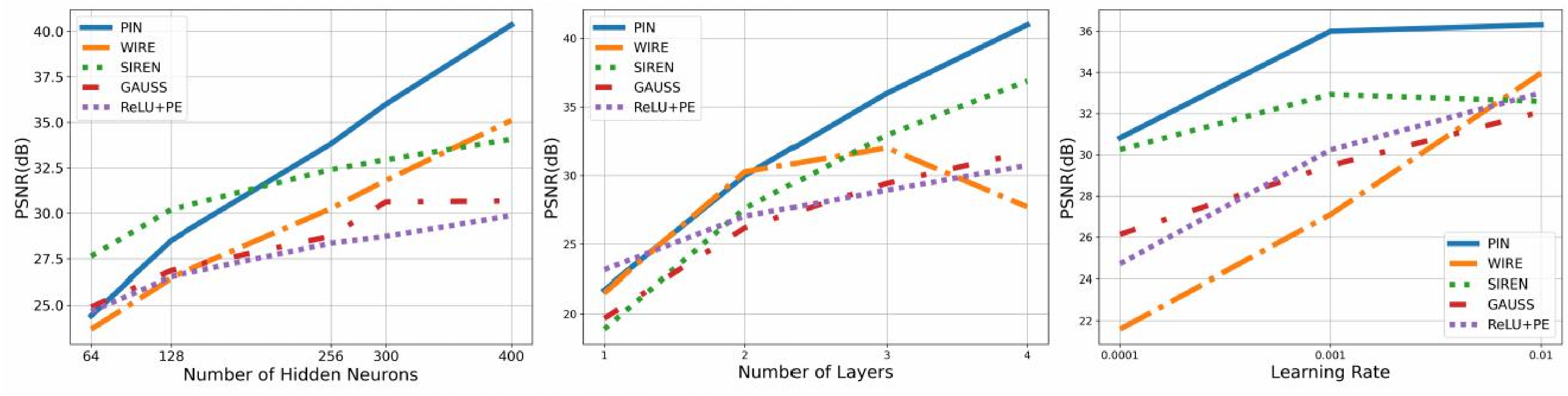


Figure 7: **Hyperparameter Turning of PIN**: PIN demonstrates a sharp linear increase in PSNR with the addition of more hidden neurons and layers compared to existing INRs. Instead of becoming unstable with higher learning rates, PIN stabilizes its PSNR, maintaining nearly constant performance.

What is a Wavelet?

- Traditional basis functions like sine and cosine are global:

$$\psi(t) = e^{i\omega t}$$

    - They extend over all space or time. Wavelets are localized in both time and frequency

- Morlet wavelet:

$$\psi(t) = e^{i\omega t} \cdot e^{-\frac{t^2}{2\sigma^2}}$$

    - It combines oscillation (via e^{iωt}) and localization (via Gaussian window)

Key Theoretical Contributions

- The expressivity of INRs is governed by the Fourier transform of the first-layer activation.
- Wavelet-INRs retain time-frequency localization and multi-scale structure, even after stacking layers with non-linearities.
- Proposes a split INR architecture to decouple smooth (low-frequency) and singular (high-frequency) components.