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 [4]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） [7]。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来将一些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TensoRF [1]，CCNeRF [8]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，INS [5]、StyleRF [6]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TensoRF [1]的方法进行原理剖析，3.3将针对CCNeRF [8]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。

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）[3]。

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：

$$\mathbf{M}_{m\times n} = \mathbf{U}_{m\times m}\mathbf{\Sigma}_{m\times n}\mathbf{V}^{\top}_{n\times n} \tag{1}$$

其中，$\mathbf{U}, \mathbf{V}^{\top}$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：

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

上述过程如图1所示。

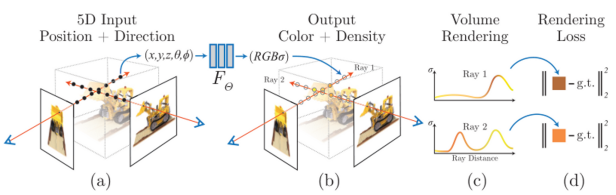

图 1. NeRF方法原理示意图

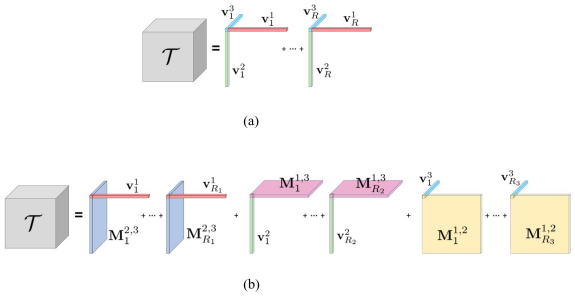

图 2. 两种经典张量分解方式。(a)CP分解，(b)VM分解

NeRF中需要学习的是对σ以及$\mathbf{c}$的预测模型。实践中的经典损失函数是真实颜色与预测颜色的平方误差损失，即只需要通过拍摄到的图像就可以进行监督与训练。

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）[2]。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：

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

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型，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TensoRF [1]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，CCNeRF [8]中提出的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因子矩阵。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）。如图3所示，HY分解实质上是对CP分解和TP分解的加权组合，同时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：

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

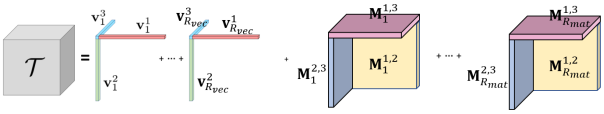

图 3. HY分解图示

4. 辐射场应用与实验分析

4.1. 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$\mathcal{G}_\sigma$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：

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

具体而言，$\mathcal{G}_\sigma \in \mathbb{R}^{I \times J \times K}$，分别代表特征网格沿XYZ轴的分辨率，$\mathcal{G}_c \in \mathbb{R}^{I \times J \times K \times P}$，其中$P$是外观特征通道的数量。3D的几何张量可以由式3.2进行标准的VM分解，4D的外观张量多了一个模式，用来表示

特征通道的维度，通常这个维度要远小于XYZ分辨率，即$P \ll I \approx J \approx K$，因此这个张量的秩会比较低。因此，我们不将该模式与矩阵因子中的其他模式相结合，而是仅使用向量，用$\mathbf{b}_r$表示，用于因子分解中的该模式。两个张量的分解公式分别如下：

$$\begin{aligned}\mathcal{G}_\sigma &= \sum_{r=1}^{R_\sigma} \mathbf{v}_{\sigma,r}^X \circ \mathbf{M}_{\sigma,r}^{YZ} + \mathbf{v}_{\sigma,r}^Y \circ \mathbf{M}_{\sigma,r}^{XZ} + \mathbf{v}_{\sigma,r}^Z \circ \mathbf{M}_{\sigma,r}^{XY} \\ &= \sum_{r=1}^{R_\sigma} \sum_{m \in XYZ} \mathcal{A}_{\sigma,r}^m\end{aligned} \tag{8}$$

$$\begin{aligned}\mathcal{G}_c = &\sum_{r=1}^{R_c} \mathbf{v}_{c,r}^X \circ \mathbf{M}_{c,r}^{YZ} \circ \mathbf{b}_{3r-2} + \mathbf{v}_{c,r}^Y \circ \mathbf{M}_{c,r}^{XZ} \circ \mathbf{b}_{3r-1} + \\ &\mathbf{v}_{c,r}^Z \circ \mathbf{M}_{c,r}^{XY} \circ \mathbf{b}_{3r} \\ = &\sum_{r=1}^{R_c} \mathcal{A}_{c,r}^X \circ \mathbf{b}_{3r-2} + \mathcal{A}_{c,r}^Y \circ \mathbf{b}_{3r-1} + \mathcal{A}_{c,r}^Z \circ \mathbf{b}_{3r}\end{aligned} \tag{9}$$

综上，总共有$3R_\sigma + 3R_c$个矩阵以及$3R_\sigma + 6R_c$个向量来对辐射场进行建模。分解的组分个数通远小于XYZ维度（如8/16/32等），这样会使得产生高度紧凑的表示，可以对高分辨率密集网格进行编码。本质上，XYZ模式的向量和矩阵因子（上式中所有的$\mathbf{v}, \mathbf{M}$）描述了场景几何体和外观沿着它们对应的轴的空间分布。而外观特征模式的向量$\mathbf{b}_r$表示了全局外观相关性。将所有的$\mathbf{b}_r$向量按列堆叠，得到一个$P \times 3R_c$的外观矩阵$\mathbf{B}$，也可以被视为一个全局外观字典，它抽象了整个场景的外观共性。辐射场的建模如图4所示。

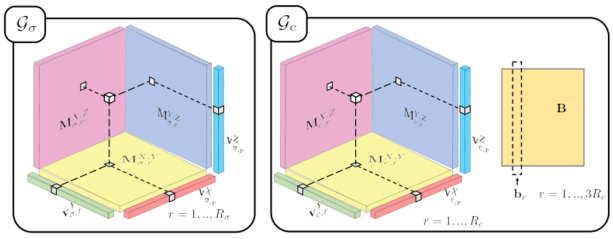

图 4. TensoRF中基于VM分解辐射场建模表示。注意到，每个原始网格中的体素对应着VM分解中的一个XYZ模式上的向量或矩阵因子。右侧的矩阵代表将若干$\mathbf{b}_r$向量拼接而成的全局矩阵$\mathbf{B}$。

基于因子分解的模型可以低成本计算每个体素的特征向量，每个XYZ模式向量/矩阵因子只需要一个值即可进行估计。对张量进行三线性插值则可以构建更为精细的连续场从而提高重建质量。根据该建模得到的辐射场表示为：

$$\sigma, c = \sum_r \sum_m \mathcal{A}_{\sigma,r}^m(\mathbf{x}), S\left(\mathbf{B}\left(\oplus\left[\mathcal{A}_{c,r}^m(\mathbf{x})\right]_{m,r}\right), d\right) \tag{10}$$

其中$\oplus$可以看作是将所有标量值（1通道向量）连接为$3R_c$通道向量的连接运算符。实践中需要并行计算大量体素的时候，只需要首先计算并拼接所有体素，按列向量生成一个矩阵$\mathcal{A}_{c,r,ijk}^m$，然后直接与$\mathbf{B}$做矩阵乘法即可。在给定任何3D位置和观看方向的情况下，都可以获得连续的体积密度和与观测方向相关的颜色。这允许高质量的辐射场重建和渲染。

在训练中，除了真实颜色与预测颜色的误差损失外，为了鼓励张量因子参数的稀疏性，还应用了标准的L1正则化，从而有效地提高外推视图的质量，并在最终渲染中去除噪点。对于输入图像很少或捕获条件不完美的真实数据集，加入了TV损失以加强监督。

4.2. 基于HY分解的张量辐射场

VM分解由于将二者成分进行了耦合，因此实际比例是固定的。HY分解相较于VM分解，更易于灵活控制向量分量和矩阵分量的比例，这也增加了对于辐射场建模的灵活性以及可压缩、可组合性。

为了减少MLP参数过大对于模型的影响以及分解紧凑型的损失，可以使用球谐函数对5-DoF输入中的观测方向进行拟合建模。具体来说，对于一个有最大degree为$\ell_{\max}$ 的球谐函数，需要$(\ell_{\max}+1)^2$ 个球谐系数来对每个通道的颜色进行建模。对于这些系数，也可以用张量来进行表示：

$$\ell_{\max} \mathcal{T}_{i,j,k}^\kappa \in \mathbb{R}^{(\ell_{\max}+1)^2}, \kappa \in \{r, g, b\} \tag{11}$$

每个位置的密度和颜色可以通过以下公式得到：

$$\sigma = \phi\left(\mathcal{T}^{\text{density}}\right) c^\kappa = \psi\left(\sum_{\ell=0}^{\ell_{\max}} \sum_{m=-\ell}^{\ell} \mathcal{T}_{\ell,m}^\kappa Y_\ell^m(\mathbf{d})\right), \kappa \in \{r, g, b\} \tag{12}$$

其中$\phi(\cdot), \psi(\cdot)$ 都是激活函数，$\mathbf{d}$ 是观测方向，Y_ℓ^m 是球谐函数基。通过以上方式，可以在可微渲染的过程中进行分解模式的学习。

除此之外，基于HY分解的张量辐射场还可以通过分解成分的秩的残差来进行场景压缩与组合。参

考SVD分解的原理，秩越低，分解越紧凑，成分所表达的信息越重要。这也是诸如PCA等降维方法的重要原理。假设总共的分解得到R个秩，训练阶段数量为M，可以将所有的秩成分按序列分到M个组当中，记每个组的秩数量为R_m，$m \in \{1, 2, ..., M\}$。在训练过程中，每一次收敛后对秩成分进行固定并附加下一个秩残差项$R_m - R_{m-1}$。通过对秩残差损失的监督，我们可以更快学习到每一个组分对于基本组分的补充信息，从而在很小的模型容量的情况下获得很高的视觉质量。在实践中，这对于LOD（Level-of-Details）的设置有着很重要的效果，能够根据场景的复杂程度以及重要程度进行高保真压缩。

4.3. 实验与分析

我们首先以numpy为后端实现了基础的CP分解作为原理分析的练习，并与经典的张量学习库tensorly的结果进行了对比。我们用一个简单的任务来进行CP分解的性能比较。CP分解的求解用到的是交替优化方法ALS，对于每个因子矩阵相关的子问题，则使用共轭梯度法求解。模拟三阶CP分解以及在维度为$(2, 3, 4)$的三维张量上进行分解与重建精度对比的实验结果如表1所示。

Backend	iters	total time(s)	avg time(s)	error
numpy	30	3.08016	0.10267	0.18030
tensorly	30	5.28228	0.17608	0.16475

表 1. CP分解实验结果

另外，我们根据TensoRF提供的基于VM分解的NeRF代码与原始NeRF的代码进行了运行对比，实验证明张量辐射场的训练和推理时间都有极大的提升，同时建模质量也有了很大的提升。重建效果与深度可视化如图5所示。

从CP分解的实验中可以看出，通过CP分解与重建后张量损失很小，在小规模张量的分解，每次迭代求解只需要大约0.1s。将辐射场进行张量化，利用分量因子进行特征表示与重建，既在体素网格的框架下对目标场景进行了显式特征存储，减少了训练和推理时间，又能够保证很小的几何形体损失，这也说明了这一方法具有非常好的可扩展性。

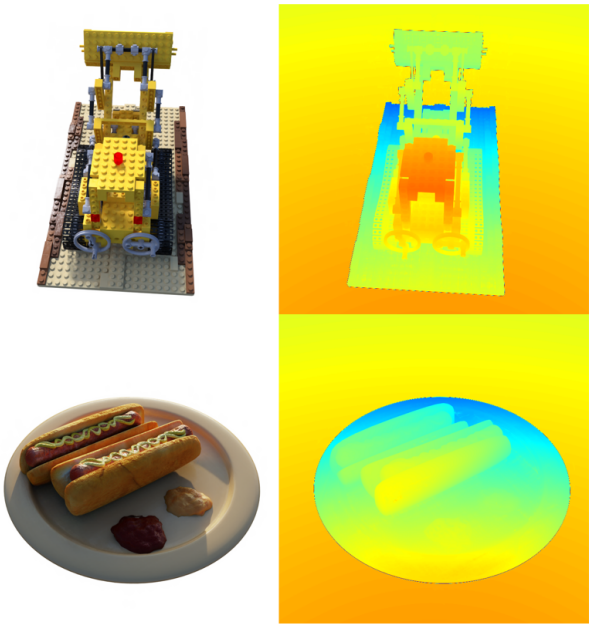

图 5. TensoRF在blender synthetic数据集上的重建效果。每个模型训练时间大约为10 13分钟。

5. 总结与展望

本文针对几篇关于张量辐射场的最新工作的数学原理与应用原理进行了分析，并梳理了从CP分解到VM分解，再到结合了TP分解的HY分解的脉络。在三维建模的应用任务中，张量建模与分解起到了体素建模以及显式存储三维信息的作用。在方法迭代的过程中，张量分解的紧凑性与张量分解的灵活性处于一个动态平衡的过程。从加速与质量的角度考虑，使用四维张量进行辐射场建模是非常有效的方法，这也为三维渲染和图形建模等工作提供了新的思路。

然而正如前面所说，这个动态平衡的问题其实仍然存在许多挑战。例如，TY分解的计算量要大于VM分解，因为其中包含了一个类似的CP分解的向量成分分解。同时，这种辐射场建模方式将所有外观信息耦合，不能够提取场景的光照等信息，因此无法进行重光照、光照分解或是纹理编辑。

未来一个可以探讨的改进方向是，在张量化低秩分解的基础上，迁移先前一些结合图形学先验的工作，例如结合BRDF先验，使用MLP学习或使用球面高斯进行模拟建模，提取辐射场中的可解耦外观信息。

参考文献

[1] Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. Tensorf: Tensorial radiance fields. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXII*, pages 333–350. Springer, 2022. 1, 3

[2] Richard A Harshman et al. Foundations of the parafac procedure: Models and conditions for an" explanatory" multimodal factor analysis. 1970. 2

[3] Andreas Hoecker and Vakhtang Kartvelishvili. Svd approach to data unfolding. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 372(3):469–481, 1996. 2

[4] Tamara G Kolda and Brett W Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009. 1

[5] Shaoxu Li and Ye Pan. Instant neural radiance fields stylization. *arXiv preprint arXiv:2303.16884*, 2023. 1

[6] Kunhao Liu, Fangneng Zhan, Yiwen Chen, Jiahui Zhang, Yingchen Yu, Abdulmotaleb El Saddik, Shijian Lu, and Eric Xing. Stylerf: Zero-shot 3d style transfer of neural radiance fields. *arXiv preprint arXiv:2303.10598*, 2023. 1

[7] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021. 1

[8] Jiaxiang Tang, Xiaokang Chen, Jingbo Wang, and Gang Zeng. Compressible-composable nerf via rank-residual decomposition. *arXiv preprint arXiv:2205.14870*, 2022. 1, 3