

张量分解在神经辐射场渲染中的应用

刘佳润

22221290

计算机科学与技术学院

jiarunliu@zju.edu.cn

Abstract

张量分解 (*Tensor Decomposition*) 是一种在标量分解、向量分解以及矩阵分解算法的基础上进行泛化的经典算法, 其在降维处理、稀疏性分析以及隐性关系挖掘中具有重要的作用。近年来随着三维计算机视觉的高速发展, 一种基于隐式表达的三维场景新视角渲染的方法——神经辐射场 (*Neural Radiance Field, NeRF*) 在该领域取得了极大的成功。然而大多数的 *NeRF* 方法在训练速度、神经网络复杂程度等方面存在许多缺陷。本文结合近两年来前沿工作, 分析张量分解在 *NeRF* 中的应用与改进, 分析其优越性并进行推演梳理和总结。

1. 引言

在计算机科学中, 张量是对矩阵这一概念的扩充, 作为一种数据容器, 在诸如计算机图形学、计算机视觉、并行计算、深度学习等领域有非常重要的意义。相应地, 张量分解技术在其中也具有非常重要的地位。与矩阵分解类似, 张量分解具有数据降维、稀疏数据分析与填充、隐性关系挖掘的重要作用, 尤其是对于稀疏张量, 张量分解是一种常用的分析方法。

在三维计算机视觉领域, 一种近年来发展迅速的解决新视角合成问题的方法逐渐被人们所重点关注, 即神经辐射场 (*Neural Radiance Field, NeRF*)。NeRF 是一种面向三维隐式空间建模的基于多层感知机 (*Multi-Layer Perception, MLP*) 的深度学习模型。由于 MLP 的隐式表达, 这种方法可以非常简洁的对三维场景进行建模。然而, 由于深度学习的网络参数量大、模型结

构复杂, 导致许多经典 NeRF 算法难以避免训练、推理速度慢、模型空间大等问题。针对此, 许多近期工作在 MLP 的基础上, 加入一些显式的混合表达方式将一些 3D 特征存储在显示的数据结构中, 从而优化其效果。

将张量分解技巧应用到 NeRF 问题中, 是一种非常优雅的数学方法。由于辐射场和三维空间的特性十分契合高阶张量的表示方法, 同时具有很高的稀疏性质, 因此利用张量分解, 将辐射场分解为若干紧凑度高的成分, 可以很好的对辐射场的特征值、表征值等进行建模。近年如【1】、【2】等工作将张量分解应用到 NeRF 场景中, 获得了很好的提升效果, 并证明了这一工作的可扩展性, 【3】、【4】等工作在此基础上的扩展与优化也验证了该方法的高效。

本文结合以上提到的若干工作进行脉络梳理与方法总结, 并探索使用张量表达以及张量分解在辐射场视觉问题上的原理和可优化原因。本文第 2 节将讨论张量分解的原始形式数学原理以及 NeRF 的基本原理。本文第 3 节将具体讨论三种张量分解方法以及其在 NeRF 中的结合成果, 其中 3.1、3.2 将针对【TensorRF】的方法进行原理剖析, 3.3 将针对【CCNeRF】的方法进行原理剖析, 并指出其与前两节的联系。本文第 4 节将具体分析以上方法在 NeRF 中的应用, 以及通过实验结果进行分析并探讨该方法的优越性。本文第 5 节将做出总结, 并讨论未来可扩展研究方向。

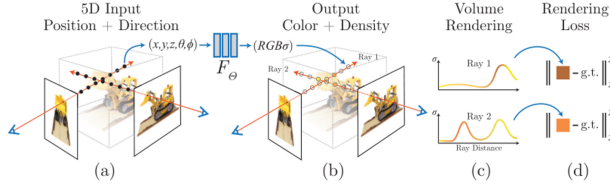


图 1. NeRF方法原理示意图

2. 背景知识

2.1. 低阶张量分解

矩阵是一种典型的低阶（二阶）张量，因此可以从矩阵分解中理解张量分解的意义。常见的矩阵分解方式很多，本文应用的分解算法思路主要基于奇异值分解（Single Value Decomposition, SVD）。

SVD分解的基本思想是将矩阵分解为两个因子矩阵和一个奇异值对角矩阵。对于任意一个矩阵 $\mathbf{M} \in \mathbb{R}^{m \times n}$ ，可以将其写作如下形式：

$$\mathbf{M}_{m \times n} = \mathbf{U}_{m \times m} \mathbf{\Sigma}_{m \times n} \mathbf{V}_{n \times n}^T \quad (1)$$

其中， \mathbf{U}, \mathbf{V}^T 是酉矩阵， $\mathbf{\Sigma}$ 主对角线上的元素即为原矩阵 \mathbf{M} 的奇异值。这样的分解方式可以寻找数据分布的主要维度，将原始的高维数据映射到低维子空间中，从而实现数据降维，或是寻找到原始数据的主要特征。

2.2. 神经辐射场（NeRF）

给定一组图像集，NeRF旨在渲染新视角下的照相级别的逼真图像。NeRF的基本思想是使用一个5自由度参数（目标点位置 $\mathbf{x} \in \mathbb{R}^3$ ，观测方向向量 $\mathbf{d} \in \mathbb{R}^2$ ）来表示任意一张新视角合成图像，并利用体渲染方法来提高渲染质量。

NeRF利用MLP对场景进行软编码，通过上述输入分别学习体密度（不透明度场） σ 以及依赖于视图的表面纹理的辐射场（颜色） \mathbf{c} 。通过体渲染方式，根据训练出的不透明度场和辐射场以及给定光线和观测方向，通过采样和积分的方法可以得到像素颜色：

$$\mathbf{C}(\mathbf{r}) = \int_{t=0}^{\infty} \sigma(\mathbf{o} + t\mathbf{d}) \cdot \mathbf{c}(\mathbf{o} + t\mathbf{d}, \mathbf{d}) \cdot e^{-\int_{s=0}^t \sigma(\mathbf{o} + s\mathbf{d}) ds} dt \quad (2)$$

上述过程如图1所示。

测试 [1]

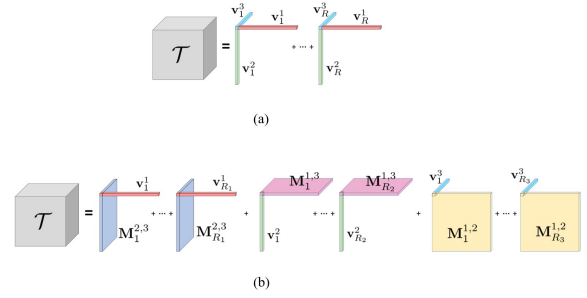


图 2. 两种经典张量分解方式。(a)CP分解，(b)VM分解

3. 张量分解方法

3.1. CP分解

在对SVD分解进行高阶扩张的过程中，一种推演方法被称为CP分解（Canonical Polyadic Decomposition）。SVD分解将矩阵分解为奇异值矩阵和正交矩阵因子之积，而CP分解则是将张量分解为若干秩一张量的和，其中每一个秩一张量可以表示为若干向量的外积。如图2(a)所示，对于一个三维张量 $\mathcal{T} \in \mathbb{R}^{I \times J \times K}$ ，可以对其进行 R 个成分的CP分解：

$$\mathcal{T} = \sum_{r=1}^R \mathbf{v}_r^1 \circ \mathbf{v}_r^2 \circ \mathbf{v}_r^3 \quad (3)$$

其中 \circ 代表矩阵外积，每一个 $\mathbf{v}_r^1 \circ \mathbf{v}_r^2 \circ \mathbf{v}_r^3$ ，对应一个秩一张量成分，其中 $\mathbf{v}_r^1 \in \mathbb{R}^I, \mathbf{v}_r^2 \in \mathbb{R}^J, \mathbf{v}_r^3 \in \mathbb{R}^K$ ，代表第 r 个成分三个分解向量。

CP分解的计算可以通过交替最小二乘法（ALS）实现。CP分解将张量分解为多个向量，表示多个紧致的秩一分量。这个过程是一个对高阶张量进行简化和重点成分提取的过程，因此常用于神经网络的压缩，同时也可能作为辐射场场景的建模。

3.2. VM分解

CP分解的方式保证了多个张量分量的低秩性与紧凑性，然而由于紧凑性太高，CP分解可能需要许多组件来对复杂场景进行建模，导致辐射场重建的计算成本很高。在此基础上TensorRF【引用】工作进行了扩展，引入了更加灵活的VM分解（Vector-Matrix Decomposition）。

如图2(b)所示，与利用纯向量因子的CP分解不同，

VM分解将张量分解为多个向量和矩阵：

$$\mathcal{T} = \sum_{r=1}^{R_1} \mathbf{v}_r^1 \circ \mathbf{M}_r^{2,3} + \sum_{r=1}^{R_2} \mathbf{v}_r^2 \circ \mathbf{M}_r^{1,3} + \sum_{r=1}^{R_3} \mathbf{v}_r^3 \circ \mathbf{M}_r^{1,2} \quad (4)$$

具体而言，该方法不是使用单独的向量，而是将每两种模式组合起来，并用矩阵表示它们，从而允许用较少的分量对每个模式进行充分的参数化。对每个分解成分，以第一项为例，第一个模式 \mathbf{v}_r^1 一定是秩为1的，而第二模式和第三模式的秩可以随意，取决于组合的矩阵 $\mathbf{M}_r^{2,3}$ 的秩。这样就降低了CP分解的紧凑性要求，同时可以指定分解的成分数量 R_1, R_2, R_3 ，根据每个模式的复杂性来另外设置。

注意到在CP分解中，每个分量张量都比一个分量有更多的参数。虽然VM分解会导致较低的紧凑性，但VM分量张量可以表达比CP分量更复杂的高维数据，从而在对同一复杂函数建模时减少所需的分量数量。另一方面，与体素或网格的表示方法相比，VM分解仍然具有非常高的紧凑性，但是能够将计算的复杂性从 $O(N^3)$ 降低为 $O(N^2)$ 。

3.3. HY分解

VM分解虽然放宽了CP分解的约束，同时发明了一种符合三维场景以及特征通道结合的建模方式，但是仍然具有一定的缺陷。首先，VM分解中使用MLP进行场景建模会扰动秩分解的结果并使得其无法进行压缩与组合。其次，VM分解虽然有效结合了秩一向量和矩阵的组合，但是每一个成分内的向量和矩阵的结合过于耦合，导致整体建模的灵活程度收到一定影响。因此，【】提出的HY分解（Hybrid Variant Decomposition）本质上是对VM分解的进一步解耦，将向量因子和矩阵因子分别分解。

具体而言，在CP分解这种向量分子分解的方法基础上，可以进行扩展。首先，从分解成分上，可以将分解成分由低秩紧密的向量放松为向量组合的矩阵，可以将式3.1修改为：

$$\mathcal{T} = \sum_{r=1}^{R_2} \mathbf{M}_r^{1,2} \circ \mathbf{M}_r^{1,3} \circ \mathbf{M}_r^{2,3} \quad (5)$$

其中的 \mathbf{M} 代表沿着三个平面的银子矩阵。直观地，这种变式指的是首先沿着每个轴对原始3D空间进行切片和平铺，然后在 $\mathbb{R}^{XY \times XZ \times YZ}$ 上进行CP分解。将这种变式被称为TP分解（Triple Plane Decomposition）。

此外，注意到对于每个单独的秩，基本的基于向量或矩阵的分解可以独立选择。因此，提出了一种将上述CP和TP分解相结合的混合变体分解（Hybrid Variant, HY）。我们可以灵活的设置两种成分的分解个数，记为 $R = R_{\text{vec}} + R_{\text{mat}}$ 。HY分解的公式如下：

$$\mathcal{T} = \alpha \sum_{r=1}^{R_{\text{vec}}} \mathbf{v}_r^1 \circ \mathbf{v}_r^2 \circ \mathbf{v}_r^3 + \beta \sum_{r=1}^{R_2} \mathbf{M}_r^{1,2} \circ \mathbf{M}_r^{1,3} \circ \mathbf{M}_r^{2,3} \quad (6)$$

4. 辐射场应用与原理分析

4.1. 基于VM分解的张量辐射场

显然，VM分解是对于CP分解的一种泛化表示。在NeRF的建模中，可以通过利用具有每体素多通道特征的规则3D网格 \mathcal{G} 来对这样的函数进行建模。通过特征通道可将其拆分为几何网格 \mathcal{G} 和外观网格 \mathcal{G}_c ，分别对体积密度和视图相关颜色进行建模。其中， \mathcal{G}_c 支持各种类型的外观特征，这取决于预先选择的函数 S ，该函数将外观特征向量和视角方向 d 转换为颜色 c 。函数可以选择小的MLP或者球谐函数。同时，只需要用一个单通道的几何网格 \mathcal{G}_σ 来进行体密度表示，不需要其他任何的转换函数。基于连续网格的辐射场可以表示为：

$$\sigma, c = \mathcal{G}_\sigma(\mathbf{x}), S(\mathcal{G}_c(\mathbf{x}), d) \quad (7)$$

4.2. 基于HY分解的张量辐射场

5. 总结与展望

参考文献

- [1] Giacomo Chiarot and Claudio Silvestri. Time series compression: a survey. *arXiv preprint arXiv:2101.08784*, 2021. 2