Nonparametric Teaching

Nonparametric teaching (NT) (Zhang et al., 2023b;a) presents a theoretical framework to facilitate efficient example selection when the target function is nonparametric, i.e., implicitly defined. T_{max} T_{max} and T_{max} and T_{max} is given a mathematic learning algorithm and al

Specifically, *machine teaching* (Zhu, 2015; Liu et al., 2017; Zhu et al., 2018) considers the design of a training set (dubbed the teaching set) for the learner, with the goal of enabling speedy convergence towards target functions. **nparametric Teaching**
 parametric teaching (NT) (Zhang et al., 2023b;a)
 al framework to facilitate efficient example selectior

tion is nonparametric, i.e., implicitly defined.

stifically, *machine teaching* (Zhu, 2

NT (Zhang et al., 2023b;a) relaxes the assumption of target functions[†] f being parametric (Liu et al., 2017; 2018), which is f can be represented by a set of parameters w , $e.g., f(x) = \langle w, x \rangle$ with input x , to encompass the teaching of nonparametric target functions. Nonparametric Iterative Machine Teaching

(b) Nonparametric IMT

. Figure 1 provides and $\mathcal{L}_{\mathcal{A}}$ provides and $\mathcal{L}_{\mathcal{A}}$ provides and $\mathcal{L}_{\mathcal{A}}$

∗

Nonparametric Teaching of Implicit Neural Representations Chen Zhang^{1*}, Steven Tin Sui Luo^{2*}, Jason Chun Lok Li¹, Yik-Chung Wu¹, Ngai Wong¹ ¹The University of Hong Kong ²The University of Toronto $P = \{P \mid \text{V} \cap \text{V} \cap \text{V} \mid \text{V} \mid \text{V} \cap \text{V} \mid \text{$ \mathbb{R}^{n} (see Fig.). The contract of \mathbb{R}^{n} 0.5 0.25 tions two products presentati

Figure 1. Comparison between parameterization between parameterization ϵ jeheral ior different tasks, $e.g.,$ square loss for reg \overline{a} $\frac{1}{2}$ is a derivative of GFT is based on the maximization of $\frac{1}{2}$ is based on the maximization of $\frac{1}{2}$ $t_{\rm{max}}$ introduced introduced introduced in P der to steepen gradients. These two teaching algorithms are two teaching algorithms algorithms algorithms are
The steepen gradients are two teaching algorithms algorithms are two teaching algorithms and the steepen gradi $^\top$ The loss ${\cal L}$ can be general for different tasks, *e.g.*, square loss for regression and hinge loss for classification.

to a nonparametric target function function function $\mathcal{L}_\mathcal{A}$ **Implicit Neural Representations**

 $\sum_{i=1}^{n}$ **Implicit neural representation** (INR) (Sitzmann et al., 2020b; Tancik et So on mouding a given signal, willow id through the use of an overparameterized multilayer perceptron (MLP) nal is accurately fitted by this MI P prese 2020 , and then derive two algorithms (one picks examηϵ˜ al., 2020) focuses on modeling a given signal, which is often discrete, such that the signal is accurately fitted by this MLP preserving great details.

amatarizad MI D innute low-dimanejana Such an overparameterized MLP inputs low-dimensional coordinates of the given signal and outputs corresponding values for each input loca-LP maps 2D input coordinates to their t_{total} ando of stochastic gradient descent (Ruder, 2016; Hardt descent (Ruder, 2016; Hardt descent (Ruder, 2016; Hard tion, *e.g.*, the MLP maps 2D input coordinates to their respective 8-bit **Example 1914** 1914 of the section levels for a grayscale image.

The Bridge Between I

Its a theo-

The evolution of an MLP

the target

ters, whereas nonparam-

means of function evolut

Zhu et al.,

Bridging this (theoretical

examination prior to the

context of INR. 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.

metric learners in IMT.

sufficient to increase the gradient, which means ▶ **INT on multiple real-world modalities.** $(M|P)$ The open directions is presentational evoluations de

• We theoretically analyze the asymptotic behavior of both **Main Contribution**

b; Tancik et **Our key contributions are**:

The Bridge Between NT and INRs: Neural Tangent Kernel

\mathbf{I} , or obtained principal projections and second principal component projections are \mathbf{I} \blacktriangle √ 2 and − 4 √ 2 , respectively. Moreover, the discrepancy between 4 **Experiments and Results**

f_{eff} and θ are a rate of f_{eff} . ▶ Toy 2D Cameraman fitting.

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.

∴ Reconstruction quality of SIREN. (b) trains SIREN wi andomly selected pixels. (d) trains it us ט נומוווא (
דומו $\frac{1}{2}$ $\overline{\mathbf{a}}$ ing INT $\overline{}$ of 20% sing selection rate progressively from 20% to increasing selection rate progressively from 20% to 100%). derstallers of ϵ is a constant of ϵ of ϵ of ϵ is the entire theories. μ (W/U) into using an pixels. (U) trains it W/U into using the the basis correspondence that μ schon rale. (σ) trains it using progressive in Fig. , 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.*,

Neural Tangent Kernel (Jacot et al., 2018; Lee et al., 2019) 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).

- bound expressed by the discrepancy of items in the dis definition in Definition 10, and we define the ITD of T representation (INR) via the theoretical lens of nonparametric teaching, which in ig great de-

turn enables the utilization of greedy algorithms from the latter to effectively bolster tha </sup> ▶ We propose Implicit Neural Teaching (INT) that novelly interprets implicit neural the training efficiency of INRs.
- \mathcal{L} miput toda postive of the contract of the analysis of nonparametric teaching provided in the applicability of nonparametric teaching towards deep learning. We further show that the dynamic NTK, derived from graers (Chen et al., 2018; Tabibian et al., 2018; Tabibian et al., 2019), linear learners (Chen et al., 2019), li
1918; Tabibian et al., 2019), linear learners (Chen et al., 2019), linear learners (Chen et al., 2019), linear (Liu et al., 2016), reinforcement learners (Kamalaruban learners (Kamalaruban learners (Kamalaruban learners (
Etimologische learners (Kamalaruban learners (Kamalaruban learners (Kamalaruban learners (Kamalaruban learners et al., 2019; 2019; 2019; 2020b) along with forest along with forest along with forest and q radient descent. ▶ 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 teachdient descent on the parameters, converges to the canonical kernel of functional tea 389
389 - Julie II
389 - Julie II (1890 - 1890 - 1890 - 1890 - 1890 - 1890 - 1890 - 1890 - 1890 - 1890 - 1890 - 1890 - 1890 - 18
- av Scale **•• •• We showcase the effectiveness of INT through extensive experiments in INR train**dy Judiu et al., 2017 ing 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. $2021; 2021$: I VI
... 50111 \overline{a} (Right) Red - R-Cosine; Blue - Cosine; Yellow - Incremental.

 F_{IV} INT saves training time for 1D audio shapes (-35.54%), while upkeeping its

INT Workflow and Algorithm

 \overline{a} : Prograpcion of INIT cologitad pivole (marked as black e. Filogression of intrasted corresponding the corresponding to the corresponding to the corresponding to the c fθ International International allows with the applied variant of the applied variant of the greedy control of the
International allows with the greedy control of the greedy control of the greedy control of the greedy control Figure: Progression of INT selected pixels (marked as black) at corresponding iterations when training with INT 20%.

∗ <u>ှ</u> \mathfrak{p} $\overline{\mathsf{S}}$ urec $\overline{}$ $cluc$ \mathcal{L} entire space to the subspace of concern (*i.e.*, spanned by alundy.
———————— The encoding time is measured excluding data I/O latency.

stant (Kakade & Tewari, 2008; Arjevani et al., 2008; Arjevani et al., 2008; Arjevani et al., 2016; Arjevani et
Distrikt (Kakade & Tewari, 2008; Arjevani et al., 2016; Arjevani et al., 2016; Arjevani et al., 2016; Arjevani

