

# 计算机视觉与深度学习中的张量方法 笔记+心得p1

张新鹏

October 21, 2023

## 1 摘要

张量（又称多维数组）是一种可以自然地表示多维视觉数据的数据结构。本质上能够有效地捕获结构化的、潜在的语义空间和高阶交互。张量在计算机视觉中早已广泛应用。计算机视觉领域转向深度学习后张量被更加广泛的运用。事实上，现代深度学习架构中的基本成分，如卷积和注意力机制，可以很容易地被认为是张量映射。实际上，张量方法在深度学习中越来越多地找到重要的应用，包括设计存储和计算高效的网络架构，提高对随机噪声和对抗性攻击的鲁棒性，以及帮助对深度网络的理论理解。

这篇文章是对张量在计算机视觉和深度学习中的运用的综述。

### 1.1 索引术语

张量方法，计算机视觉，深度学习

## 2 I.介绍

张量是矩阵的多维泛化，和向量或者矩阵那样靠索引来确定元素，矩阵可以看成二阶张量，矩阵需要靠行列来确定元素，张量也类似，N阶张量需要N个指标来进行索引。

张量的运用广泛。在数据科学和机器学习、信号处理、计算机视觉、统计学等相关领域，张量被用于表示和分析隐藏在多维数据(如图像和视频)中的信息，或捕获和利用向量值变量之间的高阶相似性或依赖关系。一阶张量是一个向量，表示一个样本的特征或属性。二阶张量是一个矩阵，表示两个样本之间的关系或相似性。三阶张量则可以表示三个样本之间的相似性，以此类推，更高阶的张量可以表示更多样本之间的关系。这样的张量通常被称为高阶统计量。在这种背景下，张量方法主要专注于扩展基于矩阵的学习模型。在深度学

习中张量的概念是核心。除了表示数据和统计之外，张量可以被视为多线性映射(即函数)，它是由矩阵表示的线性映射的高阶推广。

这篇文章概述了在表示学习和深度学习背景下的张量和张量方法，特别关注视觉数据分析和计算机视觉应用。

## 2.1 可视化数据中的张量结构

视觉数据是多维数据的一个突出例子，它的张量结构有两种主要形式：

显示/可以测量的张量结构：由物理上有意义的数组成，描述数据的坐标，颜色，光强等。比如常见彩色的图片是三阶张量由三个储层不同颜色（红黄蓝）的二阶张量堆叠成，每个二阶张量储存空间上数据，空间和光谱/颜色模式。四阶张量还被用于表示包括时间模式的视觉数据。

潜在信息张量结构：图像形成依赖于与外观(例如，光照、姿态)甚至图像语义(例如人脸中的性别和年龄)相关的多个潜在变异因素的相互作用。这些潜在的信息映射在张量上，通过张量可以识别出这些信息。如果我们保持除一个因素外的所有因素不变，视觉变化是线性的。例如，在面部图像中，改变光照不会改变被描绘的人的身份（改变光照也能通过图片信息认出这人）。

完成计算机视觉任务需要有物理或语义意义的视觉数据，用的传统机器学习模型处理这些数据的时候会出现一些问题，多维视觉数据样本需要被平展成非常高维的向量，测量的不同模式(例如，空间和时间)之间的自然拓扑结构和依赖关系被丢失。除了结构损失之外，当在训练基于矩阵的机器学习模型中使用高维向量时，在给定精度水平内估计任意函数(或机器学习模型的参数)所需的数据样本数量随着数据维度呈指数级增长。这种现象被称为维数灾难。

## 2.2 计算机视觉表示学习中的张量方法：

张量方法已经成为从多维数据中学习表示的强大工具。张量方法并在保留数据的结构的同时减少维度灾难，保留视觉数据的有用信息（上文提到的潜在信息）。张量分解和张量成分分析方法在计算机视觉应用中产生了广泛的影响，涵盖了从人类感知（如人脸和身体分析）到医学和高光谱成像等领域。

在计算机视觉中，降维、聚类和数据压缩等任务都依赖于对视觉数据进行张量分解来进行表示。通过对数据张量进行非负分解，可以得到与视觉对象的局部部分相对应的表示。将视觉数据分解为形状和运动因子，或者表面法向和反射因子，有助于从运动中恢复结构，并提供光度立体的解决方案。

最近，深度学习模型在各种计算机视觉应用上取得了质的突破，其中包括许多其他机器学习任务。深度神经网络充当了高维非线性函数的可学习近似。人们试图学习一个未知的高维目标函数，该函数将输入图像映射到输出标签。它们成功的关键因素是深度学习模型通过局部统计利用了数据的统计属性，如平稳性(例如，移位不变性)和组合性(例如，图像的层次结构)，这些都存在于视觉数据中。这些属性被卷积架构[13]，[16]利用，这些卷积架构由交替多维卷积层、逐点非线性函数(例如ReLU)、下采样(池化)层构建，同时也包含张量结构

的全连接层。在深度神经网络中使用多维卷积，可以提取跨图像域共享的局部特征。反过来，这大大减少了可学习参数的数量，从而缓解了维数灾难的影响而不牺牲逼近目标函数的能力。

## 2.3 计算机视觉深度学习架构中的张量方法:

尽管深度神经网络的组成结构减轻了维数灾难，但深度学习模型通常是过度参数化的，涉及数量巨大的未知参数(通常为数千万甚至数十亿)。其中包括许多其他机器学习任务[12]。深度神经网络充当了高维非线性函数的可学习近似。尽管深度神经网络的过参数化允许找到良好的局部最小值，如此大量的可训练参数使深度网络容易过度拟合并干扰，阻碍了对其泛化特性(即它们在未见过的数据上的表现如何)的分析。张量分解可以显著减少深度模型的未知参数数量，并进一步缓解维数灾难。量分解可以应用于神经网络层的权重来压缩它们，并且在某些情况下，加速它们。此外，在深度神经网络中使用基于张量的操作，可以保留和利用数据中的拓扑结构，同时产生简约的模型。在卷积神经网络(cnn)中，(不可分离的)多通道卷积核可以使用低秩张量分解分解为模式可分离卷积的总和，从而导致网络压缩和参数数量的减少。深度网络压缩在资源有限的设备中部署深度学习模型方面意义重大，同时减少参数数量作为隐式正则化，可以提高跨任务和域的泛化能力。此外,神经网络构建模块的张量分解已被证明在增强神经网络对抗抗性攻击和各种类型的随机噪声(如捕获数据时自然产生的噪声，如MRI)的鲁棒性方面是有效的。

然后文章举了几篇关于深度学习不同领域的综述论文（在原文第三页第二段）

## 2.4 论文的结构

论文的其余部分组织如下。在第二节中，介绍张量和矩阵代数的基础知识。在第三节中，讨论了现有的计算工具和必要的基础设施的概述，重点是文章的笔记本所基于的TensorLy库。第IV节介绍了用张量进行表示学习。在第五节中，详细介绍了最近在深度学习架构中利用张量方法的研究，而在第六节中讨论了计算机视觉中的相关应用，最后，在第七节中给出了实际挑战的总结，以及对从业者的建议。

# 3 II.关于矩阵和张量的初步介绍

张量代数的符号和操作对该领域的新人来说可能显得复杂。这篇文章这块接受文中使用的符号约定，然后回顾线性和多线性代数的一些基本概念。使读者了解张量方法中用于学习的数学工具。

然后文章指出了描述术语/符号的综述论文（在原文第三页右侧）

### 3.1 A.符号

标量用普通字母表示，即 $x, I, j$ 。一阶张量是向量，记为 $x$ 。二阶张量是矩阵，记为 $X$ 。 $I$ 表示相容维数的单位矩阵。 $X$ 的转置记为 $x_j$ 。 $X$ 的第 $i$ 列记为 $x_i$

实数和整数的集合分别用 $R$ 和 $Z$ 表示。

三阶或更高阶的张量用 $x$ 表示。张量的阶数是表示其元素所需的指标(维度)的数量。每个维度被称为一个模式。一个 $N$ 阶张量有 $N$ 个指标，每个指标寻址一个模式 $X$ 。给定对应于 $X$ 的每个模态的 $N$ 个向量(或矩阵)的集合，第 $N$ 个向量(或矩阵)记为 $u^{(N)}$ (或 $U^{(N)}$ )

纤维是矩阵的行和列概念到张量的一种推广。它们是通过固定除一个以外的所有指标来获得的。用冒号表示一个模态的所有元素，例如， $X$ 的模1光纤表示为 $X_{:,i_2,i_3}$

对于一个三阶张量，通过固定其中一个指标，可以得到切片。

## 4 心得

这篇论文是对计算机视觉和深度学习中的张量方法进行的综述。首先得明确什么是张量，张量是将矩阵进行泛化，矩阵是二阶的，有行列，张量的阶数可以比矩阵高阶，比如存储一张彩色图片用三阶张量。向量可以看成一阶张量，矩阵可以看成二阶张量。张量确定元素也和矩阵类似，矩阵靠行列来确定元素， $N$ 阶张量用 $N$ 个指标来进行索引。

然后是计算机视觉使用张量的必要性。张量被用于表示和分析隐藏在多维数据(如图像和视频)中的信息，或捕获和利用向量值变量之间的高阶相似性或依赖关系。张量能有效地捕获结构化的、潜在的语义和高阶交互。一阶张量是一个向量，表示一个样本的特征或属性。二阶张量是一个矩阵，表示两个样本之间的关系或相似性。三阶张量则可以表示三个样本之间的相似性，更高阶的张量可以表示更多样本之间的关系。视觉数据就是这样的多阶数据，用张量的方式表示和储存。显示/可观测的数据之间的高阶关系，比如储存彩色图片，不同颜色是靠不同数值的红黄蓝颜色叠加呈现的，三阶张量除了表示空间上的关系和还有不同矩阵堆叠组成的不同颜色。如果用传统的矩阵储存，颜色这种信息就会丢失。除了结构，还有图像语义等潜在的信息，比如图片中物品的类型，这种信息会映射到张量上。

传统机器学习模型处理计算机视觉会出现一些问题，多维视觉数据被展开成二维的时候，信息会丢失，并且机器学习模型需要学习的参数会特别夸张，指数级的增加，这样需要提供的样本数据会特别多，并且过拟合和干扰的问题会很严重（只拟合了样本数据，没泛化能力）。

所以深度学习广泛使用张量方法。深度学习模型应用了张量数据中的统计属性，大大减少了学习参数的数量。深度神经网络中使用基于张量的操作，可以保留和利用数据中的拓扑结构，同时产生简约的模型。深度神经网络中使用的张量方法还在阅读中。