

双加权截断核范数张量填充

李峥儿, 尚紫微

辽宁师范大学, 辽宁 大连

收稿日期: 2021年9月13日; 录用日期: 2021年10月6日; 发布日期: 2021年10月15日

摘要

低秩张量填充根据数据的低秩性来恢复所有丢失的元素来更好地描述高维数据结构, 逐渐得到各领域学者的重视。先前的学者们已经提出了张量核范数的几种定义, 但是可能无法正确地近似张量真正的秩, 并且在优化中这种方法没有明确地使用低秩性质, 基于此, 有学者提出了张量截断核范数(T-TNN), 但需要大量的迭代才能收敛。因此, 本文提出了一种新的方法双加权截断核范数张量填充(DW-T-TNN), 为张量每片分配不同的权重, 以加速收敛, 并获得可接受的性能。本文设计了一种简洁的梯度下降方法, 代替了T-TNN第二步的迭代更新方案。在实验过程中, 我们的方法有良好的视觉效果并且所用时间更少。

关键词

张量填充, 截断核范数, 低秩, 双加权

Double Weighted Truncated Nuclear Norm Regularization for Low-Rank Tensor Completion

Zheng'er Li, Ziwei Shang

Liaoning Normal University, Dalian Liaoning

Received: Sep. 13th, 2021; accepted: Oct. 6th, 2021; published: Oct. 15th, 2021

Abstract

To better describe the high dimensional data structure, low-rank tensor completion based on the low rank property has gradually attracted the attention of scholars in various fields whose partial elements are missing. Previous scholars have proposed several definitions of tensor nuclear norm, but it may not be able to correctly approximate the real rank of tensor, and this method does not explicitly use the low rank property in optimization. Based on this, some scholars have proposed

Tensor Truncated Nuclear Norm (T-TNN), but it requires numerous iterations to converge. Therefore, this paper proposes a new method, Double Weighted Truncated Nuclear Norm Regularization for Low-Rank Tensor Completion (DW-T-TNN) which assigns different weights to each piece of tensor to accelerate convergence and obtain acceptable performance. In this paper, an efficient gradient descent strategy is designed to replace the iterative update scheme of the second step of T-TNN. In the experiment of real-world image, our method has good visual effect and takes less time.

Keywords

Tensor Completion, Truncated Nuclear Norm, Low-Rank, Double Weighted

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

近年来, 随着现代网络技术的急速发展, 计算机视觉、图像处理以及推荐系统等领域得到了前所未有的关注与高度, 与此同时获取大量的高维数据也逐渐变得简单起来。但是获取高维数据过程中难免会有部分数据丢失的情况出现。虽然矩阵填充也可恢复出丢失的元素, 但是当待恢复的高维数据具有相对复杂的结构时, 利用矩阵填充可能会造成维度灾难、过拟合以至于最终破坏数据结构。因此作为矩阵填充的推广, 根据数据的低秩性来恢复出所有丢失的元素来更好地描述高维数据结构的低秩张量填充, 逐渐得到各领域学者的重视。

通过将彩色图像或者视频看做三维张量, 先前的学者们已经提出了张量核范数的几种定义。但是这些定义可能无法正确地近似于张量真正的秩而且在优化中没有明确地使用低秩性质, 这就造成了局限性。最近提出的截断核范数(TNN)被证明可以代替传统的核范数来更好的近似秩。基于此, [1]提出了张量截断核范数(T-TNN)将截断核范数从矩阵情形推广到张量情形。但是这种方法需要多次迭代才能收敛。本文在熟悉截断核范数和张量相关知识的基础上, 认真分析国内外学者在张量填充领域的研究现状, 提出了双加权截断核范数张量填充(DW-T-TNN)方法, 将不同的权重分别分配给张量每个前片的行和列, 以加快收敛速度并获得可接受的性能。同时提出了一种简单的梯度下降方法, 代替了 T-TNN 中第二步的迭代更新方法。在真实数据上进行的有效实验证明, DW-T-TNN 具有良好的性能, 在完成的速度和视觉效果上都具有优势。

2. 预备知识

定义 1 [2]张量核范数: 张量 $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ 的核范数记为 $\|\mathcal{A}\|_*$, 表示 $\bar{\mathcal{A}}$ 所有正面切片核范数的平均值,

$$\text{即: } \|\mathcal{A}\|_* = \frac{1}{\sqrt{n_3}} \|\bar{\mathcal{A}}\|_F$$

定义 2 [2]张量 \mathcal{F} 范数: 张量 $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $\bar{\mathcal{A}}$ 表示张量进行傅里叶变换之后的块循环矩阵, 张量 \mathcal{F}

$$\text{范数即: } \|\mathcal{A}\|_{\mathcal{F}} = \frac{1}{\sqrt{n_3}} \|\bar{\mathcal{A}}\|_{\mathcal{F}}$$

定义 3 [1]张量截断核范数:

张量 $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ 的截断核范数为 $\|\mathcal{X}\|_*$, 被定义为

$$\|\mathcal{X}\|_r = \|\mathcal{X}\|_* - \max_{\substack{\mathcal{A}\mathcal{A}^T=I \\ \mathcal{B}\mathcal{B}^T=I}} tr(\mathcal{A} * \mathcal{X} * \mathcal{B})$$

其中 $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $\mathcal{B} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ 是由 TSVD 生成的,

定义 4 [3] 张量奇异值分解:

设 $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ 的 TSVD 是 $\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T$, 其中 $\mathcal{U} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$, $\mathcal{V} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ 是正交的, $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ 是 f -对角张量。

定义 5 [2] $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $\mathcal{B} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, 则有 $\langle \mathcal{A}, \mathcal{B} \rangle = \frac{1}{n_3} \langle \bar{\mathcal{A}}, \bar{\mathcal{B}} \rangle$ 。

3. 双加权截断核范数张量填充

3.1. 模型的建立

由[1]易证明, 张量截断核范数等价于

$$\begin{aligned} \min \quad & \max_{\mathcal{A}\mathcal{A}^T=\mathcal{B}\mathcal{B}^T=I} tr(\mathcal{A} * \mathcal{X} * \mathcal{B}^T) - \max_{\mathcal{C}\mathcal{C}^T=\mathcal{D}\mathcal{D}^T=I} tr(\mathcal{C} * \mathcal{X} * \mathcal{D}^T) \\ \text{st} \quad & \mathcal{X}_\Omega = \mathcal{M}_\Omega \end{aligned} \tag{1}$$

其中 $\mathcal{A} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$, $\mathcal{B} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $\mathcal{C} \in \mathbb{R}^{n_1 \times n_3 \times n_3}$, $\mathcal{D} \in \mathbb{R}^{n_2 \times n_3 \times n_3}$

显而易见, 直接求解上式并不容易, 所以我们把这个优化分为两步进行, 指定 $\mathcal{X}_1 = \mathcal{M}_\Omega$ 作为初始化. 在第 k 次迭代中, 第一步我们固定 \mathcal{X}_k 来更新 \mathcal{A}_k 、 \mathcal{B}_k 、 \mathcal{C}_k 和 \mathcal{D}_k 。

$$\mathcal{A}_k = \mathcal{U}^T \tag{2}$$

$$\mathcal{B}_k = \mathcal{V}(:, 1:n_1, :)^T \tag{3}$$

$$\mathcal{C}_k = \mathcal{U}(:, 1:r, :)^T \tag{4}$$

$$\mathcal{D}_k = \mathcal{V}(:, 1:r, :)^T \tag{5}$$

接下来在第二步中, 我们通过保持其他变量不变, 通过下列算式更新优化 \mathcal{X}_{k+1} 。

$$\begin{aligned} \min \quad & tr(\mathcal{A}_k * \mathcal{X} * \mathcal{B}_k^T) - tr(\mathcal{C}_k * \mathcal{X} * \mathcal{D}_k^T) \\ \text{st} \quad & \mathcal{X}_\Omega = \mathcal{M}_\Omega \end{aligned} \tag{6}$$

然后引入辅助变量, 在这里我们采用交替乘子法进行求解,

$$\begin{aligned} \min \quad & tr(\mathcal{A}_k * \mathcal{X} * \mathcal{B}_k^T) - tr(\mathcal{C}_k * \mathcal{X} * \mathcal{D}_k^T) \\ \text{st} \quad & \mathcal{X}_\Omega = \mathcal{W}_\Omega \quad \mathcal{W}_\Omega = \mathcal{M}_\Omega \end{aligned} \tag{7}$$

随后转化成为一个无约束的增广拉格朗日函数进行求解,

$$\mathcal{L}(\mathcal{X}, \mathcal{W}, \mathcal{Y}) = tr(\mathcal{A}_k * \mathcal{X} * \mathcal{B}_k^T) - tr(\mathcal{C}_k * \mathcal{X} * \mathcal{D}_k^T) + \frac{\mu}{2} \|\mathcal{X} - \mathcal{W}\|_F^2 + tr(\mathcal{Y}^T (\mathcal{X} - \mathcal{W})) \tag{8}$$

为了更准确, 有更高优先级的进行填充, 我们选择在这里加入权重

$$\mathcal{L}(\mathcal{X}, \mathcal{W}, \mathcal{Y}) = tr(\mathcal{A}_k * \mathcal{X} * \mathcal{B}_k^T) - tr(\mathcal{C}_k * \mathcal{X} * \mathcal{D}_k^T) + \frac{\mu}{2} \|\mathcal{P}(\mathcal{X} - \mathcal{W})\mathcal{Q}\|_F^2 + tr(\mathcal{Y}^T \mathcal{P}(\mathcal{X} - \mathcal{W})\mathcal{Q}) \tag{9}$$

其中 \mathcal{P} , \mathcal{Q} 是权重对角张量。

3.2. 模型的优化

假设 \mathcal{X}_k , \mathcal{Y}_k 表示第二步中第 K 次迭代的结果, 我们通过保持其他变量不变, 来更新 \mathcal{W}_{k+1} 如下

$$\begin{aligned}\mathcal{W}_{k+1} &= \arg \min \mathcal{L}(\mathcal{X}, \mathcal{W}, \mathcal{Y}) \\ &= \arg \min_{\mathcal{W}} -tr(\mathcal{C}_k * \mathcal{W} * \mathcal{D}_k^T) + \frac{\mu}{2} \|\mathcal{P}(\mathcal{X} - \mathcal{W})\mathcal{Q}\|_F^2 + tr(\mathcal{Y}^T \mathcal{P}(\mathcal{X} - \mathcal{W})\mathcal{Q}) \\ &= \mathcal{X}_k + \frac{1}{\mu} (\mathcal{P}^{-2} \mathcal{C}_k^T \mathcal{D}_k \mathcal{Q}^{-2} + \mathcal{P}^{-1} \mathcal{Y}_k \mathcal{Q}^{-1})\end{aligned}\quad (10)$$

由约束条件可知, 观测元素的值保持不变, 在此获得

$$\mathcal{W}_{k+1} = (\mathcal{W}_{k+1})_{\Omega^c} + \mathcal{M}_{\Omega} \quad (11)$$

\mathcal{X} 的更新如下

$$\begin{aligned}\mathcal{X}_{k+1} &= \arg \min_{\mathcal{X}} \mathcal{L}(\mathcal{X}, \mathcal{W}_{k+1}, \mathcal{Y}_k) \\ &= \arg \min_{\mathcal{X}} tr(\mathcal{A}_k * \mathcal{X} * \mathcal{B}_k^T) + \frac{\mu_k}{2} \left\| \mathcal{P}(\mathcal{X} - \mathcal{W}_{k+1})\mathcal{Q} + \frac{\mathcal{Y}_k}{\mu_k} \right\|_F^2 \\ &= \mathcal{W}_{k+1} - \frac{1}{\mu_k} (\mathcal{P}^{-2} \mathcal{A}_k^T \mathcal{B}_k \mathcal{Q}^{-2} + \mathcal{P}^{-1} \mathcal{Y}_k \mathcal{Q}^{-1})\end{aligned}\quad (12)$$

\mathcal{Y} 的更新如下

$$\mathcal{Y}_{k+1} = \mathcal{Y}_k + \beta_k \mathcal{P}(\mathcal{X}_{k+1} - \mathcal{W}_{k+1})\mathcal{Q} \quad (13)$$

其中 β_k 是单调递增序列。

这三项更新步骤在实际应用中需要大量的迭代才能收敛, 计算代价很大。在此我们根据其中的内在关联, 推出一种简洁的梯度下降方式, 即在不需大量迭代步骤的情况下, 就可以有效地更新 \mathcal{X} , 如下定理 1 所示。

定理 1 [4]: 如果 $0 < \mu_t < \mu_{t+1}$, $\forall t = 1, 2, \dots, N$, $\frac{1}{\alpha_k} = \sum_{t=1}^{N-1} \frac{1}{\mu_t}$, 由(10)~(13), \mathcal{X}_k 的更新可以由下的一步迭代更新完成

$$\mathcal{X} = \mathcal{X}_k - \frac{1}{\alpha_k} \mathcal{P}^{-2} (\mathcal{A}_k^T \mathcal{B}_k - \mathcal{C}_k^T \mathcal{D}_k) \mathcal{Q}^{-2} \quad (14)$$

$$\mathcal{X}_{k+1} = (\mathcal{X})_{\Omega^c} + \mathcal{M}_{\Omega} \quad (15)$$

其中 $\frac{1}{\alpha_k}$ 代表补步长, 证明在此处不详述。

此外, 为了简便表示, 下面我们定义两个两个对角张量 $\mathcal{P}^{-2} = \tilde{\mathcal{P}}$, $\mathcal{Q}^{-2} = \tilde{\mathcal{Q}}$,

$$\tilde{\mathcal{P}} \text{ 的每一个前片 } \tilde{\mathcal{P}}^{(i)} = \text{diag}(p_1, \dots, p_{n_1}), i = 1, \dots, n_3 \quad (16)$$

$$\tilde{\mathcal{Q}} \text{ 的每一个前片 } \tilde{\mathcal{Q}}^{(i)} = \text{diag}(q_1, \dots, q_{n_2}), i = 1, \dots, n_3 \quad (17)$$

$\{p_i\}_{i=1}^{n_1} \geq 0$, $\{q_i\}_{i=1}^{n_2} \geq 0$ 分别根据每行和每列中观察到的元素的数量精确确定, 如下

$$\begin{aligned}p_i &= \exp\left(-\theta_1 \left(\frac{N_i^r}{n} - 1\right)\right) - 1, i = 1, \dots, n_1 \\ q_i &= \exp\left(-\theta_2 \left(\frac{N_j^c}{m} - 1\right)\right) - 1, j = 1, \dots, n_2\end{aligned}\quad (18)$$

其中 θ_1 和 θ_2 是缩放权重, 这意味着具有更多的观察元素的行(列)具有更小的权重值, 并且 $p_i = 0 (q_i = 0)$ 意味着此时没有元素丢失。

基于上述, 我们引入梯度下降法。

首先我们介绍一个引理[5]: 假设对于任意两个张量 $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $\mathcal{B} \in \mathbb{R}^{n_2 \times n_4 \times n_3}$, 令 $\mathcal{F} = \mathcal{A} * \mathcal{B}$, 则有下面性质成立

$$1) \|\mathcal{A}\|_F^2 = \frac{1}{n_3} \|\bar{\mathcal{A}}\|_F^2。$$

$$2) \mathcal{F} = \mathcal{A} * \mathcal{B} \text{ 和 } \bar{\mathcal{F}} = \bar{\mathcal{A}} * \bar{\mathcal{B}} \text{ 相互相等。}$$

$$\bar{\mathcal{A}}_k^{(i)} = \mathcal{U}_k^{(i)\top} = [u_1, \dots, u_r, \dots, u_{n_1}]^\top = [C_K^\top, u_{r+1}, \dots, u_{n_1}]^\top = [C_K^\top, H_k^\top]^\top, i = 1, 2, \dots, n_3 \quad (19)$$

$$\bar{\mathcal{B}}_k^{(i)} = [v_1, \dots, v_r, \dots, v_{n_4}]^\top = [D_K^\top, v_{r+1}, \dots, v_{n_4}]^\top = [D_K^\top, L_k^\top]^\top, i = 1, 2, \dots, n_3 \quad (20)$$

由引理 1, $\mathcal{A}^\top * \mathcal{B} = \bar{\mathcal{A}}^\top * \bar{\mathcal{B}}$, 容易证得

$$\mathcal{A}_K^\top \mathcal{B}_K - C_K^\top D_K = [C_K^\top, H_K^\top] \begin{bmatrix} \bar{D}_K \\ \bar{L}_K \end{bmatrix} = \bar{H}_K^\top \bar{L}_K = \mathcal{H}_K^\top \mathcal{L}_K, i = 1, 2, \dots, n_3 \quad (21)$$

因此, 通过定理 1, (16)~(18), 和(21), \mathcal{X}_{k+1} 可以不用大量迭代, 即通过一个简单的梯度下降方法就可以有效地计算, 即:

$$\mathcal{X}^* = \mathcal{X}_k - \frac{1}{\alpha_k} \mathcal{P} \mathcal{H}_K^\top \mathcal{L}_K \mathcal{Q} \quad (22)$$

$$\mathcal{X}_{k+1} = (\mathcal{X}^*)_{\Omega^c} + \mathcal{M}_\Omega \quad (23)$$

在上面两个式子的基础上, 我们发现它们等价于函数 $tr(\mathcal{H}_k \tilde{\mathcal{P}} \mathcal{X} \tilde{\mathcal{Q}} \mathcal{L}_k^\top)$ 的梯度下降法搜索的解。

由于上述一步梯度下降法的优点, 代替了在第二步中迭代计算 \mathcal{X}_k , 因此, 我们可以将(1)转换为以下公式:

$$\begin{aligned} \min \quad & tr(\mathcal{H} \tilde{\mathcal{P}} \mathcal{X} \tilde{\mathcal{Q}} \mathcal{L}^\top) \\ \text{s.t.} \quad & \mathcal{X}_\Omega = \mathcal{M}_\Omega \end{aligned} \quad (24)$$

注意到当 $r = n_1$ 意味着 $\mathcal{H} = \mathcal{O}$, $\mathcal{L} = \mathcal{O}$ 这就意味着梯度消失, 所以在此我们选择 $r \ll n_1$ 这样就会避免这种情况的发生, 这也符合视觉数据(例如真实图像)普遍具有低阶结构的事实。换句话说, 实际的秩 r 要比图像的尺寸小得多。

初始值 $\mathcal{X}_1 = \mathcal{M}_\Omega$, (24)的优化是通过一个梯度下降法解决如下

$$\mathcal{X}_{k+1} = \mathcal{X}_k - \frac{1}{\alpha_k} \tilde{\mathcal{P}} \mathcal{H}_K^\top \mathcal{L}_K \tilde{\mathcal{Q}} \quad (25)$$

$$\mathcal{X}_{k+1} = (\mathcal{X}_{k+1})_{\Omega^c} + \mathcal{M}_\Omega$$

其中 $\frac{1}{\alpha_k}$ 是递减的步长, 满足

$$\alpha_{k+1} = \rho \alpha_k \quad (26)$$

其中 $\rho > 1$ 是一个常数。

4. 实验部分

实验部分将本文模型(DW-T-TNN)与 TNN (截断核范数)、T-TNN (张量截断核范数)、DW-TNN (加权截断核范数)这 3 个模型来作比较, 在此选取图片大小为 $300 \times 300 \times 3$, 椒盐噪声为 50% 的彩色图片。我们通过对彩色图片的恢复情况的视觉效果, 展示本文模型的恢复效果。下面列出在单张彩色图片上这四种模型恢复上的视觉效果, 如图 1 所示。并且对比了四种模型恢复图片所用的时间, 如表 1 所示。

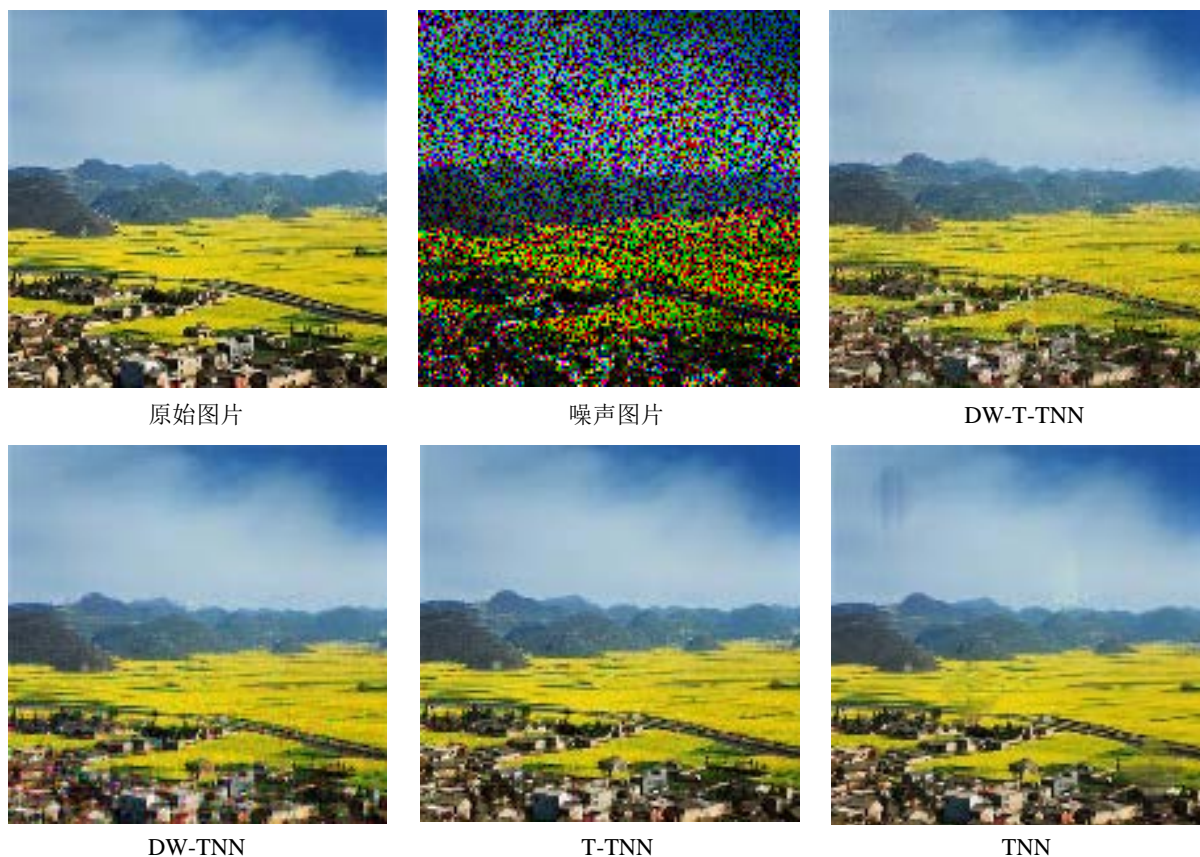


Figure 1. Comparison of visual effects of four different denoising algorithms (a) Original image; (b) Noisy image; (c) DW-T-TNN; (d) DW-TNN; (e) T-TNN; (f) TNN

图 1. 四种不同算法去噪视觉效果对比 (a) 原始图片; (b) 噪声图片; (c) DW-T-TNN; (d) DW-TNN; (e) T-TNN; (f) TNN

Table 1. Comparison of denoising time of four different algorithms

表 1. 四种不同算法去噪时间对比

算法	时间(s)
DW-T-TNN	7.8
DW-TNN	33.6
TNN	377
T-TNN	1138.7

5. 结论

在本文中, 我们提出了双加权截断核范数张量填充模型, 以及一种简洁的梯度下降方法。实验表明

在相同噪声情况下, 视觉效果和时间都优于其他算法, 所以该方法对图像恢复是有优化效果的。

参考文献

- [1] Xue, S.K., Qiu, W.Y., Liu, F. and Jin, X. (2018) Low-Rank Tensor Completion by Truncated Nuclear Norm Regularization. 2018 *24th International Conference on Pattern Recognition (ICPR)*, Beijing, 20-24 August 2018, 2600-2605. <https://doi.org/10.1109/ICPR.2018.8546008>
- [2] Lu, C.Y., Feng, J.S., Chen, Y.D., Liu, W., Lin, Z.C. and Yan, S.C. (2016) Tensor Robust Principal Component Analysis: Exact Recovery of Corrupted Low-Rank Tensors via Convex Optimization. 2016 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, 27-30 June 2016, 5249-5257. <https://doi.org/10.1109/CVPR.2016.567>
- [3] Jolliffe, I. (2002) *Principal Component Analysis*. Wiley Online Library, New York.
- [4] Xue, S.K., Qiu, W.Y., Liu, F. and Jin, X.Y. (2019). Double Weighted Truncated Nuclear Norm Regularization for Low-Rank Matrix Completion. arXiv:1901.01711v1 [cs.CV].
- [5] Kilmer, M.E. and Martin, C.D. (2011) Factorization Strategies for Third-Order Tensors. *Linear Algebra and Its Applications*, **435**, 641-658. <https://doi.org/10.1016/j.laa.2010.09.020>