

热点追踪

神经场的网格模型正切核理论 (GTK)

赵泽林¹ 范凤磊² 廖文龙³ 严骏驰⁴¹上海交通大学 ²香港中文大学 ³上海交通大学 酷哇科技 ⁴上海交通大学

本文获得CVPR24最佳论文提名，在盲审(pre-rebuttal)阶段获得三个审稿人的一致满分意见(5/5/5)。许多当代研究利用基于网格的模型来表示神经场，但对这些模型的系统分析仍然缺失，阻碍了这些模型的改进。因此，本文引入了一个基于网格模型的理论框架。该框架指出，这些模型的逼近和泛化行为由网格切线核(GTK)决定，GTK是基于网格模型的内在属性。所提出的框架促进了对各种基于网格模型的一致和系统的分析。此外，该框架还激发了一个名为乘法傅里叶自适应网格(MulFAGrid)的新型基于网格模型的发展。数值分析表明，MulFAGrid相较于其前身具有更低的泛化界限，表明其具有强大的泛化性能。如图1所示，MulFAGrid在包括二维图像拟合、三维符号距离场(SDF)重建和新视图合成在内的各种任务中实现了最先进的性能，展示了卓越的代表能力。本工作也即将在Jittor深度学习框架平台进行实现和开源。

一、研究背景

首先，我想介绍一下神经场及其广泛的应用。神经场是基于坐标的网络，表示一个场，实质上是一种连续参数化，代表一个物体或场景的物理量。神经场在计算机视觉和其他研究领域的各种任务中显示出了显著的进展和潜力。

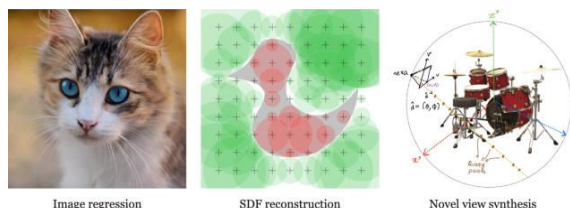


图1 神经场的应用场景举例

我们的工作主要针对网格模型(grid-based models)，这是一类主要的神经场模型。网格模型在参数化和功能上与传统的神经网络(如MLP)有根本不同。主要的不同包括MLP往往包含多层非线性神经网络，并且MLP没有显示查询的过程，同时MLP的输入也不一定非是位置坐标。如图2所示，网格模型以查询坐标为输入，该坐标被发送到下标函数以从网格中获取一组特征向量。然后，模型输出核函数和这些特征向量的加权平均值。该模型需要学习的主要是特征向量。最简单的核函数是不含参数的插值算法(如最近邻算法或者双线性插值算法)。核函数里面也可以包含可学习的参数。

图3是英文版的网格模型的示意图，选择不同的下标函数可以让我们的算法适配不同的网格模型。

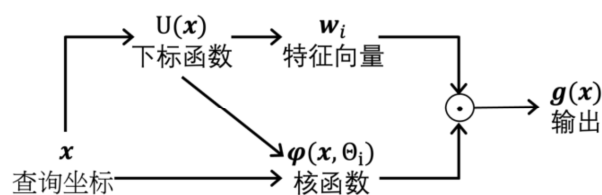


图2 网格模型的通用示意图

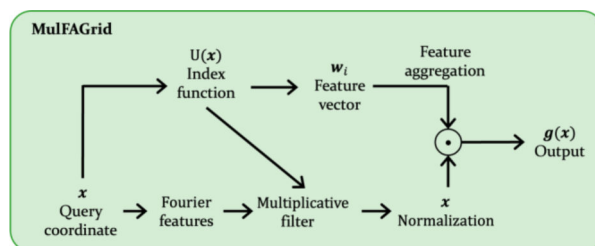


图3 我们提出的新的网格模型 MulFAGrid 的示意图

二、理论介绍

1. 主要研究问题

我们的理论旨在通过三个主要问题来理解和增强网格模型：

- 1、我们如何理解网格模型的训练动态？
- 2、我们如何衡量网格模型的泛化性能？
- 3、我们如何设计一个更好的网格模型？

为了解决这些问题，我们提出了一个基于正切核的理论框架。

这里要先介绍一下什么是正切核。正切核这一概念来自于著名的深度学习理论文章神经正切核^[1] (Neural Tangent Kernels, NTK)。NTK 是一种核函数，最初由研究者在研究神经网络的训练过程时提出的。当神经网络在参数空间中靠近其初始值时，通过对神经网络梯度下降优化过程的分析，发现网络的行为可以用一个固定的核函数来描述，这个核函数就是神经正切核。比如神经网络的输出可以用其参数的梯度来表示。在训练过程中，网络参数的更新会导致输出的变化，而这种变化在参数空间中的变化速率可以用梯度来表示。NTK 定义了两个输入数据点的输出变化之间的相似度。形式上，对于输入数据点 x_i 和 x_j ，神经正切核 $\theta(x_i, x_j)$ 可以定义为网络输出对参数的梯度的内积：

$$\theta(x_i, x_j) = \left\langle \frac{\partial f(x_i, \theta)}{\partial \theta}, \frac{\partial f(x_j, \theta)}{\partial \theta} \right\rangle$$

其中， $f(x, \theta)$ 是神经网络的输出， θ 是网络的参数。

2. 理论成果

我们的理论结果表明，网格模型的近似和泛化性能与网格切线核 (GTK) 有关。GTK 被定义为一个正半定矩阵，它测量梯度空间中两个数据点之间的距离。这里我们展示了 GTK 的定义： g 是由 $w(t)$ 参数化的网格模型， X 是一个数据集，其中 X_i 是第 i 个数据。GTK 可以这样表示：

$$[G_g(t)]_{i,j} = \left\langle \frac{\partial g(X_i, w(t))}{\partial w}, \frac{\partial g(X_j, w(t))}{\partial w} \right\rangle,$$

注意这个形式跟神经正切核 (NTK) 的形式是吻合的，因为他们都是正切核，他们的主要区别是适用的模型不同，GTK 主要适用于网格模型。后面可以看出，因为网格模型本质上比较简单纯粹，所以 GTK 的理论基本不需要近似，但是 NTK 的理论需要网络无穷宽的假设才能成立。

我们的定理一（网格模型优化定理）说明，网格模型的模型参数根据微分方程（如下面方程所示）演化。

$$\frac{dO(t)}{dt} = -G_g(t)(O(t) - Y)$$

这里 $O(t)$ 表示网格模型的输出， $G(t)$ 表示网格模型的 GTK，而这里的 Y 表示数据集的标签（向量化， Y_i 表示第 i 个数据的标签），这个定理有什么意义呢？直观地讲，有了这个定理我们就可以预测模型的效果（不用亲自炼丹即可确定模型的好坏），这一定理在理论上是有很高价值的，也是朴素成立的，不依赖于具体的模型细节。

接下来，我们提出了另一个定理 (GTK 不变定理)。定理 2 指出，网格模型的 GTK 在训练期间保持不变。（这是一个非正式的说法，如果了解正式版本的话请参考我们的论文，相关证明在论文的附录中有，论文以及附录在 arxiv 可以下载）这意味着无论网格模型的大小如何，初始 GTK 在整个训练过程中保持恒定。这一定理揭示了 GTK 是由模型和数据集决定的一个内在特性，与模型的训练过程无关，有了这个定理，自然也不难理解网格模型的很多性质都与 GTK 有关了。

定理三，描述的是网格模型的泛化性能。在理论深度学习中，泛化性能的好坏通常由泛化界 (generalization bound) 来刻画。该定理揭示了网格模型的泛化界由一个特定的度量 Δ 决定，而 $\Delta = Y^T G^{-1} Y$ ，与网格模型的 GTK 和数据集的标注有关。形式化的说，该泛化界提供了模型性能的概率保证。该定理说明了模型的泛化性能既与 GTK 有关，也与数据集的结构有关。结合该定理与 GTK 的特征值，我们能获得更多关于泛化性能的信息。

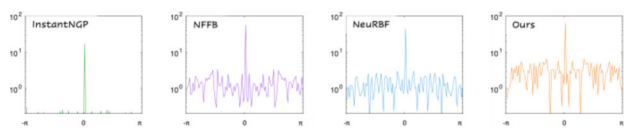


图4 傅立叶频谱分析结果

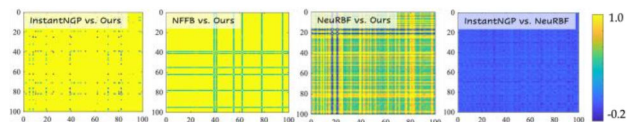


图5 泛化性能对比可视化

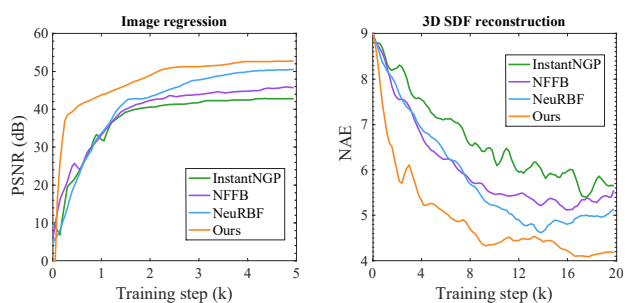


图6 2D神经场和3D神经场不同方法训练曲线

3. 新的网格模型

我们的 GTK 理论可以激发未来设计具有更好训练和泛化性能的网格模型。我们介绍了一种新的网格模型,名为 MulFAGrid 如图 3 所示。该模型使用傅里叶特征来提升高频信号的学习,并采用乘法滤波器来为模型提供节点信息。我们的模型示意图如图 3 所示。

三、实验分析与结果

然后,我们基于 GTK 理论展示了一组数值实验。我们对比较常见的网格模型如图 4,包括 InstantNGP^[2], NFFB^[3], NeuRBF^[4]等等。首先,在频谱分析中, MulFAGrid 显示了比较宽的频谱,特别是在高频域。这一特性导致它的高频成分的收敛速度更快。

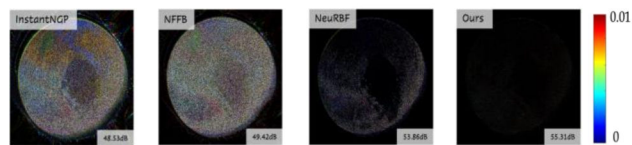


图7 图片拟合结果对比

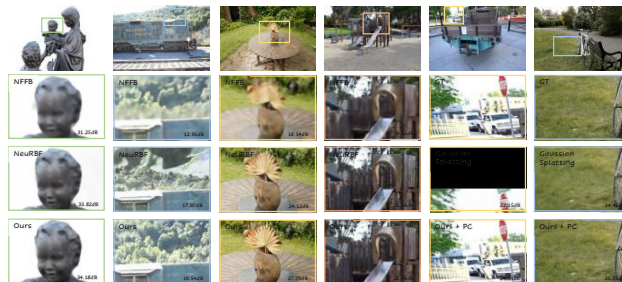


图8 新视角生成的结果对比

在下一个实验中,我们构建了一个包含两个数据点及其对应标签的数据集。如图 5 所示 MulFAGrid 对于大多数标签值表现出更紧的泛化界表明其性能更好。

图 6 展示了各种基线方法和我们的误差图。误差图衡量预测图像与真实图像的差异。MulFAGrid 提供了更准确的拟合,展示了其优越的性能。

图 7 我们展示了我们的模型在拟合二维图像和三维符号距离函数 (SDF) 方面的性能。结果突出了 MulFAGrid 的准确性和效率。

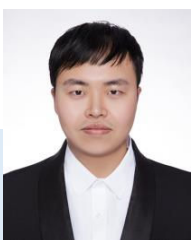
最后,我们在图 8 展示了我们在新视角合成方面的结果。详细结果表明, MulFAGrid 在生成高质量的新视角方面表现出色,突显了其实际应用性。

我们论文的 Jittor 版本会在 Project Page 里放出,链接为 <https://sites.google.com/view/cvpr24-2034-submission/home>。敬请关注。

责任编辑 魏秀参

参考文献

- [1] Jacot, Arthur, Franck Gabriel, and Clément Hongler. "Neural tangent kernel: Convergence and generalization in neural networks." Advances in neural information processing systems 31 (2018).
- [2] Müller, Thomas, et al. "Instant neural graphics primitives with a multiresolution hash encoding." ACM transactions on graphics (TOG) 41.4 (2022): 1-15.
- [3] Wu, Zhijie, Yuhe Jin, and Kwang Moo Yi. "Neural fourier filter bank." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.
- [4] Chen, Zhang, et al. "Neurfb: A neural fields representation with adaptive radial basis functions." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.



赵泽林

赵泽林，获得上海交通大学计算机系本科学位，即将加入佐治亚理工学院攻读博士学位。上海交通大学本科毕业，本科专业排名 2/150，本科期间获得国家奖学金，上海市奖学金，曾在 NeuRIPS, ECCV, CVPR, AAAI 等顶会发表四篇一作论文，引用数超过 600。

Email: sjtuytc@gmail.com



范凤磊

本文通讯作者来自香港中文大学数学系研究助理教授范凤磊博士，他所在的 Center for Mathematical AI 由曾铁勇教授创立。中心自 2018 年成立以来，在中心主任曾铁勇教授的带领下，先后承担科技部国家重点研发计划项目等一系列关键项目。

范凤磊博士于美国伦斯勒理工学院 (Rensselaer Polytechnic Institute) 获得博士学位，导师为国际知名影像专家王革教授，主要研究方向是脑启发智能以及神经网络的数学理论，在 JMLR, TMI, TNNLS, TCI 等杂志发表论文二十余篇，引用数过千。曾获得 IBM AI Horizon Scholarship 和国际神经网络协会 (INNS) 2021 年杰出博士论文奖。

Email: flfan@math.cuhk.edu.hk



严骏驰

严骏驰教授带领实验室发表第一/通讯作者 CCF-A 类论文超百篇，谷歌引用过万次，获 PaperDigest 评选的最具影响力 AAAI21、IJCAI23 论文榜首。

严骏驰教授长期任机器学习三大会议 ICML/NeurIPS/ICLR 领域主席，模式识别旗舰期刊 TPAMI、PRJ 编委。实验室学生获得挑战杯特等奖、CCF 优博/CV 新锐奖、交大学术之星等荣誉和本科生自然科学基金。

Email: yanjunchi@sjtu.edu.cn