非線形波動方程式に基づくニューラルネットワークにおける波面整合学習

Wavefront Matching Training for Nonlinear Wave Equation Based Neural Network

NTT 先デ研　○中島　光雅，橋本　俊和

NTT Device Technology Labs., Mitsumasa Nakajima and Toshikazu Hashimoto

E-mail: mitsumasa.nakajima.wc@hco.ntt.co.jp

【背景】

近年、ディープニューラルネットワーク(DNN)を常微分方程式(ODE)としてみなす、Neural-ODE が機械学習分野で注目されている[1]。この枠組みでは、微分方程式 $x(t+1)=f[x(t),\theta,t]$を層の発展とみなし、出力演算や学習は ODE を解くことによって行われる[図 1(b)]。これまでに、Neural-ODE で取り扱う微分方程式を光導波路内の光波伝搬を記述する波動方程式であるシュレディンガー方程式へと拡張してきた[2]。これによって、小型な光導波路中に大規模なネットワークを実装することが可能となる。本稿では、学習法に波面整合法[3]を用いる影響について調べたので報告する。

【波動方程式に基づく深層ニューラルネット】

光導波路中の波動伝搬等をあらわすシュレディンガー方程式を Neural-ODE の枠組みで考える[図 1 (a)]。この発展は、複素数型の畳み込みニューラルネットワーク(CNN)と等価であり、局所的な屈折率の分布が学習重みに相当する。非線形活性化関数は、カー効果などの非線形ハミルトニアンを適用することで物理実装可能である。順伝搬はシュレディンガー方程式を解くことで計算できる。逆伝搬は、誤差逆伝搬法の連続近似と同等である Adjoint 法によって計算可能である。いずれも有限差分法などの物理計算法を用いて計算可能である。屈折率は実部項と虚部項が存在し、それぞれ光波の位相と増幅・吸収に相当する。本稿では、屈折率の虚部項の更新を無視し、実部のみを更新する手法を波面整合法[3]と呼び、本手法のニューラルネットワーク性能への影響を調べた。

【数値実験】

本手法を汎用的な機械学習プラットフォームである Pytorch 上に実装し、MNIST タスクを用いて導波路部の長さ（ネットワーク深さに相当）に対する分類精度の依存性を調べた。結果を図 2 に示す。屈折率の実部のみ(赤線、波面総合法に相当)、虚部のみ（緑線）、両方（青線）を学習する場合を同時にプロットしている。図からわかるように、虚部を学習する場合は学習が安定せず、ネットワーク深さが増すほどに精度が低下する。これは、重み行列の強度成分を学習したことにより、勾配爆発が発生し、学習が安定していないことを示唆している。一方、波面整合法で学習する場合は、このような不安定性は観測されなかった。これは、重み行列がユニタリ行列となり、強度情報が保存されるため勾配爆発が発生しなかったためと考えられる。また、精度は深さに対して単調的に増加し、得られた精度は 95.8%に達した。以上から、波面整合法による学習の有効性が示唆された。

[1] T. Q. Chen *et al.*, “Neural ordinary differential equations,” NeurIPS 2018, pp. 6571-6583 (2018).

[2] M. Nakajima *et al.*, “Neural Schrodinger Equation: Physical Law as Neural Network,” arXiv:2006.13541.

[3] T. Hashimoto *et al.*, Optics letters 30, 2620-2622 (2015).

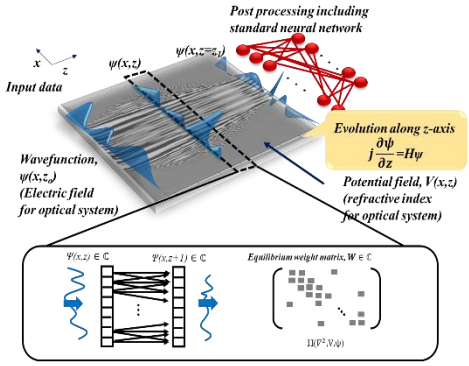

Fig. 1. Neural network based on Schrodinger equation.

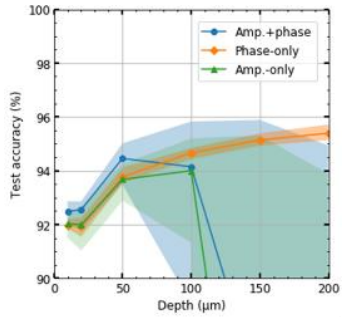

Fig. 2. Depth dependency of MNIST test accuracy.