ICML
International Conference
On Machine Learning

Nonparametric Teaching of Implicit Neural Representations

Chen Zhang[1*], **Steven Tin Sui Luo**[2*], **Jason Chun Lok Li**[1], **Yik-Chung Wu**[1], **Ngai Wong**[1]

[1]The University of Hong Kong
[2]The University of Toronto

May 13, 2024

Overview

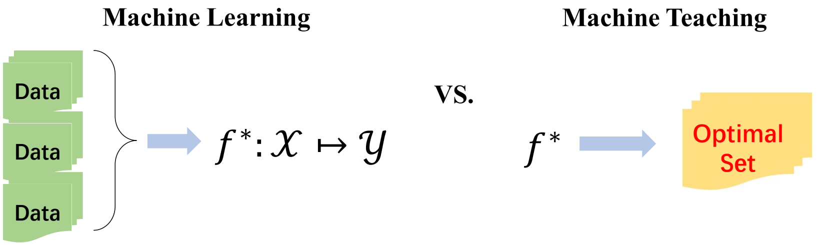
ICML
International Conference
On Machine Learning

Nonparametric Iterative Machine Teaching

What is Machine Teaching?

Machine teaching (MT) [17, 18] is the study of how to design the optimal teaching set, typically with minimal examples, so that learners can quickly learn target models based on these examples.

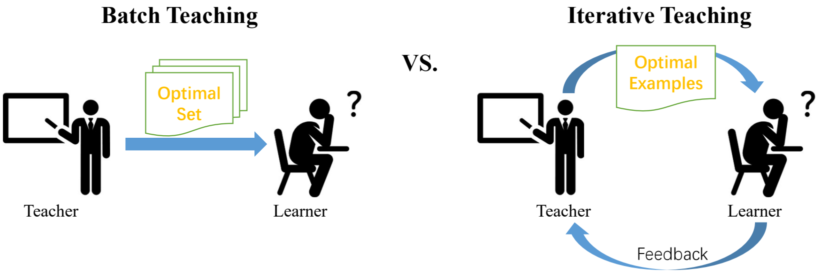
It can be considered as an inverse problem of machine learning, where machine learning aims to learn model parameters from a dataset, while MT aims to find a minimal dataset from the target model parameters.



What does "Iterative" mean?

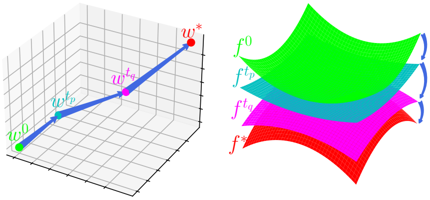
Considering the interaction manner between teachers and learners, MT can be conducted in either

- batch fashion [17, 9, 4, 10] where the teacher is allowed to interact with the learner once, or
- iterative fashion [6, 7, 8] where an iterative teacher would feed examples sequentially based on current status of the iterative learner.



“Parametric” VS. “Nonparametric”

Parametric Teaching [6, 7, 14, 13] assumes that f can be represented by a set of parameters $\boldsymbol{w}$, *e.g.*, $f(\boldsymbol{x}) = \langle \boldsymbol{w}, \boldsymbol{x} \rangle$ with input $\boldsymbol{x}$[1].



(a) Parametric IMT (b) Nonparametric IMT

Parametric assumption results in difficulty when the target models are defined to be functions without dependency on parameters (viz. non-closed-form functions). Such a limitation is addressed by **Nonparametric Teaching** [15, 16], which generalizes model space from a finite dimensional one to an infinite dimensional one.

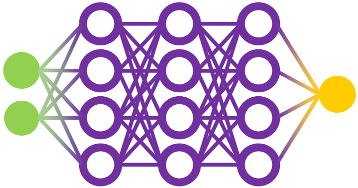
[1] The loss $\mathcal{L}$ can be general for different tasks, *e.g.*, square loss for regression and hinge loss for classification.

Implicit Neural Teaching (INT)

Implicit Neural Representations

Implicit neural representation (INR) [11, 12] focuses on modeling a given signal, which is often discrete, through the use of an overparameterized multilayer perceptron (MLP) such that the signal is accurately fitted by this MLP preserving great details.

Such an overparameterized MLP inputs low-dimensional coordinates of the given signal and outputs corresponding values for each input location, *e.g.*, the MLP maps 2D input coordinates to their respective 8-bit levels for a grayscale image.

2D Coordinates  Gray Scale

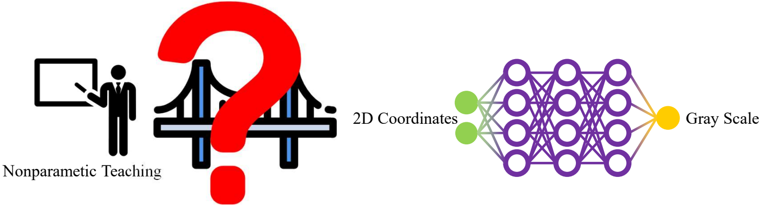
Motivation

The motivation comes from two folds:

- Lower the training cost and enhance the training efficiency of INR, which is urgently needed when dealing with high-definition signals. For instance, consider the case of a 2D grayscale image with a resolution of 1024×1024, which leads to a training set comprising 10^6 pixels
- Expand the applicability of nonparametric teaching towards deep learning. "Nonparametric" is a quite abstract concept, which may be of interest for theoretical analysis but less practical.

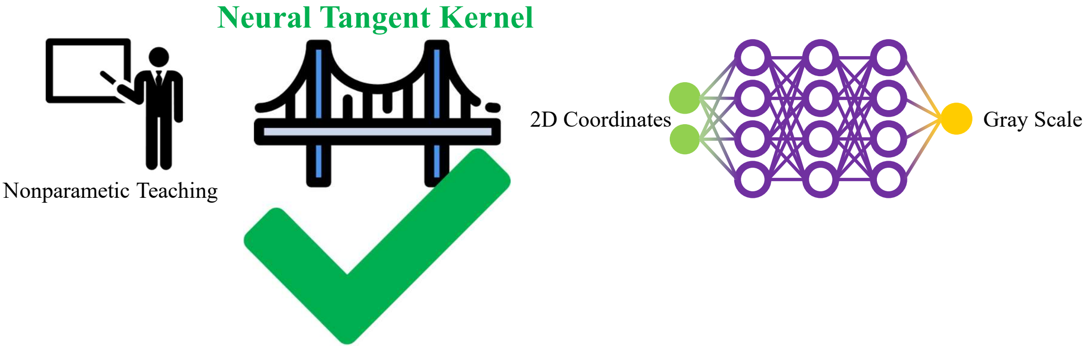
Cont.

- If we can connect nonparametric teaching to MLP training, both problems including training efficiency and applicability are addressed.
- Unfortunately, the evolution of an MLP is typically achieved by gradient descent on its parameters, whereas nonparametric teaching involves functional gradient descent as the means of function evolution.



Bridging this (theoretical + practical) gap is of great value and calls for more examination prior to the application of nonparametric teaching algorithms in the context of INR. ***Can we do that***?

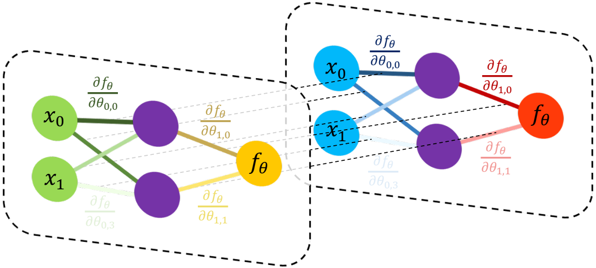
Cont.


Neural Tangent Kernel
Nonparametric Teaching
2D Coordinates
Gray Scale

Neural Tangent Kernel

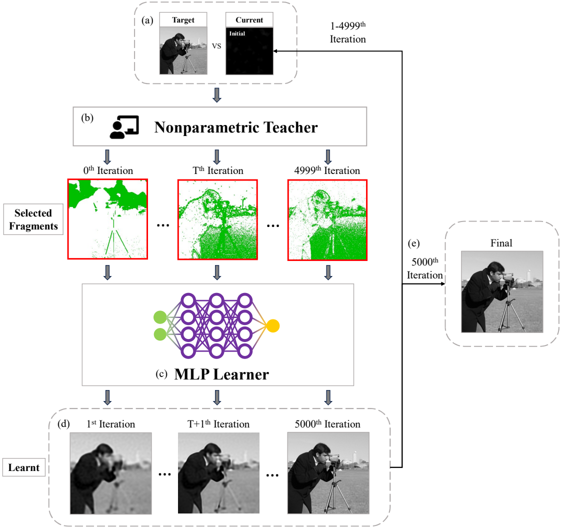
Neural Tangent Kernel [3, 5, 1, 2] is a symmetric and positive definite kernel function, which is derived from the analysis of the evolution of a neural network (the MLP is considered).

$$K_{\theta^t}(\boldsymbol{x}_i, \cdot) = \left\langle \left.\frac{\partial f_\theta}{\partial \theta}\right|_{\cdot,\theta^t}, \left.\frac{\partial f_\theta}{\partial \theta}\right|_{\boldsymbol{x}_i,\theta^t} \right\rangle \quad (1)$$



NTK = $[\sum_{l=0}^{L}\sum_{p=0}^{P^l} \frac{\partial f_\theta(x)}{\partial \theta_{l,p}} \frac{\partial f_\theta(x)}{\partial \theta_{l,p}}]_{1\mathrm{x}1} = [\frac{\partial f_\theta(x)}{\partial \theta_{0,0}} \frac{\partial f_\theta(x)}{\partial \theta_{0,0}} + \cdots + \frac{\partial f_\theta(x)}{\partial \theta_{0,3}} \frac{\partial f_\theta(x)}{\partial \theta_{0,3}} + \frac{\partial f_\theta(x)}{\partial \theta_{1,0}} \frac{\partial f_\theta(x)}{\partial \theta_{1,0}} + \frac{\partial f_\theta(x)}{\partial \theta_{1,1}} \frac{\partial f_\theta(x)}{\partial \theta_{1,1}}]$

Intuitive Illustration of INT Workflow



By comparing the disparity between the given signal and the current MLP output (a), the nonparametric teacher (b) selectively chooses examples (pixels) of the greatest disparity (red boxes), instead of a raster scan, to feed to the MLP learner (c) who undergoes learning (*i.e.*, training) (d) and outputs the final (e).

Experiments and Results

We conduct extensive experiments to validate the effectiveness of INT.

- **Toy 2D Cameraman fitting.**

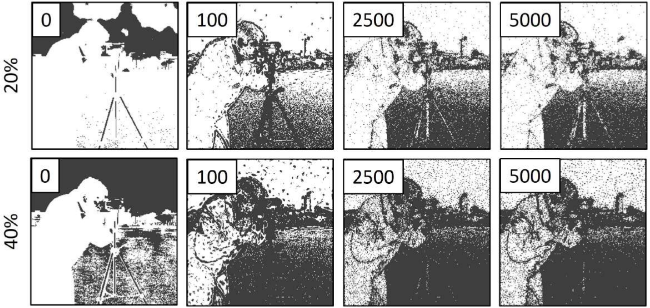


Figure: Progression of INT selected pixels (marked as black) at corresponding iterations when training with INT 20% (top) and 40% (bottom).

Cont.

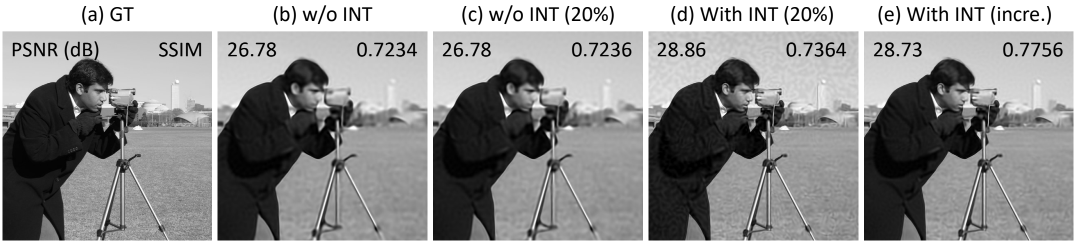


Figure: Reconstruction quality of SIREN. (b) trains SIREN without (w/o) INT using all pixels. (c) trains it w/o INT using 20% randomly selected pixels. (d) trains it using INT of 20% selection rate. (e) trains it using progressive INT (*i.e.*, increasing selection rate progressively from 20% to 100%).

Cont.

- **INT on multiple real-world modalities.**

INT	Modality	Time (s)	PSNR(dB) / IoU(%) ↑
✗	Audio	23.05	48.38±3.50
	Image	345.22	36.09±2.51
	Megapixel	16.78K	31.82
	3D Shape	144.58	97.07±0.84
✓	Audio	15.76 (-31.63%)	48.15±3.39
	Image	211.04 (-38.88%)	36.97±3.59
	Megapixel	11.87K (-29.26%)	33.01
	3D Shape	93.19 (-35.54%)	96.68±0.83

Table: Signal fitting results for different data modalities. The encoding time is measured excluding data I/O latency.

Contribution Summary

Contributions Summary

Main Contribution:

- We propose Implicit Neural Teaching (INT) that novelly interprets implicit neural representation (INR) via the theoretical lens of nonparametric teaching, which in turn enables the utilization of greedy algorithms from the latter to effectively bolster the training efficiency of INRs.
- We unveil a strong link between the evolution of a multilayer perceptron (MLP) using gradient descent on its parameters and that of a function using functional gradient descent in nonparametric teaching. This connects nonparametric teaching to MLP training, thus expanding the applicability of nonparametric teaching towards deep learning.
- We showcase the effectiveness of INT through extensive experiments in INR training across multiple modalities. Specifically, INT saves training time for 1D audio (-31.63%), 2D images (-38.88%) and 3D shapes (-35.54%), while upkeeping its reconstruction quality.

Thank you for listening!

References I

[1] Alberto Bietti and Julien Mairal. On the inductive bias of neural tangent kernels. In NeurIPS, 2019.

[2] Xialiang Dou and Tengyuan Liang. Training neural networks as learning data-adaptive kernels: Provable representation and approximation benefits. Journal of the American Statistical Association, 116(535):1507–1520, 2021.

[3] Arthur Jacot, Franck Gabriel, and Clément Hongler. Neural tangent kernel: Convergence and generalization in neural networks. In NeurIPS, 2018.

[4] Akash Kumar, Hanqi Zhang, Adish Singla, and Yuxin Chen. The teaching dimension of kernel perceptron. In AISTATS, 2021.

[5] Jaehoon Lee, Lechao Xiao, Samuel Schoenholz, Yasaman Bahri, Roman Novak, Jascha Sohl-Dickstein, and Jeffrey Pennington. Wide neural networks of any depth evolve as linear models under gradient descent. In NeurIPS, 2019.

[6] Weiyang Liu, Bo Dai, Ahmad Humayun, Charlene Tay, Chen Yu, Linda B Smith, James M Rehg, and Le Song. Iterative machine teaching. In ICML, 2017.

[7] Weiyang Liu, Bo Dai, Xingguo Li, Zhen Liu, James Rehg, and Le Song. Towards black-box iterative machine teaching. In ICML, 2018.

References II

[8] Weiyang Liu, Zhen Liu, Hanchen Wang, Liam Paull, Bernhard Schölkopf, and Adrian Weller. Iterative teaching by label synthesis. In NeurIPS, 2021.

[9] Farnam Mansouri, Yuxin Chen, Ara Vartanian, Jerry Zhu, and Adish Singla. Preference-based batch and sequential teaching: Towards a unified view of models. In NeurIPS, 2019.

[10] Hong Qian, Xu-Hui Liu, Chen-Xi Su, Aimin Zhou, and Yang Yu. The teaching dimension of regularized kernel learners. In ICML, 2022.

[11] Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. In NeurIPS, 2020.

[12] Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. In NeurIPS, 2020.

[13] Pei Wang, Kabir Nagrecha, and Nuno Vasconcelos. Gradient-based algorithms for machine teaching. In CVPR, 2021.

[14] Zhaozhuo Xu, Beidi Chen, Chaojian Li, Weiyang Liu, Le Song, Yingyan Lin, and Anshumali Shrivastava. Locality sensitive teaching. In NeurIPS, 2021.

References III

[15] Chen Zhang, Xiaofeng Cao, Weiyang Liu, Ivor Tsang, and James Kwok. Nonparametric iterative machine teaching. In ICML, 2023.

[16] Chen Zhang, Xiaofeng Cao, Weiyang Liu, Ivor Tsang, and James Kwok. Nonparametric teaching for multiple learners. In NeurIPS, 2023.

[17] Xiaojin Zhu. Machine teaching: An inverse problem to machine learning and an approach toward optimal education. In AAAI, 2015.

[18] Xiaojin Zhu, Adish Singla, Sandra Zilles, and Anna N Rafferty. An overview of machine teaching. arXiv preprint arXiv:1801.05927, 2018.