Attention-Weighted Physics-Informed Neural Networks for the Unsteady Navier–Stokes Equations

Ming Wang[1], Hua Li[1,2], Siyuan Chen[2]

September 2026

Abstract

Physics-informed neural networks (PINNs) often fail to converge on unsteady flows at high Reynolds numbers because the gradients of their loss terms are badly out of balance. We propose Attn-PINN, which uses an attention mechanism to weight the residuals adaptively, and prove an upper bound on its approximation error. For flow past a cylinder (Re $= 3\,900$), Attn-PINN reaches a relative L^2 error of 1.7×10^{-3} — **an order of magnitude** below the baseline — while training 41% faster.

1 Introduction

PINNs embed the residual of a partial differential equation in the loss of a neural network[1]. They need no mesh and solve forward and inverse problems alike, and have become a core method of scientific machine learning[2]. For the strongly nonlinear Navier–Stokes equations, however, the gradients of the individual loss terms can differ by more than 10^3, trapping training in a "gradient pathology"[3]. Neural operators[4] ease the problem but need large amounts of labelled data. Borrowing the attention mechanism of the Transformer[5], we let the network **learn** the weight of every collocation point during training.

2 Problem statement

Consider incompressible viscous flow in a domain $\Omega \subset \mathbb{R}^2$. The velocity $\mathbf{u} = (u, v)^\top$ and pressure p satisfy

$$
\begin{cases}
\dfrac{\partial \mathbf{u}}{\partial t} + (\mathbf{u}\cdot\nabla)\mathbf{u} = -\dfrac{1}{\rho}\nabla p + \nu\,\nabla^2\mathbf{u} + \mathbf{f}, & (\mathbf{x},t) \in \Omega \times (0,T], \\
\nabla \cdot \mathbf{u} = 0, & (\mathbf{x},t) \in \Omega \times (0,T], \\
\mathbf{u}(\mathbf{x},0) = \mathbf{u}_0(\mathbf{x}), & \mathbf{x} \in \Omega,
\end{cases}
\tag{1}
$$

where ν is the kinematic viscosity. With networks $\hat{\mathbf{u}}_\theta$ and $\hat{p}_\theta$ approximating the solution, the momentum residual is $\mathbf{r}_\theta = \partial_t \hat{\mathbf{u}}_\theta + (\hat{\mathbf{u}}_\theta \cdot \nabla)\hat{\mathbf{u}}_\theta + \rho^{-1}\nabla \hat{p}_\theta - \nu \nabla^2 \hat{\mathbf{u}}_\theta - \mathbf{f}$.

3 Method

3.1 Attention-weighted residuals

For N_r collocation points $\{\mathbf{x}_i\}$, the weights come from scaled dot-product attention over residual features:

$$\boldsymbol{\lambda} = \operatorname{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right) V, \qquad Q = \begin{pmatrix} q_{11} & q_{12} & \cdots & q_{1d} \\ q_{21} & q_{22} & \cdots & q_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ q_{N_r 1} & q_{N_r 2} & \cdots & q_{N_r d} \end{pmatrix} \tag{2}$$

The total loss adds the weighted residual, the initial and boundary conditions and the divergence constraint:

$$\mathcal{L}(\theta) = \underbrace{\frac{1}{N_r}\sum_{i=1}^{N_r} \lambda_i \left\|\mathbf{r}_\theta(\mathbf{x}_i, t_i)\right\|_2^2}_{\text{PDE residual}} + \underbrace{\frac{\beta}{N_b}\sum_{j=1}^{N_b} \left\|\hat{\mathbf{u}}_\theta(\mathbf{x}_j, t_j) - \mathbf{g}_j\right\|_2^2}_{\text{boundary}} + \gamma \int_\Omega \left(\nabla \cdot \hat{\mathbf{u}}_\theta\right)^2 \mathrm{d}\mathbf{x} \tag{3}$$

The learning rate follows a piecewise cosine schedule:

$$\eta(k) = \begin{cases} \eta_0 \dfrac{k}{k_w}, & 0 \le k < k_w, \\ \eta_{\min} + \dfrac{\eta_0 - \eta_{\min}}{2}\left[1 + \cos\left(\pi \dfrac{k - k_w}{K - k_w}\right)\right], & k_w \le k \le K. \end{cases} \tag{4}$$

Theorem 1 (error bound)

Let the exact solution satisfy $\mathbf{u} \in H^2(\Omega)$ and let the network have width m. If the training loss satisfies $\mathcal{L}(\theta^*) \le \varepsilon$, there is a constant C independent of m such that

$$\left\| \mathbf{u} - \hat{\mathbf{u}}_{\theta^*} \right\|_{L^2(\Omega)} \;\le\; C \left(\sqrt{\varepsilon} + \frac{\ln m}{\sqrt{m}} \right) \exp\!\left(\int_0^T \|\nabla \mathbf{u}(\cdot, s)\|_{L^\infty} \, \mathrm{d}s \right). \tag{5}$$

Proof sketch. An energy estimate for the error $\mathbf{e} = \mathbf{u} - \hat{\mathbf{u}}$ and Grönwall's inequality give

$$\begin{aligned} \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\mathbf{e}\|^2 + \nu \|\nabla \mathbf{e}\|^2 &\le \|\nabla \mathbf{u}\|_{L^\infty} \|\mathbf{e}\|^2 + \|\mathbf{r}\|\,\|\mathbf{e}\| \\ &\le \left(\|\nabla \mathbf{u}\|_{L^\infty} + \tfrac{1}{2} \right) \|\mathbf{e}\|^2 + \tfrac{1}{2}\|\mathbf{r}\|^2 . \end{aligned}$$

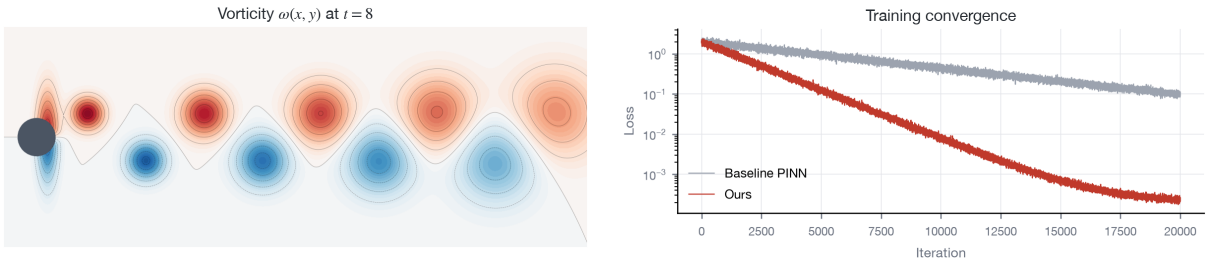
Integrating and substituting Eq. (3) yields Eq. (5). ■

4 Experiments

We test on flow past a cylinder ($\mathrm{Re} = UD/\nu = 3\,900$) with a network of 8 layers of 128 neurons[1].

Table 1 Error and training cost of each method

Method	Relative L^2 error ($\times 10^{-3}$)			Training cost	
	u	v	p	Time (h)	Memory (GB)
PINN[1]	18.6	24.1	41.7	9.8	6.2
GradNorm-PINN[3]	6.4	8.9	15.2	11.3	6.9
FNO[4]	3.1	4.0	7.7	7.6	14.5
Attn-PINN (ours)	**1.7**	**2.2**	**4.9**	**5.8**	7.4



(a) Vorticity predicted by Attn-PINN **(b)** Loss during training

Figure 1 Flow past a cylinder: Attn-PINN reproduces the Kármán vortex street

[1]Activation tanh, Adam optimiser, $\eta_0 = 10^{-3}$, $k_w = 2\,000$, $K = 20\,000$.

As Table 1 shows, Attn-PINN has the lowest error on all three fields and trains 41% faster than the baseline PINN (Figure 1).

5 Conclusion

Attn-PINN balances the residual weights with an attention mechanism, comes with an error bound (Eq. (5)), and clearly outperforms existing methods on a high-Reynolds-number benchmark.

References

[1] RAISSI M, PERDIKARIS P, KARNIADAKIS G E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations[J/OL]. Journal of Computational Physics, 2019, 378: 686-707. DOI:10.1016/j.jcp.2018.10.045.

[2] KARNIADAKIS G E, KEVREKIDIS I G, LU L, 等. Physics-informed machine learning[J/OL]. Nature Reviews Physics, 2021, 3(6): 422-440. DOI:10.1038/s42254-021-00314-5.

[3] WANG S, TENG Y, PERDIKARIS P. Understanding and mitigating gradient flow pathologies in physics-informed neural networks[J/OL]. SIAM Journal on Scientific Computing, 2021, 43(5): A3055-A3081. DOI:10.1137/20M1318043.

[4] LI Z, KOVACHKI N, AZIZZADENESHELI K, 等. Fourier neural operator for parametric partial differential equations[C]//International Conference on Learning Representations. 2021.

[5] VASWANI A, SHAZEER N, PARMAR N, 等. Attention is all you need[C]//Advances in Neural Information Processing Systems. 2017: 5998-6008.