
FCS 文章精要 国防科技大学李东升等：精度随层容量递减的深度神经网络高效训练优化方法

作者：writer 来源：科学网

本文原地址：<https://www.iikx.com/news/progress/35797.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

FCS 文章精要

国防科技大学李东

升等：精度随层容量递减的深度神经网络高效训练优化方法。论文标题：Efficient deep neural network training via decreasing precision with layer capacity

期刊：Frontiers of Computer Science

作者：Ao SHEN, Zhiquan LAI, Tao SUN, Shengwei LI, Keshi GE, Weijie LIU, Dongsheng LI

发表时间：30 Oct 2024

DOI：10.1007/s11704-024-40669-3

微信链接：点击此处阅读微信文章

Front. Comput. Sci., 2025, 19(10): 1910355
<https://doi.org/10.1007/s11704-024-40669-3>

RESEARCH ARTICLE

Efficient deep neural network training via decreasing precision with layer capacity

Ao SHEN (✉)¹, Zhiquan LAI (✉)¹, Tao SUN¹, Shengwei LI¹, Keshi GE^{1,2},
Weijie LIU¹, Dongsheng LI (✉)¹

1 National Key Laboratory of Parallel and Distributed Computing, National University of Defense Technology,
Changsha 410073, China

2 College of Computer, National University of Defense Technology, Changsha 410073, China

© The Author(s) 2024. This article is published with open access at link.springer.com and journal.hep.com.cn

引用格式：

Ao SHEN, Zhiquan LAI, Tao SUN, Shengwei LI, Keshi GE, Weijie LIU, Dongsheng LI Efficient deep neural network training via decreasing precision with layer capacity[J]. Frontiers of Computer Science, 2025, 19(10): 1910355

阅读原文：



问题概述

现代深度神经网络（DNN）取得的卓越性能依赖于大量的训练数据和参数，导致了巨大的计算成本。因此，低精度训练已成为一种实际的方法来优化训练过程，节省时间、内存和能量成本。然而，之前的方法普遍将低精度视为需要最小化的噪声源，而没有考虑其潜在的优势。

技术步骤

受近期研究表明噪声在DNN训练中存在潜在好处的启发，研究人员系统地研究了利用较低精度的优势。他们发现,低精度训练理论上可以优化DNN性能，并起到正则化的作用，从而增强模型泛化能力。基于这些分析，作者提出了精度随层容量递减的训练方法DPC，该方法可自动确定精度边界，并根据层容量以对数调度模式来变化精度。

实验结果

针对多个数据集和各种应用领域的模型进行的广泛实验，证明了所提出的DPC技术的持续有效性。令人惊讶的是，作者发现即使没有明确追求降低低精度相关误差，低精度训练也能优化模型性能，从而实现DNN准确性和训练效率的双赢。实验结果显示，在效率和优化性能方面都有所改善。

期刊简介

Frontiers of Computer Science (FCS) 是由教育部主管、高等教育出版社和北京航空航天大学共同主办，南京大学支持，SpringerNature 公司海外发行的英文学术期刊。本刊于 2007 年创刊，月刊，全球发行。主要刊登计算机科学领域具有创新性的综述论文、研究论文等。本刊主编为周志华教授，共同主编为熊璋教授。编委会及青年 AE 团队由国内外知名学者及优秀青年学者组成。本刊被 SCI、Ei、DBLP、INSPEC、SCOPUS 和中国科学引文数据库 (CSCD) 核心库等收录，为 CCF 推荐 B 类期刊；两次入选中国科技期刊国际影响力提升计划；入选第 4 届中国国际化精品科技期刊；两次入选中国科技期刊卓越行动计划（一期梯队、二期领军）。



Frontiers of Computer Science

2024 IF = 4.6

Fast and high quality peer review
 Rapid publication upon acceptance
 Fast Track of Revised Top Conference Submission

SCI (E) , EI
 CCF B
 Scopus, DBLP, INSPEC

Paper types

- Perspective/News&Highlights
- Review/Viewpoints
- Research Article
- Letter/Others

Subject areas, (not limited to)

- Architecture
- Software
- Artificial intelligence
- Theoretical computer science
- Networks and communication
- Information systems
- Image and graphics
- Information security
- Interdisciplinary

主 编



周志华
南京大学

共同主编



熊 璋
北京航空航天大学

副主编



许 可
北京航空航天大学

副主编



周 磊
浙江大学

执行编委



陈 力
浙江大学



高 跃
复旦大学



廖小飞
华中科技大学



张 健
中科院软件所



张敬
东南大学



陈 俊
南京大学

在线投稿网址:
<https://mc.manuscriptcentral.com/hepfc>
 Tel: +86(10) 8233 8176
 E-mail: FCS@pub.hep.cn






中国学术前沿期刊网

<http://journal.hep.com.cn>

来源：Frontiers of Computer Science

更多 科学进展 请访问 <https://www.iikx.com/news/progress/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](https://www.iikx.com)转发