CauchyNet

Compact and Data-Efficient Learning Using Holomorphic Activation Functions

Hong-Kun Zhang Xin Li Sikun Yang Zhihong Xia

Great Bay University · University of Massachusetts Amherst

Presented by **Hong-Kun Zhang**

International Conference on Machine Learning (ICML) 2026

When “scale for accuracy” breaks

Edge devices, scarce data, and partially observed signals all defeat the big-model recipe. Targets with **sharp peaks** or **near-singular** structure are especially hard:

- **ReLU networks** need $\Omega(1/\delta)$ linear pieces to resolve a peak of width $\sim\delta$ — each unit gives just one bend.
- **SIREN** (sinusoidal) captures smooth oscillation but rings (Gibbs) near sharp transitions.
- **Transformers / N-BEATS** are flexible but parameter-heavy — out of reach for edge or data-scarce settings.

Can a tiny network resolve spikes exactly, with provable guarantees?

Idea: borrow the pole from Cauchy's integral formula

For f holomorphic inside a contour C,

$$f(z) = \frac{1}{2\pi i} \oint_C \frac{f(\xi)}{\xi - z}\, d\xi.$$

The **inversion term** $1/(\xi - z)$ is the whole engine. In $\mathbb{C}^N$ it becomes a **Cauchy-kernel atom**

$$K(\boldsymbol{\xi}, \mathbf{x}) = \prod_{i=1}^{N} \frac{1}{\xi_i - x_i}.$$

One hidden unit = one kernel atom = one pole $\boldsymbol{\xi}_k$.
The network only learns where to place the poles.

The CauchyNet architecture: bias-only, single layer

Cauchy activation: $\mathscr{X}(\boldsymbol{z}) = \prod_{i=1}^{N} z_i^{-1}$.

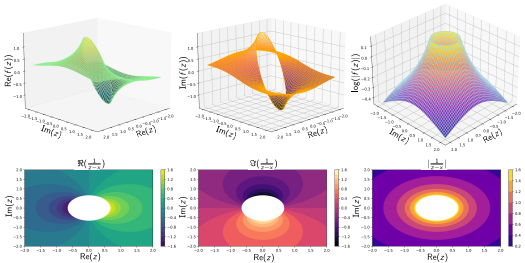
A single hidden layer with **no weight matrix** W — only complex bias shifts:

$$\mathbf{H} = \mathbf{Z} + \mathbf{B}, \qquad \mathbf{B} \in \mathbb{C}^{h \times N}$$

$$h_k = \prod_{i=1}^{N} (H_{k,i} + \varepsilon)^{-1}$$

$$o = \tfrac{1}{h}\, \mathbf{C}^{\top} \boldsymbol{h}, \qquad \mathbf{C} \in \mathbb{C}^{h \times 1}$$

$\approx 2h(N{+}1)$ real parameters, $\mathcal{O}(hN)$ compute.



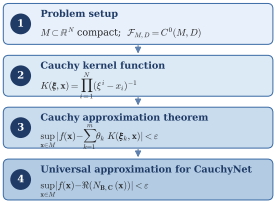
Cauchy activation $\mathscr{X}(\boldsymbol{z})$: real / imaginary / magnitude.

Theory: bias-only is already universal

Why no W? Translation covariance makes a pre-activation weight redundant:

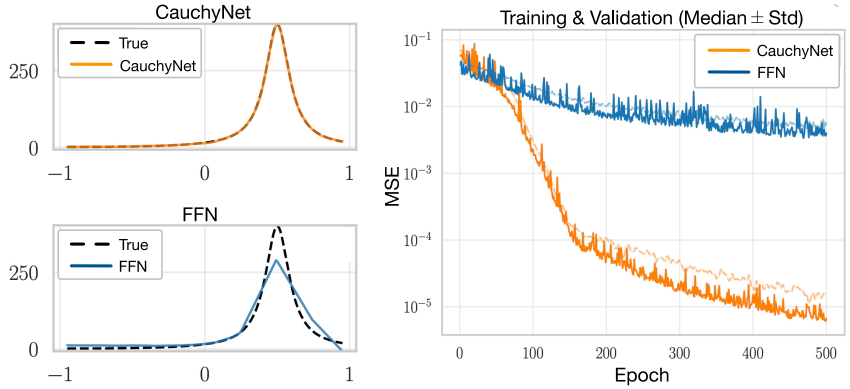
$$\frac{1}{wx+b} = \frac{1}{w} \cdot \frac{1}{x-(-b/w)}.$$

- **Thm (universality).** Finite Cauchy-kernel sums are **dense in $C(M)$** on compact $M \subset \mathbb{R}^N$ — so a width-h CauchyNet approximates any continuous f to arbitrary ε, *using biases* **B** *and output weights* **C** *only*.
- **Rates.** Analytic targets: $\mathcal{O}\left(e^{-2\pi d\, h^{1/N}}\right)$ (exponential). C^k targets: $\mathcal{O}\left(h^{-k/N}\right)$.



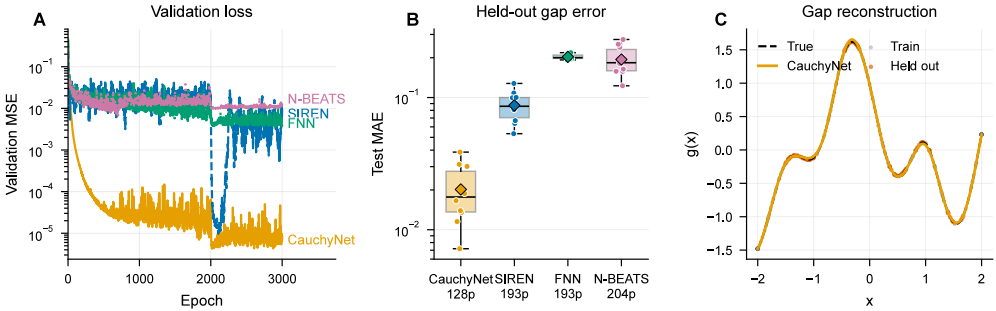
Constructive proof: contour quadrature → kernel sum → CauchyNet realization.

Experiment 1 — sharp peaks under limited data



CauchyNet tracks the steep rational spike; a width-matched **ReLU FFN** underestimates it. With a **fixed-pole** dictionary and ridge-fitted coefficients (only $n_{\text{train}}=20$ samples): dense-grid MSE **0.0027** vs. **0.0403** for a 256-parameter ReLU FFN — **$\sim$15$\times$** lower error at half the learned parameters.

Experiment 2 — missing-value imputation (gap-filling)



Model	Test MAE	Params
CauchyNet	**0.020**	**128**
SIREN	0.087	193
N-BEATS	0.194	204
ReLU FNN	0.203	193

Held-out gaps at turning points:

4.3× lower error than SIREN,
9.6× than N-BEATS, **10.1**× than FNN —
at **34% fewer** trainable parameters.

Where it wins — and where it doesn't

Strong gains

- Near-singular targets: **20**× over ReLU, up to **102**× at moderate pole distance.
- Wins on chirp, piecewise-smooth, and Gibbs targets.
- Hybrid FNN→Cauchy best on tabular (Diabetes, Cal. Housing, Wine).

Honest limits

- ReLU wins the piecewise-affine "step-ramp" by **1.6**×.
- Advantage collapses when the pole sits extremely close to the axis (δ=0.01) and samples under-resolve the spike.
- Trainable poles can be fragile at very small n; fixed dictionaries are more stable.

A principled inductive bias for near-singular structure — not a universal drop-in.

Takeaways

- **CauchyNet**: a single-layer, *bias-only*, complex-valued network of Cauchy-kernel atoms — each unit is a pole.
- **Provable** universality with exponential (analytic) and algebraic (C^k) approximation rates.
- **Compact & data-efficient**: up to **10–100**$\times$ lower error at fewer parameters on near-singular and gap-filling tasks.
- Classical complex analysis still has much to offer modern deep learning.

Future: higher dimensions, longer sequences, deep / residual variants, integration with neural operators and PDE solvers.

Thank you! Paper & code released with the ICML 2026 camera-ready.