ACCEPTED VERSION

Lei Zhang, Wei Wei, Qinfeng Shi, Chunhua Shen, Anton van den Hengel, and Yanning Zhang

Accurate tensor completion via adaptive low-rank representation

IEEE Transactions on Neural Networks and Learning Systems, 2020; 31(1):4170-4184

© 2019 IEEE

Published version at: http://dx.doi.org/10.1109/tnnls.2019.2952427

PERMISSIONS

https://www.ieee.org/publications/rights/author-posting-policy.html

Author Posting of IEEE Copyrighted Papers Online

The IEEE Publication Services & Products Board (PSPB) last revised its Operations Manual Section 8.1.9 on Electronic Information Dissemination (known familiarly as "author posting policy") on 7 December 2012.

PSPB accepted the recommendations of an ad hoc committee, which reviewed the policy that had previously been revised in November 2010. The highlights of the current policy are as follows:

- The policy reaffirms the principle that authors are free to post their own version of their IEEE periodical or conference articles on their personal Web sites, those of their employers, or their funding agencies for the purpose of meeting public availability requirements prescribed by their funding agencies. Authors may post their version of an article as accepted for publication in an IEEE periodical or conference proceedings. Posting of the final PDF, as published by IEEE *Xplore*®, continues to be prohibited, except for open-access journal articles supported by payment of an article processing charge (APC), whose authors may freely post the final version.
- The policy provides that IEEE periodicals will make available to each author a preprint version of that person's article that includes the Digital Object Identifier, IEEE's copyright notice, and a notice showing the article has been accepted for publication.
- The policy states that authors are allowed to post versions of their articles on approved third-party servers that are operated by not-for-profit organizations. Because IEEE policy provides that authors are free to follow public access mandates of government funding agencies, IEEE authors may follow requirements to deposit their accepted manuscripts in those government repositories.

IEEE distributes accepted versions of journal articles for author posting through the Author Gateway, now used by all journals produced by IEEE Publishing Operations. (Some journals use services from external vendors, and these journals are encouraged to adopt similar services for the convenience of authors.) Authors' versions distributed through the Author Gateway include a live link to articles in IEEE *Xplore*. Most conferences do not use the Author Gateway; authors of conference articles should feel free to post their own version of their articles as accepted for publication by an IEEE conference, with the addition of a copyright notice and a Digital Object Identifier to the version of record in IEEE *Xplore*.

17 May 2021

http://hdl.handle.net/2440/130272

Accurate Tensor Completion via Adaptive Low-Rank Representation

Lei Zhang, *Member*, *IEEE*, Wei Wei, *Senior Member*, *IEEE*, Qinfeng Shi, Chunhua Shen, Anton van den Hengel and Yanning Zhang, *Senior Member*, *IEEE*

***Abstract*—Low-rank representation-based approaches that assume low-rank tensors and exploit their low-rank structure with appropriate prior models, have underpinned much of the recent progress in tensor completion. However, real tensor data only approximately comply with the low-rank requirement in most cases, viz., the tensor consists of low-rank (e.g., principle part) as well as non-low-rank (e.g., details) structures, which limits the completion accuracy of these approaches.**

To address this problem, we propose an adaptive low-rank representation model for tensor completion that represents low-rank and non-low-rank structures of a latent tensor separately in a Bayesian framework. Specifically, we reformulate the CANDECOMP/PARAFAC (CP) tensor rank, and develop a sparsity-induced prior for the low-rank structure that can be used to determine tensor rank automatically. Then, the non-low-rank structure is modeled using a mixture of Gaussians prior that is shown to be sufficiently flexible and powerful to inform the completion process for a variety of real tensor data. With these two priors, we develop a Bayesian minimum mean squared error estimate framework for inference. The developed framework can capture the important distinctions between low-rank and non-low-rank structures, thereby enabling more accurate model, and ultimately, completion. For various applications, compared with the state-of-the-art methods, the proposed model yields more accurate completion results.

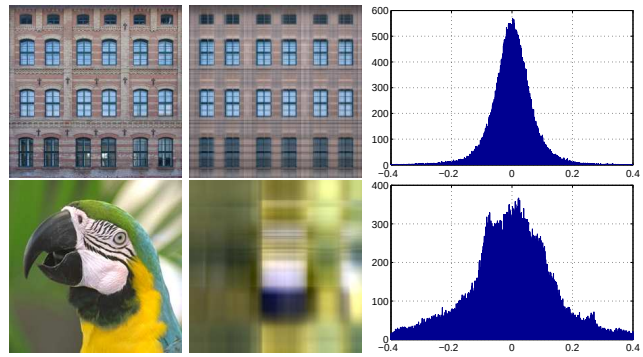
***Index Terms*—Adaptive low-rank representation, automatic tensor rank determination, tensor completion.**

Manuscript received November 28, 2018; revised May 13, 2019; accepted November 02, 2019. Date of publication XXX, 2019; date of current version XXX, 2019. This work was supported in part by the National Natural Science Foundation of China (No. 61671385, 61571354). Lei Zhang and Wei Wei contributed equally to this work. *(Corresponding author: Wei Wei.)*

L. Zhang is with the School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China, with the Shaanxi Provincial Key Laboratory of Speech and Image Information Processing, Northwestern Polytechnical University, Xi'an 710072, China, and also with the Inception Institute of Artificial Intelligence (IIAI), Abu Dhabi 51133, United Arab Emirates (e-mail: zhanglei211@mail.nwpu.edu.cn).

W. Wei and Y. Zhang are with the School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China, with the National Engineering Laboratory for Integrated Aero-Space-Ground-Ocean Big Data Application Technology, Northwestern Polytechnical University, Xi'an 710072, China, and also with the Shaanxi Provincial Key Laboratory of Speech and Image Information Processing, Northwestern Polytechnical University, Xi'an 710072, China (e-mail: weiweinwpu@nwpu.edu.cn; ynzhang@nwpu.edu.cn).

Q. Shi, C. Shen and A. van den Hengel are with the School of Computer Science, The University of Adelaide, Adelaide, SA 5005, Australia, and also with Australian Institute for Machine Learning, Adelaide, SA 5005, Australia (e-mail: javen.shi@adelaide.edu.au, chunhua.shen@adelaide.edu.au; anton.vandenhengel@adelaide.edu.au).



(a) (b) (c)

Fig. 1. Low rank and non-low-rank structures in two color images (top: 'facade', bottom: 'parrot'). (a) Images. (b) Low rank structure obtained by employing CANDECOMP/PARAFAC (CP) factorization [23] on each image with rank = 3. (c) Histograms of entries in the residual non-low-rank structure. For the highly structured 'facade', entries in the non-low-rank structure fit an approximately sparse distribution, while that in 'parrot' is more complex (e.g., heavy-tailed and multimodal).

I. Introduction

Representation of multidimensional data has become increasingly important in various computer vision tasks, such as visual data restoration [1]–[3], discriminative feature learning [4]–[9], compression [10]–[13]. Tensors provide an effective way to represent multidimensional data without loss of structural characteristics, e.g., a video can be viewed as a 4-mode tensor with dimensionality of *height* $\times$ *width* $\times$ *channel* $\times$ *time*. However, a wide range of practical applications (such as social networks [14], recommender systems [15], image processing [16], [17] and face synthesis [18]–[20].) often produce incomplete tensor data where partial entries are missing, e.g., incomplete social relations, unknown user-item correlation or corrupted videos. Missing entries often cause the performance of related applications to decrease significantly, particularly when the missing ratio is high. Thus, significant attention has been paid [1], [21], [22] to estimating missing entries in an incomplete tensor by exploiting its intrinsic structural relations [1]. This is often referred as tensor completion.

A promising way to address tensor completion is to adopt a low-rank representation model [24] that assumes the latent tensor to be of low rank, and recovers missing entries by exploiting the low-rank structure. Specifically, for a latent K-mode tensor $\mathcal{L} \in \mathbb{R}^{n_1 \times \cdots \times n_K}$ ($n_1, \cdots, n_K$ are dimensions in each mode), $y_{\boldsymbol{i}}$ represents the observation of the entry in $\mathcal{L}$ at position $\boldsymbol{i} = (i_1, \cdots, i_K)$. Let $\mathcal{Y}_\Omega = \{y_{\boldsymbol{i}}\}_{\boldsymbol{i} \in \Omega}$ be the set of all such observations, where Ω contains the indices

of all observations. Therefore, the observation model can be formulated as follows:

$$\mathcal{Y}_\Omega = \mathcal{L}_\Omega + \mathcal{M}_\Omega \tag{1}$$

where $\mathcal{M}_\Omega$ is the noise corruption in an observation. In the low-rank representation model, $\mathcal{L}$ can be recovered via the maximum a posteriori (MAP) estimate as follows:

$$\hat{\mathcal{L}} = \arg\max_{\mathcal{L}} p\left(\mathcal{L}|\mathcal{Y}_\Omega\right) \propto p\left(\mathcal{Y}_\Omega|\mathcal{L}\right) p\left(\mathcal{L}|\Theta\right) \tag{2}$$

where $p\left(\mathcal{Y}_\Omega|\mathcal{L}\right)$ denotes the likelihood induced by Eq. (1) and $p(\mathcal{L}|\Theta)$ represents a Θ-parametrized low-rank prior. Various low-rank priors have been proposed [1], [21], [25]–[27]. A brief review can be found in Section II.

However, in practice, real data only approximately comply with the low-rank requirement in most cases, because low-rank structure only covers the major information in the data and fails to represent the residual details information, as shown in Fig. 1. For simplicity, we refer to this problem as *applicability to real data*. To distinguish the residual details information from the low-rank structure, we refer to it as *non-low-rank structure*. Note that these two types of structure are not distinguished by their rank, and we employ the non-low-rank structure to capture the remaining (or residual) information in the real data except the low-rank structure. Inspired by this, the latent tensor $\mathcal{L}$ can be factorized as follows:

$$\mathcal{L} = \mathcal{X} + \mathcal{E}. \tag{3}$$

where $\mathcal{X}$ and $\mathcal{E}$ represents the low-rank and non-low-rank structures, respectively[1]. Most previous methods assume $\mathcal{L}$ to be of low rank in its entirety [1], [21], [24], [25], [27], [28]. They thus implicitly assume $\mathcal{E}$ to be zero [1], [21], [25] or explicitly model a sparse $\mathcal{E}$ [24], [27]–[30]. This assumption is explained as resulting from highly-structured data (e.g., 'facade' in Figure 1 with extensive repeated textures); however, this assumption neglects the fact that few real data exhibit the structural regularity required to support such a model, i.e., most real data are beyond the low-rank assumption and shows a more complex $\mathcal{E}$, e.g., 'parrot' in Figure 1. ***Accurate tensor completion requires explicit modeling of all aspects of the latent $\mathcal{L}$, i.e., those that do, or do not conform to the low rank prior***. Therefore, modeling the complex non-low-rank structure $\mathcal{E}$ is as import to the latent tensor representation as the low-rank structure $\mathcal{X}$. However, few efforts have invested in leveraging this for tensor completion.

In addition, a previous study [27] indicates that ***accurate low-rank representation necessitates the tensor rank estimate***. However, it is difficult for most previous low-rank representation models to determine the tensor rank. We refer to this problem as *tensor rank determination*. In most previous studies [1], [21], [25], tensor rank is determined by unfolding the tensor into a collection of matrices and then minimizing matrix rank norms or setting them manually. However, matrix rank norms cannot capture the multidimensional structure of the latent tensor; thus misleading the rank determination. Moreover, the resulting erroneous rank estimate can cause over-fitting in tensor completion [27].

To jointly address these two problems above with one stone, we present an adaptive low-rank representation model for tensor completion that represents the low-rank and non-low-rank structures of a latent tensor separately in a Bayesian framework. Firstly, we reformulate the CANDECOMP/PARAFAC (CP) tensor rank in a new form, wherein the low-rank structure $\mathcal{X}$ is depicted by a sparsity-induced low-rank prior model. By exploiting the sparsity in the factorization weights probabilistically, the proposed model can determine the tensor rank automatically. Then, we model the non-low-rank structure $\mathcal{E}$ with a mixture of Gaussians (MOG). The powerful ability of MOG to fit a wide range of $\mathcal{E}$s (e.g., zero, sparse or mixed) enables the proposed model to adapt to a variety of real tensor data that is often beyond the low-rank assumption with complex non-low-rank structures. Both of these two advantages contribute to the robustness of modeling the latent tensor. To harness them in a principled way, we adopt the Bayesian minimum mean squared error estimate (MMSE) framework for inference over the proposed model. In contrast to most previous studies that only produce a point estimation on each missing entry with the MAP estimate [9], [17], we infer the posterior mean of missing entries as well as their uncertainty using Gibbs samplers. Experimental results on synthetic tensor data sufficiently demonstrate the capacity of the proposed model in tensor rank determination as well as fitting various non-low-rank structures, the effectiveness of each ingredient, convergence and recovery performance. In addition, the proposed model is evaluated in image inpainting, video completion and facial image synthesis applications to validate its superiority over other state-of-the-art tensor completion methods in terms of recovery accuracy.

The primary contribution of this study can be summarized as follows. 1) Differing from previous methods that only consider the low-rank structure, we propose a general tensor completion model that recovers both the low-rank and non-low-rank structures in tensor data. 2) The proposed model can automatically determine the unknown tensor rank, even with high missing ratios. 3) The proposed model can data-adaptively fit the complex non-low-rank structure in the latent tensor. 4) Spatial coherence is considered for visual tensor data to further improve recovery accuracy. 5) We present state-of-the-art results for various tensor completion applications.

II. Related work

In this section, we briefly review previous low-rank representation based tensor completion methods in terms of their applicability to real data and tensor rank determination.

Applicability to real data. In most cases, real tensor data $\mathcal{L}$ only ever approximately comply with the low-rank assumption and contains both the low-rank and the non-low-rank structures, i.e., $\mathcal{X}$ and $\mathcal{E}$. Nevertheless, most previous studies adopt the low-rank tensor representation model [1], [21], [25] and totally neglect the non-low-rank $\mathcal{E}$ (i.e., $\mathcal{E} = 0$). They mainly focus on exploiting effective low-rank models. For example, Liu et al. [31] propose a new core tensor trace-

[1]Note that $\mathcal{E}$ is not the observation noise that is captured by $\mathcal{M}$ in Equation (1).

norm minimization (CTNM) method to speed up the conventional low-rank tensor completion scheme. Yang et al. [32] leverage the regularized redescending M-estimators to improve the robustness of the low-rank tensor completion model to the outliers. Recently, Dian et al. [3] utilize the low tensor-train rank model for hyperspectral imagery super-resolution. Differing from these apporaches, some other literatures in [24], [27], [28] depict $\mathcal{E}$ by a sparse model. However, in real tensor data $\mathcal{E}$ can be complicated (e.g., heavy-tailed or multimodal). Thus a flexible model is required to describe $\mathcal{E}$. MOG has demonstrated powerful ability to represent complex distributed data in various applications, e.g., denoising [33], [34], deblurring [35] and image representation [36], etc. Inspired by this, we leverage MOG to fit the complex non-low-rank structure $\mathcal{E}$ as part of the latent tensor $\mathcal{L}$ for a general tensor completion model. Although MOG has been utilized in several low-rank models [33], [34], [37], they are clearly different from this work. Specifically, in [33], [34], [37], the latent tensor $\mathcal{L}$ is assumed to be entirely of low rank, and MOG is utilized to model the mixed noise corruption (e.g., $\mathcal{M}$ in Eq. (1)), which is expected to be eliminated from the latent tensor $\mathcal{L}$. In contrast, we assume $\mathcal{L}$ is beyond the low-rank assumption and leverage MOG to fit the complex non-low-rank structure $\mathcal{E}$ belonging to $\mathcal{L}$. In addition, the MOG is inferred from a fully-observed matrix for robust principle analysis in [33], [37], whereas we infer it from a handful of observed entries for tensor completion. In [34], EM algorithm is utilized to determine a point estimation for the MOG parameters. In contrast, we infer the posterior mean of these parameters with Gibbs samplers in the MMSE framework.

Tensor rank determination. According to the rank determination schemes, existing low-rank tensor representation models can be roughly divided into two categories. *1) Completion models:* These models minimize the tensor rank by unfolding a tensor into a collection of matrices and then solving a convex optimization on those matrices with matrix rank norms. For example, Liu et al. [1] propose minimizing the trace norm of a tensor, which is defined as the summation of the nuclear norm on the unfolding matrix along each mode of the tensor. Zhao et al. [26] define a tensor rank norm as the product of rank norms on all unfolding matrices of the tensor. Xu et al. [38] attempt to factorize each unfolding matrix optimally. Hu et al. [39] formulate the low-rank model based on the twist tensor nuclear norm for video completion. Zhang et al. [40] employ a family of nonconvex functions onto the singular values of the square deal matrix of the tensor to approximate the rank of the tensor. These models intrinsically exploit the low-rank structure in the unfolding matrices. However, this cannot fully represent the multidimensional structure of tensors [22]. *2) Factorization models:* They decompose the latent tensor into multiple factors with a fixed rank which is often set manually, and then infer those factors instead. Previous studies [21], [41] propose various factor priors to regularize the Tucker factorization of tensors. Another study [34] proposes a weighted CP factorization scheme. However, it is difficult to set correct tensor rank by chance. Moreover, incorrect tensor rank can induce over-fitting [22]. In this study, we reformulate the CP rank, and the tensor rank can be determined automatically by exploiting the sparsity in factorization weights. Although a similar idea has also been adopted in [22], [27], [42], the proposed model is more general, flexible and powerful. These previous studies essentially differ from the proposed model in three aspects. (i) In [22], [27], CP factorization weights are absorbed into the factor matrices; thus the tensor rank depends on the row sparsity of those matrices. We model weights and factor matrices separately to analyze the effect of various factors on the tensor more flexibly. (ii) All factor matrices follow the same distribution to give a consistent rank in [22], [27], whereas we adopt various distributions to model each factor matrix separately (Section III-A4). (iii) The sparsity is depicted by a hierarchical student-t distribution in [22], [27], while we employ a more powerful reweighed Laplace prior (Remark 1), which determines the tensor rank more accurately (Section V-A).

To the best of our knowledge, this study is the first attempt to jointly address automatic tensor rank determination and modeling the complex non-low-rank structure in tensor completion.

III. Proposed tensor completion model

Given the observation model as Eq. (1), we assume that $\mathcal{M}_\Omega$ is Gaussian white noise and each entry in $\mathcal{M}_\Omega$ is independent and identically distributed with precision τ_0. Thus, we obtain the following likelihood

$$p\left(\mathcal{Y}_\Omega|\mathcal{X},\mathcal{E}\right)=\prod\nolimits_{\boldsymbol{i}}\mathcal{N}\left(y_{\boldsymbol{i}}|x_{\boldsymbol{i}}+e_{\boldsymbol{i}},\tau_0^{-1}\right)^{\mathcal{O}_{\boldsymbol{i}}} \tag{4}$$

where $\mathcal{O}$ is an indicator tensor with entries $\mathcal{O}_{\boldsymbol{i}}=1$ if $\boldsymbol{i}\in\Omega$. $x_{\boldsymbol{i}}$ and $e_{\boldsymbol{i}}$ are the entries in $\mathcal{X}$ and $\mathcal{E}$, respectively. In this study, we propose to recover the latent $\mathcal{L}$ from $\mathcal{Y}_\Omega$ by exploiting the low-rank structure $\mathcal{X}$ and fitting the complex non-low-rank structure $\mathcal{E}$. To this end, we specifically design priors for $\mathcal{X}$ and $\mathcal{E}$ as follows.

A. Low-rank structure modeling

We reformulate the CP tensor rank in a new form, with which a sparsity-induced low-rank model is proposed to represent $\mathcal{X}$ and the corresponding tensor rank of $\mathcal{X}$ then can be determined automatically by exploiting the sparsity in the CP factorization weights.

1) CP factorization: In CP factorization, tensor $\mathcal{X}$ is factorized as a sum of R rank-one tensors as follows:

$$\mathcal{X}=\sum\nolimits_{r=1}^{R}\lambda_r\boldsymbol{u}_r^{(1)}\circ\cdots\circ\boldsymbol{u}_r^{(K)}=[\![\boldsymbol{\lambda};\boldsymbol{U}^{(1)},\cdots,\boldsymbol{U}^{(K)}]\!], \tag{5}$$

where $\boldsymbol{u}_r^{(k)}\in\mathbb{R}^{n_k}$ is the factor vector in k-th mode and $k=1,\cdots,K$. $\circ$ denotes the outer product. For simplicity, the CP factorization can be represented as the right part of Eq. (5). Here $\boldsymbol{\lambda}=[\lambda_1,\cdots,\lambda_R]^T$ is the weights vector, and $\boldsymbol{U}^{(k)}=\left[\boldsymbol{u}_1^{(k)},\cdots,\boldsymbol{u}_R^{(k)}\right]\in\mathbb{R}^{n_k\times R}$ denotes the k-th factor matrix.

2) CP rank vs. sparsity-induced rank:

Definition 1 (CP rank): The rank of tensor $\mathcal{X}$, denoted $\mathrm{rank}(\mathcal{X})$, is defined as the smallest number of rank-one tensors in the CP factorization of $\mathcal{X}$ [43].

CP rank is a specialized tensor rank that degenerates to the matrix rank when $K=2$. However, since CP factorization is ill-posed [43], the determination of CP rank is NP hard.

To address this problem, we reformulate the CP rank in a new form. Specifically, given a tensor $\mathcal{X}$ of $\operatorname{rank}(\mathcal{X}) = R_0$, assume we have all of its CP factorizations with $R \gg R_0$, and we can find the one with the sparsest weight vector $\boldsymbol{\lambda}$. According to Definition 1, only R_0 weights in $\boldsymbol{\lambda}$ will be non-zero, viz., $\operatorname{rank}(\mathcal{X}) = \|\boldsymbol{\lambda}\|_0$. Inspired by this, we give the following definition for sparsity-induced tensor rank.

Definition 2 (Sparsity induced rank): Given CP factorization of $\mathcal{X}$ with $R \gg \operatorname{rank}(\mathcal{X})$ and the sparsest weight vector $\boldsymbol{\lambda}$, $\operatorname{rank}(\mathcal{X}) = \|\boldsymbol{\lambda}\|_0$.

While being equivalent, Definition 1 and Definition 2 differ in two aspects. 1) *Definition form.* Definition 2 defines the CP rank by exploiting the sparsity in CP factorization rather than counting the number of factorized rank-one tensors in Definition 1. 2) *Difficulty of rank determination.* Definition 2 can ease the determination of CP rank. According to Definition 2, rank determination can be cast into exploiting the sparsity in the CP factorization. Following this, appropriate sparse prior is introduced into the CP factorization as regularization to effectively reduce the solution space. When the sparse solution is obtained, the CP rank is the sparseness of weight vector $\boldsymbol{\lambda}$. Although the ℓ_0 norm based sparse learning problem is also NP-hard, many techniques (e.g., ℓ_1 norm) have been proposed to provide satisfactory approximation solutions by solving a much easier problem (e.g., convex optimization or Bayesian inference). Thus, the rank determination problem resulting from Definition 2 can be addressed effectively. In contrast, according to the CP tensor rank in Definition 1, exhaustive search is required among the infinitely possible CP factorizations of a given tensor to determine its CP tensor rank. Therefore, based on the sparsity-induced tensor rank, the difficulty in rank determination can be reduced.

3) Sparsity-induced low rank model: For low-rank tensor recovery, a feasible method is to separately recover the decomposed weight vector and corresponding factor matrices generated in its CP factorization. Given these decomposed components, the low-rank tensor can then be reconstructed by fusing these components in CP factorization form. However, there often exist infinite CP factorization results for a given tensor; thus, it is difficult to find appropriate factorization for recovery. Fortunately, according to the reformulated sparsity induced tensor rank, for a low-rank tensor (e.g., rank is R_0), its CP factorization with R (e.g., $R \gg R_0$) factors contains a special form, i.e., factorization with the sparsest decomposed weight vector. Thus, to recover the low-rank tensor, we can set a sufficiently large R and search a specific CP factorization with the sparsest weight vector. Due to the sparse property, we can introduce appropriate sparsity regularizations to guide searching a unique solution for tensor completion. For simplicity, we refer to this specific CP factorization as the sparsest CP factorization. Following this idea, we can formulate the low-rank tensor recovery problem as a sparse learning problem, in which we attempt to recover both the sparsest decomposed weight vector (i.e., the sparse representation) and the corresponding factor matrices (i.e., the dictionary). This is similar to the conventional sparse learning problem to simultaneous recover the dictionary and sparse representation. In addition, according to Definition 2, the real tensor rank is equal to the sparseness of the decomposed weight vector in the sparsest CP factorization. Thus, by solving such a sparse learning problem, we can also determine the tensor rank automatically. In the following, we provide details of establishing a prior model for the low-rank structure $\mathcal{X}$ based on Definition 2.

Specifically, Definition 2 indicates that a low-rank tensor can be well represented by a sparse CP factorization (i.e., with a sparse weight vector $\boldsymbol{\lambda}$ in Eq. (5)). Moreover, the tensor rank can also be determined automatically by the sparseness of weight vector $\boldsymbol{\lambda}$. Thus, we propose to model the low-rank structure $\mathcal{X}$ by exploiting the sparsity of the weight vector $\boldsymbol{\lambda}$ in its CP factorization. Such a low rank model is referred to as a *sparsity-induced low rank model*. With a small tensor rank (i.e., $R \gg \operatorname{rank}(\mathcal{X})$), this model amounts to representing $\mathcal{X}$ on a tensor dictionary with sparse coefficients $\boldsymbol{\lambda}$, where each dictionary atom is a rank-one tensor as $\boldsymbol{u}_r^{(1)} \circ \cdots \circ \boldsymbol{u}_r^{(K)}$ in Eq. 5. These atoms can well preserve the multidimensional structure of tensors, viz., the new tensor rank intrinsically depends on the multidimensional structure of tensors, which is totally different from the rank norms defined when unfolding the matrices of a tensor [1], [25], because only the two-dimensional structure is depicted in unfolding matrices.

To model the sparsity of $\boldsymbol{\lambda}$, we adopt the following two-level hierarchical reweighted Laplace prior [44]:

$$\boldsymbol{\lambda} \sim \mathcal{N}\left(\boldsymbol{\lambda}|\mathbf{0}, \mathbf{diag}\left(\boldsymbol{\gamma}\right)\right), \boldsymbol{\gamma} \sim \prod\nolimits_{r=1}^{R} \operatorname{Ga}(\gamma_r|1, \frac{\kappa_r}{2}) \tag{6}$$

where $\boldsymbol{\lambda}$ and $\boldsymbol{\gamma} = [\gamma_1, ..., \gamma_R]^T$ follow a zero-mean Gaussian distribution and a product of Gamma distributions, respectively. γ_r and κ_r are the respective r-th entries of $\boldsymbol{\gamma}$ and $\boldsymbol{\kappa} = [\kappa_1, ..., \kappa_R]^T$. Here the motivation is twofold. First, the reweighed Laplace prior outperforms traditional sparsity priors (e.g., the Laplace prior) in exploiting the low-rank structure in a given tensor (see Remark 1). Second, the inherent two-level hierarchical prior is conjugates to the likelihood in Eq. (4), which enables the inference discussed in Section IV tractable. In the following, we discuss the former reason in detail.

To exploit the sparsity in $\boldsymbol{\lambda}$, ℓ_0 norm is the best choice; however, ℓ_0 norm often renders the related sparse learning problem NP hard. Thus, most existing literatures [45], [46] have sought an appropriate surrogate for ℓ_0 norm. A famous surrogate is the Laplace prior (i.e., ℓ_1 norm), which can be formulated as:

$$p\left(\boldsymbol{\lambda}|w\right) \propto \exp\left(-w\|\boldsymbol{\lambda}\|_1\right), \tag{7}$$

w is the scale parameter. However, it demonstrates a key difference to ℓ_0 norm, where a large magnitude entry will be penalized more heavily than a small magnitude entry [45]. This imbalance often limits the accuracy of the sparse solution. To mitigate this problem, we employ the reweighted Laplace prior given as Eq. (6). According to [44], [46], it has been proven that these two hierarchical priors in Eq. (6) amount to a reweighted Laplace prior on $\boldsymbol{\lambda}$ as:

$$p\left(\boldsymbol{\lambda}|\boldsymbol{K}\right) \propto \exp\left(-\|\boldsymbol{K}\boldsymbol{\lambda}\|_1\right), \tag{8}$$

where $\boldsymbol{K} = \mathbf{diag}\left([\sqrt{\kappa_1}, ..., \sqrt{\kappa_R}]^T\right)$. Differing from the Laplace prior, an additional diagonal weight matrix $\boldsymbol{K}$ is

introduced to adjust the magnitude of each entry in $\boldsymbol{\lambda}$. Moreover, Lemma 3 proves that each κ_r is a decreasing function over $|\lambda_r|$ in the Bayesian inference, which is similar as [44]. Therefore, the reweighted Laplace prior can well mitigate the negative influence from the entry magnitude when depicting the sparsity. By introducing such a powerful sparsity prior, we obtain the following remark:

Remark 1: The sparsity-induced low-rank model can effectively exploit the low-rank structure in the tensor.

4) Regularized factor matrix: In this study, weight vector $\boldsymbol{\lambda}$ and factor matrices $\boldsymbol{U}^{(k)}$s from the CP factorization of $\mathcal{X}$ are modeled separately. Although $\boldsymbol{\lambda}$ has been well regularized as Eq. (6), infinite solutions still exist for the CP factorization of $\mathcal{X}$, e.g., $\mathcal{X} = [\![\boldsymbol{\lambda}; c^{-1}\boldsymbol{U}^{(1)}, \boldsymbol{U}^{(2)}, \cdots, c\boldsymbol{U}^{(K)}]\!]$ with any scalar $c \neq 0$. To address this problem, we assume that each entry $u_{ir}^{(k)}$ of the factor matrix $\boldsymbol{U}^{(k)}$ follows Gaussian distribution independently and identically as:

$$u_{ir}^{(k)} \sim \mathcal{N}\left(u_{ir}^{(k)} | \mu^{(k)}, \tau^{(k)-1}\right). \tag{9}$$

which amounts to regularizing each entry with the ℓ_2 norm. Note that we adopt an individual Gaussian distribution for each $\boldsymbol{U}^{(k)}$. This enables capturing the specific characteristic for each factor matrix. Moreover, such a Gaussian prior is a conjugate prior for the likelihood in Eq. (4), thereby rendering the subsequent inference feasible and solvable (Section IV). In addition, such a prior can be easily extended to consider the spatial similarity (e.g., smoothness) of visual tensor data (Section VI-B). To complete the Bayesian model, we further introduce conjugate priors over the parameters of the Gaussian distributions, $\mu^{(k)}$s and $\tau^{(k)}$s as

$$\mu^{(k)}, \tau^{(k)} \sim \mathcal{N}\left(\mu^{(k)} | \mu_0, (\beta_0 \tau^{(k)})^{-1}\right) \mathrm{Ga}\left(\tau^{(k)} | a_0, b_0\right). \tag{10}$$

where $\mu^{(k)}$ and $\tau^{(k)}$ jointly follow a Gaussian-gamma distribution parametrized by μ_0, β_0, a_0 and b_0.

In summary, the low rank structure $\mathcal{X}$ is modeled by exploiting the sparsity in weight vector $\boldsymbol{\lambda}$ and regularizing factor matrices $\boldsymbol{U}^{(k)}$ in its CP factorization.

B. Mixture of Gaussians for non-low-rank structure

As shown in Fig. 1, the distribution of entries in $\mathcal{E}$ can be complex (e.g., heavy-tailed or multi-mode) in practice. To impose a suitable prior on complex $\mathcal{E}$, we assume that each entry $e_{\boldsymbol{i}}$ comes from a mixture of D Gaussians as follows:

$$e_{\boldsymbol{i}} \sim \sum\nolimits_{d=1}^{D} \pi_d \mathcal{N}\left(e_{\boldsymbol{i}} | \mu_d, \tau_d^{-1}\right), \tag{11}$$

where $\pi_d \geq 0$ is the mixing proportion with $\sum_{d=1}^{D} \pi_d = 1$. $\mathcal{N}\left(e_{\boldsymbol{i}} | \mu_d, \tau_d^{-1}\right)$ denotes the d-th Gaussian component with mean μ_d and precision τ_d. By introducing D indicator variables $z_{\boldsymbol{i}}^d$s for $d = 1, \cdots, D$, Eq. (11) can be represented equivalently as a two-level generative model [47]:

$$e_{\boldsymbol{i}} \sim \prod\nolimits_{d=1}^{D} \mathcal{N}\left(e_{\boldsymbol{i}} | \mu_d, \tau_d^{-1}\right)^{z_{\boldsymbol{i}}^d}, \ \boldsymbol{z}_{\boldsymbol{i}} \sim \mathrm{Multinomial}\left(\boldsymbol{z}_{\boldsymbol{i}} | \boldsymbol{\pi}\right), \tag{12}$$

where $\boldsymbol{z}_{\boldsymbol{i}} = \left(z_{\boldsymbol{i}}^1, ..., z_{\boldsymbol{i}}^D\right) \in \{0,1\}^D$ with $\sum_{d=1}^{D} z_{\boldsymbol{i}}^d = 1$ following a multinomial distribution parametrized by $\boldsymbol{\pi} = (\pi_1, ..., \pi_D)$. To model $\mathcal{E}$ flexibly, we further impose conjugate priors over the parameters of μ_ds, τ_ds, and $\boldsymbol{\pi}$ as follows:

$$\begin{aligned} \mu_d, \tau_d &\sim \mathcal{N}(\mu_d | \mu_0, (\beta_0 \tau_d)^{-1}) \mathrm{Ga}\left(\tau_d | a_0, b_0\right) \\ \boldsymbol{\pi} &\sim \mathrm{Dir}\left(\boldsymbol{\pi} | \boldsymbol{\alpha}_0\right), \end{aligned} \tag{13}$$

where $\boldsymbol{\pi}$ follows Dirichlet distribution $\mathrm{Dir}\left(\boldsymbol{\pi} | \boldsymbol{\alpha}_0\right)$ with parameter $\boldsymbol{\alpha}_0 = (\alpha_{01}, \cdots, \alpha_{0D})$. The MOG has obvious advantage relative to representing the complex $\mathcal{E}$, which leads to the following remark.

Remark 2: The MOG can fit a wide range of non-low-rank structures $\mathcal{E}$s.

It has been demonstrated that an MOG has universal ability to approximate any continuous distributions [47]. In addition, it has been proven [33] that both the Dirac delta distribution for a zero variable (i.e., $\delta(0)$) and the spike-and-slab sparsity prior [48] are special cases of the MOG. Therefore, the MOG can fit a wide range of $\mathcal{E}$s (i.e., zero, sparse or more complex), which leads to a more robust representation for the real tensor data and ultimately improves completion performance (Section V-A).

IV. Inference

According to the above likelihood and priors, we show the probabilistic graphical structure of the proposed model in Figure 2. Most previous studies have inferred latent tensor $\mathcal{L}$ with MAP estimate as Eq. (2); however, generative models often perform poorly in the context of MAP [49]. We adopt the Bayesian minimum mean squared error estimate (MMSE) suggested in [49] for tensor completion as follows:

$$\hat{\mathcal{L}} = \arg\min_{\hat{\mathcal{L}}} \int \left\| \tilde{\mathcal{L}} - \mathcal{L} \right\|_F^2 p\left(\mathcal{L} | \mathcal{Y}_\Omega\right) d\mathcal{L} = \mathbb{E}\left[\mathcal{L} | \mathcal{Y}_\Omega\right], \tag{14}$$

where $\hat{\mathcal{L}}$ equals the expectation $\mathbb{E}\left[\mathcal{L} | \mathcal{Y}_\Omega\right]$ of $\mathcal{L}$ under the posterior distribution $p\left(\mathcal{L} | \mathcal{Y}_\Omega\right)$, and $\|\mathcal{A}\|_F$ denotes the Frobenius norm on tensor $\mathcal{A}$. In contrast to MAP that only produces a point estimation, MMSE exploits the uncertainty of $\mathcal{L}$ and adopts the probabilistic mean as a solution which often demonstrates better robustness and performance. However, it is difficult to conduct the expectation in Eq. (14). To circumvent this problem, we can adopt the mean of samples drawn from the posterior of $\mathcal{L}$ as an unbiased estimate for $\hat{\mathcal{L}}$ [49]. Since we model the non-low-rank structure $\mathcal{E}$ and the CP factorization of the low-rank structure $\mathcal{X}$ rather than directly modelling $\mathcal{L}$, we draw samples from the corresponding posteriors for $\boldsymbol{\lambda}$, $\boldsymbol{U}^{(k)}$ and $\mathcal{E}$. Given their sample means, $\hat{\mathcal{L}}$ can then be obtained from Eqs. (5) (3). Specifically, based on Eqs. (4) (6) (9) (10) (12) (13), we can obtain the posterior of all involved variables as follows:

$$p\left(\boldsymbol{\lambda}, \mathcal{U}, \mathcal{Z}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{\mu}_e, \boldsymbol{\tau}_e | \mathcal{Y}_\Omega\right), \tag{15}$$

where $\mathcal{U} = \left\{U^{(k)}\right\}$, $\mathcal{Z} = \left\{\boldsymbol{z}_{\boldsymbol{i}}\right\}$, $\boldsymbol{\mu} = \left\{\mu^{(k)}\right\}$, $\boldsymbol{\tau} = \left\{\tau^{(k)}\right\}$, $\boldsymbol{\mu}_e = \{\mu_d\}$ and $\boldsymbol{\tau}_e = \{\tau_d\}$ are introduced for simplicity. Then, Gibbs sampling is employed on Eq. (15) to sample each variable as follows.

A. Gibbs samplers for low-rank structure

The low-rank structure $\mathcal{X}$ is decomposed into weight vector $\boldsymbol{\lambda}$ and factor matrices $\boldsymbol{U}^{(k)}$s.

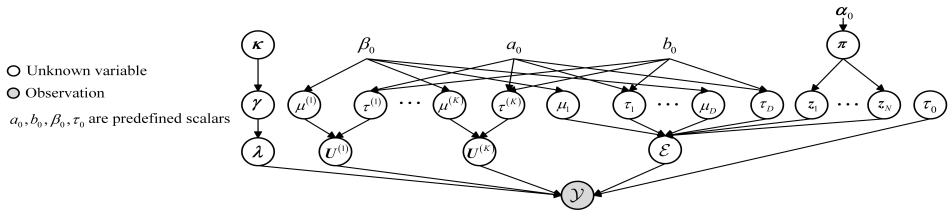


Fig. 2. Probabilistic graphical structure of the proposed model. $\boldsymbol{z}_1, \cdots, \boldsymbol{z}_N$ denote all $\boldsymbol{z}_{\boldsymbol{i}}$ with all possible $\boldsymbol{i}$s for simplicity.

1) Sampler for weight vector $\boldsymbol{\lambda}$: According to Figure 2, the posterior for $\boldsymbol{\lambda}$ can be inferred by receiving two kinds of messages. The first one comes from its parent and is represented as the prior in Eq. (6), and the second one sent by the observation $\mathcal{Y}$ and its parents (i.e., factor matrices $\boldsymbol{U}^{(k)}$, non-low-rank structure $\mathcal{E}$ and τ_0) is expressed as the likelihood in Eq. (4). Thus, given $\boldsymbol{U}^{(k)}$s, $\mathcal{E}$, $\boldsymbol{\gamma}$ and τ_0, we obtain the posterior over $\boldsymbol{\lambda}$ as follows:

$$p\left(\boldsymbol{\lambda}|-\right) \propto \mathcal{N}\left(\boldsymbol{\lambda}|\mathbf{0},\Gamma\right)\prod_{\boldsymbol{i}}\mathcal{N}\left(y_{\boldsymbol{i}}|e_{\boldsymbol{i}}+\sum_{r=1}^{R}\lambda_r\prod_{k=1}^{K}u_{i_k r}^{(k)},\tau_0^{-1}\right)^{\mathcal{O}_{\boldsymbol{i}}}. \tag{16}$$

Since all entries in $\boldsymbol{\lambda}$ are dependent on each other, we must sample each entry alternatively. To this end, let $\tilde{y}_{\boldsymbol{i}}^r = y_{\boldsymbol{i}} - e_{\boldsymbol{i}} - \sum_{t\neq r}^{R}\lambda_t\prod_{k=1}^{K}u_{i_k t}^{(k)}$ and $\tilde{b}_{\boldsymbol{i}}^r = \prod_{k=1}^{K}u_{i_k r}^{(k)}$. Then, the posterior for entry λ_r then can be given as follows:

$$p\left(\lambda_r|-\right) \propto \exp\left(-\frac{\lambda_r^2}{2\gamma_r}\right)\prod_{\boldsymbol{i}}\exp\left[-\frac{\mathcal{O}_{\boldsymbol{i}}\left(\tilde{y}_{\boldsymbol{i}}^r-\lambda_r\tilde{b}_{\boldsymbol{i}}^r\right)^2}{2\tau_0^{-1}}\right]. \tag{17}$$

where messages from other entries λ_t $(t \neq r)$ are introduced. As can be seen, we draw λ_r from a Gaussian distribution as

$$\lambda_r \sim \mathcal{N}\left(\tilde{\mu}_{\lambda_r}, \tilde{\tau}_{\lambda_r}^{-1}\right), \tag{18}$$

where the posterior parameters can be updated as

$$\tilde{\tau}_{\lambda_r} = \gamma_r^{-1} + \tau_0\sum_{\boldsymbol{i}}\mathcal{O}_{\boldsymbol{i}}\tilde{b}_{\boldsymbol{i}}^{r2},\ \ \tilde{\mu}_{\lambda_r} = \tau_{\lambda_r}^{-1}\tau_0\sum_{\boldsymbol{i}}\mathcal{O}_{\boldsymbol{i}}\tilde{b}_{\boldsymbol{i}}^r\tilde{y}_{\boldsymbol{i}}^r. \tag{19}$$

2) Sampler for hyperperameter: $\boldsymbol{\gamma}$ By integrating the messages from weight vector $\boldsymbol{\lambda}$ and its hyperprior expressed in Eq. (6), the posterior of $\boldsymbol{\gamma}$ can be formulated as

$$p\left(\boldsymbol{\gamma}|\sim\right) \propto \mathcal{N}\left(\boldsymbol{\lambda}|\mathbf{0},\Gamma\right)\prod_{r=1}^{R}\mathrm{Ga}\left(\gamma_r|1,\frac{2}{\kappa_r}\right), \tag{20}$$

where each entry λ_r is independently distributed as

$$p\left(\gamma_r|\sim\right) \propto \gamma_r^{-1/2}\exp\left(-\frac{\lambda_r^2+\kappa_r\gamma_r^2}{2\gamma_r}\right). \tag{21}$$

Thus, γ_r can be drawn from a generalized inverse Gaussian distribution as

$$\gamma_r \sim \mathcal{GIG}\left(\gamma_r|\kappa_r,\lambda_r^2,1/2\right). \tag{22}$$

Given the r-th entry λ_r and κ_r, the posterior expectation of γ_r can be given as

$$\mathbb{E}\left[\gamma_r\right] = \frac{|\lambda_r|\,\mathcal{K}_{3/2}(|\lambda_r|\sqrt{\kappa_r})}{\sqrt{\kappa_r}\mathcal{K}_{1/2}(|\lambda_r|\sqrt{\kappa_r})} \tag{23}$$

which linearly depends on λ_r and performs as a decreasing function of κ_r. To demonstrate this point, we introduce the following Lemma.

Lemma 1: Given the modified Bessel function of second kind $K_v\left(z\right)$, we define a function $f_v\left(z\right) = zK_{v+1}\left(z\right)/K_v\left(z\right)$, then, for any variable $x > 0$, $f_{1/2}\left(x\right) = x+1$.

Proof: In [50], it has been proved that

$$\begin{aligned} &K_{v+1}\left(z\right) - K_{v-1}\left(z\right) = \frac{2v}{x}K_n\left(z\right); \\ &K_{v+\frac{1}{2}}\left(z\right) = K_{-v-\frac{1}{2}}\left(z\right),\ \forall v\in\mathbb{R}. \end{aligned} \tag{24}$$

When $v = 1/2$, we have

$$\frac{K_{3/2}(z)}{K_{1/2}(z)} = \frac{1}{x} + \frac{K_{-1/2}\left(z\right)}{K_{1/2}(z)} = \frac{1}{x}+1. \tag{25}$$

Thus, $f_{1/2}\left(x\right) = x\left(1/x+1\right) = x+1$. ■

Lemma 2: Given λ_r and κ_r, $\mathbb{E}\left[\gamma_r\right] = \frac{|\lambda_r|}{\sqrt{\kappa_r}} + \frac{1}{\kappa_r}$.

Proof: In Eq. (23), $\mathbb{E}\left[\gamma_r\right] = \frac{f_{1/2}(|\lambda_r|\sqrt{\kappa_r})}{\kappa_r}$. According to Lemma 1, $\mathbb{E}\left[\gamma_r\right] = \frac{|\lambda_r|}{\sqrt{\kappa_r}} + \frac{1}{\kappa_r}$. ■

3) Sampler for hyperparameter $\boldsymbol{\kappa}$: The posterior over $\boldsymbol{\kappa}$ can be inferred by receiving the messages from hyperparameter $\boldsymbol{\gamma}$, as shown in Fig. 2, and the posterior over each entry κ_r can be formulated as follows:

$$p\left(\kappa_r|-\right) \propto \mathrm{Ga}\left(\gamma_r|1,\frac{2}{\kappa_r}\right) \propto \kappa_r\exp\left(-\frac{\gamma_r\kappa_r}{2}\right). \tag{26}$$

Thus, we can draw κ_r from a Gamma distribution as

$$\kappa_r \sim \mathrm{Ga}\left(\kappa_r|2,\frac{2}{\gamma_r}\right). \tag{27}$$

Lemma 3: Given λ_r and the corresponding $\mathbb{E}\left[\gamma_r\right]$ for Eq. (27), there exists a decreasing function ρ s.t. $\mathbb{E}\left[\kappa_r\right] = \rho(|\lambda_r|)$.

Proof: Given λ_r and the corresponding $\mathbb{E}\left[\lambda_r\right]$, the posterior for κ_r can be reformulated as $\mathrm{Ga}\left(\kappa_r|2,\mathbb{E}\left[\lambda_r\right]\right)$, and the resulted posterior expectation $\mathbb{E}\left[\kappa_r\right] = 4/\mathbb{E}\left[\lambda_r\right]$. Thus, according to Lemma 2, we obtain

$$\mathbb{E}\left[\kappa_r\right] = \rho\left(|\lambda_r|\right) \propto \left(|\lambda_r| + \mathrm{const}\right)^{-1} \tag{28}$$

where const denotes the variable relative to λ_r. Thus, $\rho\left(|\lambda_r|\right)$ is a decreasing function over λ_r. ■

With Eq. (8), weight λ_r is shrunk with $\sqrt{\kappa_r}$ while exploiting the sparsity of $\boldsymbol{\lambda}$. Based on Lemma 3, $\mathbb{E}\left[\kappa_r\right] \propto \left(|\lambda_r| + \mathrm{const}\right)^{-1}$, thus λ_r will be shrunk less than λ_j when $|\lambda_r| > |\lambda_j|$, which is consistent with Remark 1 in Section III-A3.

4) Sampler for factor matrix $\boldsymbol{U}^{(k)}$: Similarly, the posterior for $\boldsymbol{U}^{(k)}$ is jointly determined by observation $\mathcal{Y}$, other factor matrices $\boldsymbol{U}^{(j)}$ with $j \neq k$, and hyperparameters $\mu^{(k)}$, $\tau^{(k)}$ in its prior. Since the entries in each column of $\boldsymbol{U}^{(k)}$ are correlated with each other, we sample each entry $u_{ir}^{(k)}$ with the posterior

$$p(u_{ir}^{(k)}|-) \propto \prod_{\boldsymbol{i}:i_k=i} \left[\mathcal{N}\left(y_{\boldsymbol{i}}|e_{\boldsymbol{i}} + \sum_{t=1}^{R} \lambda_t \prod_{s=1}^{K} u_{i_s t}^{(s)}, \tau_0^{-1}\right)^{\mathcal{O}_{\boldsymbol{i}}} \mathcal{N}\left(u_{ir}^{(k)}|\mu^{(k)}, \tau^{(k)-1}\right)\right]. \tag{29}$$

Thus, $u_{ir}^{(k)}$ can be drawn from a Gaussian distribution as

$$u_{ir}^{(k)} \sim \mathcal{N}\left(\tilde{\mu}_{u_{ir}^{(k)}}, \tilde{\tau}_{u_{ir}^{(k)}}\right), \tag{30}$$

where let $\tilde{c}_{\boldsymbol{i}}^{rk} = \lambda_r \prod_{s\neq k}^{K} u_{i_s r}^{(s)}$ and the corresponding parameters are

$$\begin{aligned} \tilde{\tau}_{u_{ir}^{(k)}} &= \sum_{\boldsymbol{i}:i_k=i} \tau_0 \mathcal{O}_{\boldsymbol{i}} \tilde{c}_{\boldsymbol{i}}^{rk2} + \tau^{(k)}, \\ \tilde{\mu}_{u_{ir}^{(k)}} &= \tau_{u_{ir}^{(k)}}^{-1} \left(\sum_{\boldsymbol{i}:i_k=i} \tau_0 \mathcal{O}_{\boldsymbol{i}} \tilde{y}_{\boldsymbol{i}}^{r} \tilde{c}_{\boldsymbol{i}}^{rk} + \tau^{(k)} \mu^{(k)}\right). \end{aligned} \tag{31}$$

5) Sampler for hyperparameter $\mu^{(k)}$ and $\tau^{(k)}$: By receiving the messages from $\boldsymbol{U}^{(k)}$, β_0, a_0 and b_0, the posterior over $\mu^{(k)}$ can be given as

$$p\left(\mu^{(k)}|-\right) \propto \prod_{i=1}^{n_k} \prod_{r=1}^{R} \mathcal{N}\left(u_{ir}^{(k)}|\mu^{(k)}, \tau^{(k)-1}\right) \mathcal{N}\left(\mu^{(k)}|\mu_0, \left(\beta_0 \tau^{(k)}\right)^{-1}\right) \tag{32}$$

Thus, $\mu^{(k)}$ can be drawn from a Gaussian distribution as

$$\mu^{(k)} \sim \mathcal{N}\left(\tilde{\mu}_{\mu^{(k)}}, \tilde{\tau}_{\mu^{(k)}}^{-1}\right), \tag{33}$$

where the posterior parameters are

$$\tilde{\tau}_{\mu^{(k)}} = \tau^{(k)}\left(\beta_0 + n_k R\right), \tilde{\mu}_{\mu^{(k)}} = \frac{\tau^{(k)}}{\tau_{\mu^{(k)}}} \left(\sum_{i=1}^{n_k} \sum_{r=1}^{R} u_{ir}^{(k)} + \beta_0 \mu_0\right). \tag{34}$$

Similarly, we have the posterior over $\tau^{(k)}$ as

$$p\left(\tau^{(k)}|-\right) \propto \prod_{i=1}^{n_k} \prod_{r=1}^{R} \left[\mathcal{N}\left(u_{ir}^{(k)}|\mu^{(k)}, \tau^{(k)-1}\right) \mathcal{N}\left(\mu^{(k)}|\mu_0, \left(\beta_0 \tau^{(k)}\right)^{-1}\right) \mathrm{Ga}\left(\tau^{(k)}|c_0, e_0\right)\right]. \tag{35}$$

It can be seen that $\tau_d^{(k)}$ can be drawn from the following Gamma distribution

$$\tau^{(k)} \sim \mathrm{Ga}\left(\tilde{a}^{(k)}, \tilde{b}^{(k)}\right), \tag{36}$$

where the parameters are

$$\begin{aligned} \tilde{a}^{(k)} &= a_0 + \left(n_k R + 1\right)/2; \\ \tilde{b}^{(k)} &= b_0 + \left[\sum_{i=1}^{n_k} \sum_{r=1}^{R} \left(u_{ir}^{(k)} - \mu^{(k)}\right)^2 + \beta_0 \left(\mu^{(k)} - \mu_0\right)^2\right]/2. \end{aligned} \tag{37}$$

B. Gibbs sampler for non-low-rank structure

1) Sampler for non-low-rank structure $\mathcal{E}$: Since each entry $e_{\boldsymbol{i}}$ is assumed to be independent to others, given messages from $\mathcal{Y}$, $\boldsymbol{U}^{(k)}$, $\boldsymbol{\lambda}$, μ_ds and τ_ds, we obtain the following posterior:

$$p\left(e_{\boldsymbol{i}}|-\right) \propto \mathcal{N}\left(y_{\boldsymbol{i}}|e_{\boldsymbol{i}} + x_{\boldsymbol{i}}, \tau_0^{-1}\right)^{\mathcal{O}_{\boldsymbol{i}}} \prod_{d=1}^{D} \mathcal{N}\left(e_{\boldsymbol{i}}|\mu_d, \tau_d^{-1}\right)^{z_{\boldsymbol{i}}^d}, \tag{38}$$

and $e_{\boldsymbol{i}}$ can be drawn from a Gaussian distribution as

$$e_{\boldsymbol{i}} \sim \mathcal{N}\left(\tilde{\mu}_{e_{\boldsymbol{i}}}, \tilde{\tau}_{e_{\boldsymbol{i}}}^{-1}\right), \tag{39}$$

with parameters

$$\begin{aligned} \tilde{\tau}_{e_{\boldsymbol{i}}} &= \mathcal{O}_{\boldsymbol{i}} \tau_0 + \sum_{d=1}^{D} \tau_d z_{\boldsymbol{i}}^d, \\ \tilde{\mu}_{e_{\boldsymbol{i}}} &= \tilde{\tau}_{e_{\boldsymbol{i}}}^{-1} \left[\tau_0 \mathcal{O}_{\boldsymbol{i}} \left(y_{\boldsymbol{i}} - x_{\boldsymbol{i}}\right) + \sum_{d=1}^{D} \tau_d z_{\boldsymbol{i}}^d \mu_d\right]. \end{aligned} \tag{40}$$

2) Sampler for hyperparameter μ_d and τ_d: With messages from $\mathcal{E}$ and parameters β_0, a_0, b_0 in the hyperprior, the posterior over μ_d can be given as follows:

$$p\left(\mu_d|-\right) \propto \prod_{\boldsymbol{i}} \mathcal{N}\left(e_{\boldsymbol{i}}|\mu_d, \tau_d^{-1}\right)^{z_{\boldsymbol{i}}^d} \mathcal{N}\left(\mu_d|\mu_0, \left(\beta_0 \tau_d\right)^{-1}\right). \tag{41}$$

Thus, μ_d can be drawn from a Gaussian distribution as:

$$\mu_d \sim \mathcal{N}\left(\tilde{\mu}_{\mu_d}, \tilde{\tau}_{\mu_d}^{-1}\right), \tag{42}$$

with parameters

$$\tilde{\tau}_{\mu_d} = \tau_d \left(\sum_{\boldsymbol{i}} z_{\boldsymbol{i}}^d + \beta_0\right), \ \tilde{\mu}_{\mu_d} = \tilde{\tau}_{\mu_d}^{-1} \tau_d \left(\sum_{\boldsymbol{i}} z_{\boldsymbol{i}}^d e_{\boldsymbol{i}} + \beta_0 \mu_0\right). \tag{43}$$

Similarly, we have the posterior over τ_d as

$$p\left(\tau_d|-\right) \propto \prod_{\boldsymbol{i}} \mathcal{N}\left(z_{\boldsymbol{i}}|\mu_d, \tau_d^{-1}\right)^{z_{\boldsymbol{i}}^d} \mathcal{N}\left(\mu_d|\mu_0, \left(\beta_0 \tau_d\right)^{-1}\right) \mathrm{Ga}\left(\tau_d|a_0, b_0\right). \tag{44}$$

Thus, τ_d can be drawn from a Gamma distribution as

$$\tau_d \sim \mathrm{Ga}\left(\tilde{a}_d, \tilde{b}_d\right), \tag{45}$$

with parameters

$$\begin{aligned} \tilde{a}_d &= a_0 + \left(\sum_{\boldsymbol{i}} z_{\boldsymbol{i}}^d + 1\right)/2, \\ \tilde{b}_d &= b_0 + \left[\sum_{\boldsymbol{i}} z_{\boldsymbol{i}}^d e_{\boldsymbol{i}} + \beta_0 \left(\mu_d - \mu_0\right)^2\right]/2; \end{aligned} \tag{46}$$

3) Sampler for $\boldsymbol{z}_{\boldsymbol{i}}$: Similarly, we obtain the posterior over $\boldsymbol{z}_{\boldsymbol{i}}$ as

$$p\left(\boldsymbol{z}_{\boldsymbol{i}}|-\right) \propto \mathrm{Multinomial}\left(\boldsymbol{z}_{\boldsymbol{i}}|\boldsymbol{\pi}\right). \tag{47}$$

$\boldsymbol{z}_{\boldsymbol{i}}$ thus can be drawn from a multinomial distribution as

$$\boldsymbol{z}_{\boldsymbol{i}} \sim \mathrm{Multinomial}\left(\boldsymbol{z}_{\boldsymbol{i}}|\tilde{\boldsymbol{\pi}}\right), \tag{48}$$

with $\tilde{\boldsymbol{\pi}} = (\tilde{\pi}_1, ..., \tilde{\pi}_D)$ and each entry $\tilde{\pi}_d$ is given as

$$\tilde{\pi}_d = \pi_d \mathcal{N}\left(e_{\boldsymbol{i}}|\mu_d, \tau_d^{-1}\right) / \sum_{t=1}^{D} \pi_t \mathcal{N}\left(e_{\boldsymbol{i}}|\mu_t, \tau_t^{-1}\right). \tag{49}$$

4) Sampler for hyperparameter $\boldsymbol{\pi}$: By integrating messages from $\boldsymbol{z}_{\boldsymbol{i}}$ and $\boldsymbol{\alpha}_0$ in the hyperprior, we obtain the posterior over $\boldsymbol{\pi}$ as

$$p\left(\boldsymbol{\pi}|\sim\right) \propto \prod_{\boldsymbol{i}} \mathrm{Multinomial}\left(\boldsymbol{z}_{\boldsymbol{i}}|\boldsymbol{\pi}\right) \mathrm{Dir}\left(\boldsymbol{\pi}|\boldsymbol{\alpha}_0\right). \tag{50}$$

Thus, $\boldsymbol{\pi}$ can be drawn from a Dirichlet distribution as follows:

$$\boldsymbol{\pi} \sim \mathrm{Dir}\left(\boldsymbol{\pi}|\tilde{\boldsymbol{\alpha}}\right), \tag{51}$$

where $\tilde{\boldsymbol{\alpha}} = (\tilde{\alpha}_1, ..., \tilde{\alpha}_D)$ with entry $\tilde{\alpha}_d = \sum_{\boldsymbol{i}} z_{\boldsymbol{i}}^d + \alpha_{0d}$.

C. Algorithm, complexity and convergence

With the above samplers, the entire Gibbs sampling flow is summarized in Algorithm 1 where N_b, N_s are the iteration number for burn-in and samples collection, respectively.

Algorithm 1: Adaptive low-rank tensor completion

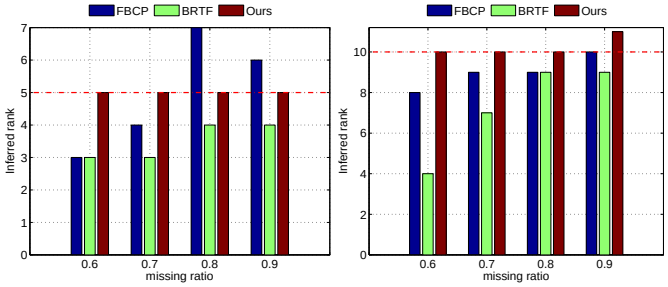
Input: Observation $\mathcal{Y}_\Omega$, parameters $a_0, b_0, \mu_0, \beta_0, \boldsymbol{\alpha}_0, \tau_0, D$.
Initialization: $\mathcal{U}, \boldsymbol{\lambda}, \boldsymbol{\tau}, \boldsymbol{\tau}_e$ are initialized by $\mathbf{1}$ with proper size, $\boldsymbol{\mu}, \boldsymbol{\mu}_e$ are initialized by $\mathbf{0}$, $\boldsymbol{\pi} = \left(D^{-1}, \cdots, D^{-1}\right)^T$ and $\boldsymbol{z_i} \sim \text{Multinomial}\left(\boldsymbol{z_i}|\boldsymbol{\pi}\right)$.
for $t \leftarrow 1$ ***to*** $N_b + N_s$ **do**
  Sample the low-rank structure:
  1. Sample λ_r, γ_r and κ_r as Eqs. (18) (22) (27);
  2. Sample $u_{ir}^{(k)}$, $\mu^{(k)}$ and $\tau_d^{(k)}$ as Eqs. (30) (33) (36);
  Sample the non-low-rank structure:
  1. Sample $e_{\boldsymbol{i}}$, μ_d and τ_d as Eqs. (39) (42) (45);
  2. Sample $\boldsymbol{z_i}$ and $\boldsymbol{\pi}$ as Eqs. (48) (51);
  if $t > N_b$ **then**
    Collect samples for $\boldsymbol{\lambda}, \mathcal{U}, \mathcal{E}$;
4. Get the sample mean $\bar{\boldsymbol{\lambda}}, \bar{\mathcal{U}}, \bar{\mathcal{E}}$ across N_s samples;
5. Complete the latent tensor $\hat{\mathcal{L}}$ with $\bar{\boldsymbol{\lambda}}, \bar{\mathcal{U}}, \bar{\mathcal{E}}$ as Eqs. (5) (3).

In each Gibbs sampling iteration, we draw samples for all unknown variables using corresponding samplers. Note that per-iteration sampling complexity is dominated by $O(RKN)$, which is only linear in the number N of entries in the latent tensor $\mathcal{L}$. Moreover, it has been demonstrated that MMSE often converges well [49] and more evidence will be provided in Section V-A.

V. Experiments

We evaluate the proposed model comprehensively in extensive experiments on synthetic and real visual tensor data. The experiments on synthetic data are conducted to validate the following five aspects of the proposed model: i) automatic tensor rank determination; ii) the ability to fit a wide range of non-low-rank structures $\mathcal{E}$s; iii) the effectiveness of the adaptive low-rank tensor model; iv) convergence; v) recovery performance. In addition, three other practical applications, i.e., image inpainting, video completion and facial image synthesis, are performed to further evaluate the recovery performance of the proposed model. To this end, the proposed model is compared with 8 state-of-the-art low rank tensor completion methods, including FaLRTC [1], HaLRTC [1], RPTC$_{\text{scad}}$ [26], TMac [38], STDC [21], t-SVD [25], FBCP [22] and BRTF [27]. Note that only BRTF considers a sparse non-low-rank structure $\mathcal{E}$, while the others adopt a zero $\mathcal{E}$.

For the proposed model, we set all hyperparameters in a non-informative manner to reduce their influence on the posterior as much as possible [33], [47]. Throughout the experiments, $\mu_0 = 0$, and $\alpha_{01}, \cdots, \alpha_{0D}$, β_0, a_0 and b_0 are 10^{-6}. In addition, we fix the number D of mixture components in $\mathcal{E}$ as 3 for simplicity. Of course, it also can be determined automatically by the tuning scheme proposed in [33] with simply setting a sufficiently large initialized D. The initialized tensor rank R is set 20 and 100 for the synthetic data and real visual data, respectively. The comparison methods are implemented by the codes of authors with tuned parameters for the best performance. For simplicity, we term the proposed model 'Ours' in the following tables and figures.



(a) Rank = 5 (b) Rank = 10

Fig. 3. Tensor rank determined by different methods on synthetic data. The vertical color bars denote the inferred tensor rank, while the horizontal red dash lines indicate the true tensor rank.

TABLE II
RRE of the proposed method and its variants on synthetic tensor with different missing ratios.

	rank = 5			rank = 10		
Method	70%	80%	90%	70%	80%	90%
Ours_nx(6)	-*	-	-	-	-	-
Ours_nx(10)	-	-	-	-	-	-
Ours_nx(20)	-	-	-	-	-	-
Ours_nx(50)	-	-	-	-	-	-
Ours_ne	0.2234	0.2242	0.2426	0.2210	0.2372	0.2702
Ours	**0.1829**	**0.1960**	**0.2038**	**0.1796**	**0.1914**	**0.2055**

*'-' denotes recovery failure.

A. Tensor completion on synthetic data

We generate the latent tensor $\mathcal{L}$ of size $30 \times 30 \times 30$ based on Eq. (3). The low-rank structure $\mathcal{X}$ of $\text{rank}\left(\mathcal{X}\right) = 5$ is generated via CP factorization in Eq. (5) with factor matrices $\boldsymbol{U}^{(k)} \in \mathbb{R}^{30\times 5}$s and weights vector $\boldsymbol{\lambda} \in \mathbb{R}^{5\times 1}$. The entries in $\boldsymbol{U}^{(k)}$ are sampled independently from $\mathcal{N}\left(0, 1\right)$, while weights in $\boldsymbol{\lambda}$ are sampled uniformly in range of $(0, 2]$. To simulate a wide range of residual components, the following 5 types of $\mathcal{E}$ are generated: (1) zero non-low-rank structure with all zero entries; (2) Gaussian non-low-rank structure with all entries sampled from $\mathcal{N}\left(0, 0.01\right)$; (3) sparse non-low-rank structure with 10% of entries sampled uniformly from range of $[-2, 2]$; (4) mixture non-low-rank structure (zero mean) with 10% of entries sampled uniformly in range of $[-2, 2]$, 30% of entries sampled from $\mathcal{N}\left(0, 0.1\right)$ and 60% of entries sampled from $\mathcal{N}\left(0, 0.005\right)$; (5) mixture non-low-rank structure (non-zero mean) with 10% of entries sampled uniformly in range of $[-1, 4]$, 20% of entries sampled from $\mathcal{N}\left(0.1, 0.1\right)$ and 70% of entries sampled from $\mathcal{N}\left(-0.1, 1/300\right)$. Additional 5 simulations for $\mathcal{L}$ are generated in a similar manner, however, $\text{rank}\left(\mathcal{X}\right) = 10$. To obtain the incomplete observation $\mathcal{Y}_\Omega$, we first add noise $\mathcal{M}$ with entries sampled from $\mathcal{N}\left(0, 0.001\right)$ on $\mathcal{L}$. Then, a certain percentage (i.e., the missing ratio) of entries are selected randomly from the noisy $\mathcal{L}$ and set to zeros. Here we choose the missing ratio from 70% to 90%.

TABLE I
QUANTITATIVE COMPARISON BETWEEN PARAMETERS IN TRUE PDFS OF $\mathcal{E}$ (DENOTED 'TRUE') AND THAT IN ESTIMATED ONES (DENOTE 'EST.').

rank = 5		Zero	Sparse		Gaussian	Mixture (zero mean)			Mixture (non-zero mean)		
		Comp.1	Comp.1	Comp.2	Comp.1	Comp.1	Comp.2	Comp.3	Comp.1	Comp.2	Comp.3
π_d	True	-	0.1	0.9	-	0.1	0.3	0.6	0.1	0.2	0.7
	Est.	-	0.102	0.898	-	0.150	0.282	0.568	0.065	0.250	0.685
μ_d	True	0	0	0	0.1	0	0	0	1.5	0.1	-0.1
	Est.	-2e-4	-0.015	4e-4	0.101	-0.017	-0.012	1e-4	2.429	0.050	-0.1
τ_d	True	1e12	0.75	1e12	100	0.75	10	200	0.48	10	300
	Est.	6.64e4	0.7081	4.13e5	100.1	1.072	13.582	220.8	1.050	7.086	314.6
rank = 10		Zero	Sparse		Gaussian	Mixture (zero mean)			Mixture (non-zero mean)		
		Comp.1	Comp.1	Comp.2	Comp.1	Comp.1	Comp.2	Comp.3	Comp.1	Comp.2	Comp.3
π_d	True	-	0.1	0.9	-	0.1	0.3	0.6	0.1	0.2	0.7
	Est.	-	0.101	0.899	-	0.149	0.281	0.570	0.062	0.239	0.699
μ_d	True	0	0	0	0.1	0	0	0	1.5	0.1	-0.1
	Est.	-2e-4	0.032	0.000	0.100	0.020	4e-4	-6e-4	2.611	0.094	-0.100
τ_d	True	1e12	0.75	1e12	100	0.75	10	200	0.48	10	300
	Est.	6.16e5	0.728	9.12e4	98.76	1.005	15.70	220.6	1.391	6.392	313.9

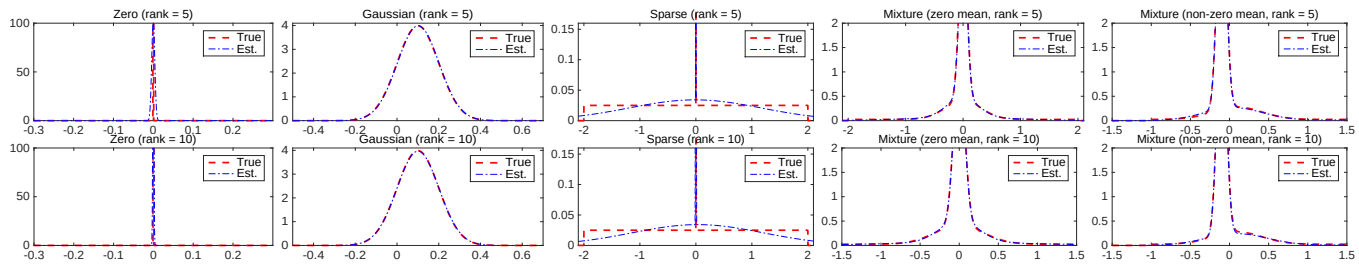
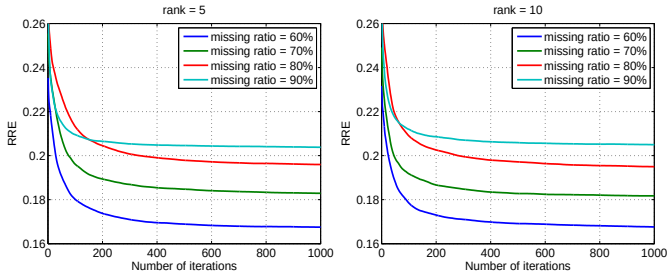


Fig. 4. Visual comparison between true PDFs of $\mathcal{E}$ (denoted 'True') and those estimated ones (denote 'Est.').



(a) Rank = 5 (b) Rank = 10

Fig. 5. RRE curves versus the iteration number of Gibbs sampling on two synthetic tensors with different missing ratios.

1) Tensor rank determination: We evaluate the proposed model on 2 different synthetic tensors that consist of two low rank $\mathcal{X}$s of $\text{rank}\,(\mathcal{X}) = 5, 10$ but a same mixture non-low-rank structure $\mathcal{E}$ with non-zero mean. Given $\mathcal{Y}_\Omega$ with different missing ratios, we infer the sparse weights vector $\boldsymbol{\lambda}$, and give the inferred tensor rank as the number of weights $|\lambda_r| > 10\text{e-}5$. Since most competitors cannot infer the tensor rank, only the inferred ranks from the proposed model, FBCP and BRTF are shown in Figure 3[2]. As can be seen, when $\text{rank}\,(\mathcal{X}) = 5$, the proposed model infers the tensor rank exactly even when 90% of entries are missing, while both FBCP and BRTF fail to estimate the real tensor rank accurately. When $\text{rank}\,(\mathcal{X}) = 10$, the proposed model only misses the true rank when the missing ratio is 90%, while FBCP and BRTF miss the true one in most cases. This is primary due to the newly defined tensor rank and the sparsity-induced low-rank model adopted. In addition, modeling the non-low-rank structure benefits separating the low-rank structure exactly, thereby determining the accurate tensor rank. Therefore, we can conclude that the proposed model can accurately estimate the tensor rank, even if the tensor data contain complex non-low-rank structure with numerous missing entries.

2) Ability to fit a wide range of $\mathcal{E}$s: Given $\mathcal{Y}_\Omega$ with 70% of entries missing, the proposed model can estimate the probability density function (PDF) (i.e., the distribution) of $\mathcal{E}$ by inferring the posterior mean of parameters $\boldsymbol{\mu}_e$, $\boldsymbol{\tau}_e$ and $\boldsymbol{\pi}$ in tensor completion. A quantitative comparison of the estimated parameters in 5 different PDFs of $\mathcal{E}$ and the corresponding ground truth are given in Table I. As shown, the proposed model produces accurate estimations for these parameters, viz., the proposed model can well fit $\mathcal{E}$ with different PDFs. To clarify this point, we plot the estimated PDFs of $\mathcal{E}$ and the corresponding ground truth in Figure 4. As can be seen, the estimated PDFs comply well with the ground truth, especially in the complicated mixture $\mathcal{E}$ cases. In addition, the proposed model performs stably with different low-rank structures (e.g., $\text{rank}(\mathcal{X}) = 5, 10$ in Table I and Figure 4) in terms of fitting $\mathcal{E}$. Therefore, we conclude that the proposed model can fit a wide range of $\mathcal{E}$, which is consistent with the theoretical analysis in Remark 2.

3) Effectiveness of adaptive low-rank tensor model: In the proposed adaptive low-rank tensor model, the low-rank and the complex non-low-rank structures are modeled separately. To evaluate the effectiveness of separate components, we compare the proposed model with two variants which we refer to as 'Ous_nx' and 'Ours_ne'. In 'Ours_nx', we remove the low-rank structure $\mathcal{X}$ and model the entire tensor with MOG,

[2]More comparisons can be found in the supplementary materials.

TABLE III
RRE ON SYNTHETIC TENSORS WITH DIFFERENT NON-LOW-RANK STRUCTURES AND MISSING RATIOS.

Zero non-low-rank structure $\mathcal{E}$						
	rank = 5			rank = 10		
Method	70%	80%	90%	70%	80%	90%
FaLRTC [1]	0.0248	0.1990	0.6942	0.4386	0.7057	0.9105
HaLRTC [1]	0.0218	0.1974	0.6934	0.4383	0.7056	0.9105
$\text{RPTC}_{\text{scad}}$ [26]	0.0063	0.1948	0.8367	0.5245	0.7485	0.9168
TMac [38]	0.8354	0.8917	0.9498	0.8352	0.8920	0.9479
STDC [21]	0.1694	0.3685	0.8714	0.5585	0.8377	0.9965
t-SVD [25]	0.3874	0.6242	0.8687	0.6224	0.7924	0.9498
FBCP [22]	**0.0027**	**0.0032**	0.1589	0.0404	0.0417	0.3928
BRTF [27]	0.1279	0.1379	0.1509	0.0352	0.1393	0.4719
Ours	0.0086	0.0041	**0.0047**	**0.0036**	**0.0047**	**0.0074**
Gaussian non-low-rank structure $\mathcal{E}$						
	rank = 5			rank = 10		
Method	70%	80%	90%	70%	80%	90%
FaLRTC [1]	0.0750	0.2170	0.6864	0.4944	0.7159	0.9067
HaLRTC [1]	0.0754	0.2179	0.6857	0.4938	0.7154	0.9068
$\text{RPTC}_{\text{scad}}$ [26]	0.0297	0.2025	0.8258	0.5564	0.7545	0.9143
TMac [38]	0.8355	0.8916	0.9483	0.8450	0.8935	0.9470
STDC [21]	0.1624	0.3947	0.8797	0.5682	0.8330	-*
t-SVD [25]	0.4044	0.6117	0.8734	0.6439	0.7907	0.9526
FBCP [22]	0.0462	0.0335	0.1663	0.0612	0.0634	0.3104
BRTF [27]	0.0889	0.1437	0.1587	0.1338	0.1449	0.4732
Ours	**0.0285**	**0.0307**	**0.0351**	**0.0292**	**0.0320**	**0.0389**
Mixture non-low-rank structure $\mathcal{E}$ (on-zero mean)						
	rank = 5			rank = 10		
Method	70%	80%	90%	70%	80%	90%
FaLRTC [1]	0.2902	0.4363	0.7834	0.5494	0.7413	0.9240
HaLRTC [1]	0.2901	0.4362	0.7842	0.5495	0.7415	0.9240
$\text{RPTC}_{\text{scad}}$ [26]	0.2903	0.4906	0.8711	0.5907	0.7704	0.9288
TMac [38]	0.8340	0.9006	0.9522	0.8374	0.8923	0.9511
STDC [21]	0.3233	0.5527	0.8995	0.6512	0.8688	0.9423
t-SVD [25]	0.5158	0.7116	0.9245	0.6759	0.8110	0.9754
FBCP [22]	0.2199	0.2416	0.2746	0.2247	0.2316	0.2996
BRTF [27]	0.1976	0.2326	0.2547	0.1837	0.2340	0.5642
Ours	**0.1829**	**0.1960**	**0.2038**	**0.1796**	**0.1914**	**0.2055**

*'-' denotes recovery failure.

while 'Ours_ne' sets the no-low-rank structure $\mathcal{E} = 0$. In the experiments, we employ these three methods to recover two synthetic tensors with different missing ratios, which consists of two respective $\mathcal{X}$s of $\text{rank}(\mathcal{X}) = 5, 10$ and a non-zero mean mixture $\mathcal{E}$. Since 'Ours_nx' only utilizes the MOG to model the entire tensor, different Ds (i.e., the number of Gaussian components) are adopted to well fit the complex tensor structure. The recovery performance is measured by the relative reconstruction error (RRE), i.e., $\|\mathcal{L}-\hat{\mathcal{L}}\|_F/\|\mathcal{L}\|_F$, where $\hat{\mathcal{L}}$ is the estimation of the true $\mathcal{L}$. The comparison results are presented in Table II.

Without modeling the non-low-rank structure, the proposed model outperforms 'Ours_ne' obviously. This demonstrates that modeling the non-low-rank structure separately can improve the performance. In addition, Ours_nx fails to recover the tensor in most cases. The reason is intuitive. To solve the ill-posed tensor completion problem, we must well represent the low-rank structure as regularization. However, the MOG is too flexible to regularize the ill-posed problem and thus causes completion failure, especially when the tensor is highly structured (i.e., the tensor rank = 5). Thus, we model the low-rank structure by exploiting the sparsity in the CP factorization, which is crucial for tensor completion. Therefore, we can conclude that the proposed data-adaptive tensor model which exploits the low-rank-structure and the non-low-rank structure separately, is effective in tensor completion.

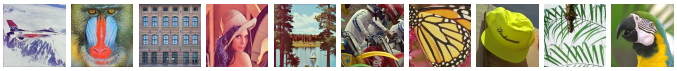


Fig. 6. Ground truth of 10 benchmark images.

4) Validation of convergence: It has been proven that samples generated by Gibbs sampling coverage to a stable probabilistic distribution [47]. To illustrate this point more intuitively, using the two synthetic tensors adopted in Section V-A3, we plot the RRE curves of the proposed model versus the Gibbs sampling iteration number with different missing ratios, as shown in Figure 5. The plots indicate that the proposed model converges in hundreds of iterations in different cases.

5) Evaluation of recovery performance: We compare the proposed model with 8 existing tensor completion methods. To ensure a comprehensive comparison, we conduct experiments on tensors consisting of a low-rank $\mathbf{X}$ with rank = $5, 10$ as well as three different types of non-low-rank $\mathcal{E}$: i) a zero non-low-rank $\mathcal{E}$; ii) a Gaussian non-low-rank $\mathcal{E}$; iii) a mixture non-low-rank $\mathcal{E}$ (non-zero-mean). The recovery results are reported in Table III. We find that the proposed method outperforms other comparison methods in most cases. Moreover, its superiority is further enhanced with the increasing missing ratio or a simpler $\mathcal{E}$, e.g., when $\mathcal{E}$ comes from mixture distribution and the missing ratio is 90%, the improvement over the second best method, i.e., FBCP, is up to 0.09 in RRE when the tensor rank is 10. When $\mathcal{E}$ is set to zero without any other changes, the improvement is up to 0.38. In addition, compared with other methods, the proposed method performs more stably with the changing missing ratios. For example, when $\mathcal{E} = 0$ and the tensor rank is 10, the RRE of the proposed model is less than 0.01. In contrast, the RRE of FaLRTC increases from 0.4 to 0.9. Based on those observed results, we can conclude that the proposed model outperforms existing methods on the synthetic tensor data.

6) Analysis of computational cost: The runtime required for all methods to complete the tensor $\mathcal{L}$ with a non-zero-mean mixture $\mathcal{E}$, when the rank is 5 and the missing ratio is 80%, are listed in Table IV. All methods are run using MATLAB software on a Laptop with 4GB RAM and 4 i5 CPU cores. Since Gibbs sampling requires extensive iterations for convergence, the proposed method consumes more time than the other methods. However, note that this study primarily focuses on validating the theoretical results. In addition, the proposed method can be further accelerated via using some common implementation techniques, e.g., the parallel programming or the variational Bayesian (VB) technique. In Table IV, the last column shows the runtime of a rough VB version of the proposed method. As can be seen, implementing VB increases the speed of the proposed method by nearly 4 times with only a slight performance drops (e.g., RRE increases from 0.1960 up to 0.1987).

TABLE IV

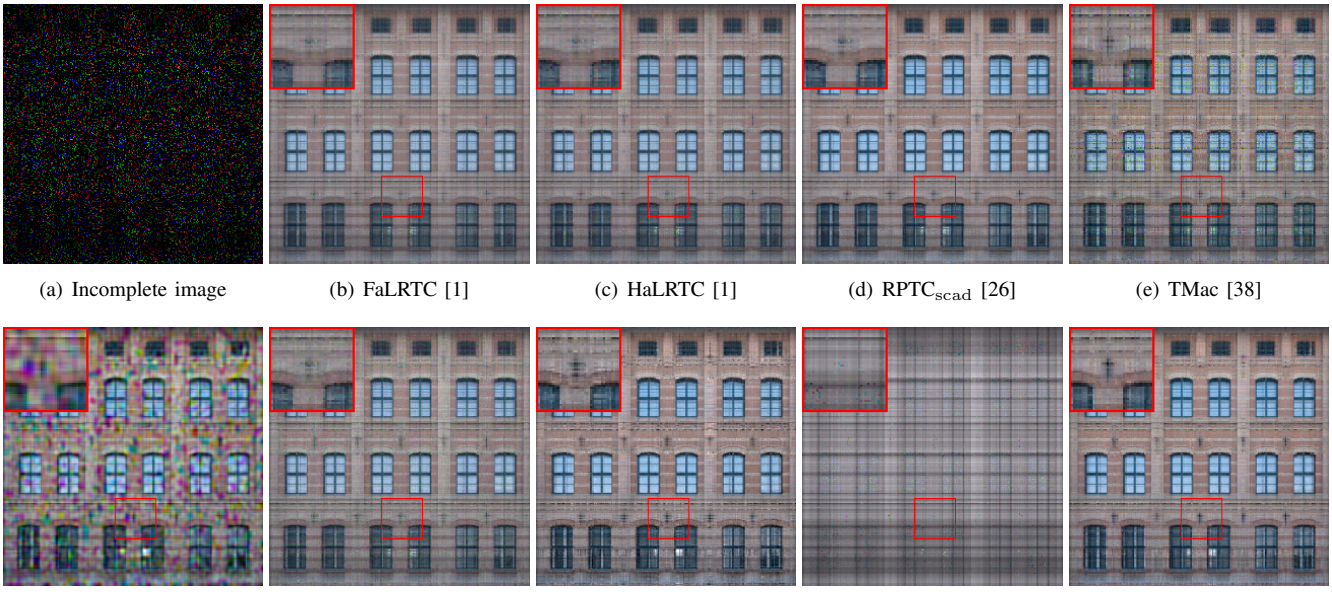
RUNTIME OF ALL METHODS ON SYNTHETIC DATA OF RANK = 5 WITH A NON-ZERO-MEAN MIXTURE $\mathcal{E}$ AND 80% MISSING RATE.

Method	FaLRTC [1]	HaLRTC [1]	$\text{RPTC}_{\text{scad}}$ [26]	TMac [38]	STDC [21]	t-SVD [25]	FBCP [22]	BRTF [27]	Ours	Ours (VB)
Time(s)	3.935	7.581	12.36	2.087	1.419	5.445	13.77	21.84	213.5	52.46

TABLE V

AVERAGE RRE, PSNR AND SSIM ON 10 BENCHMARK IMAGES WITH DIFFERENT MISSING RATIOS.

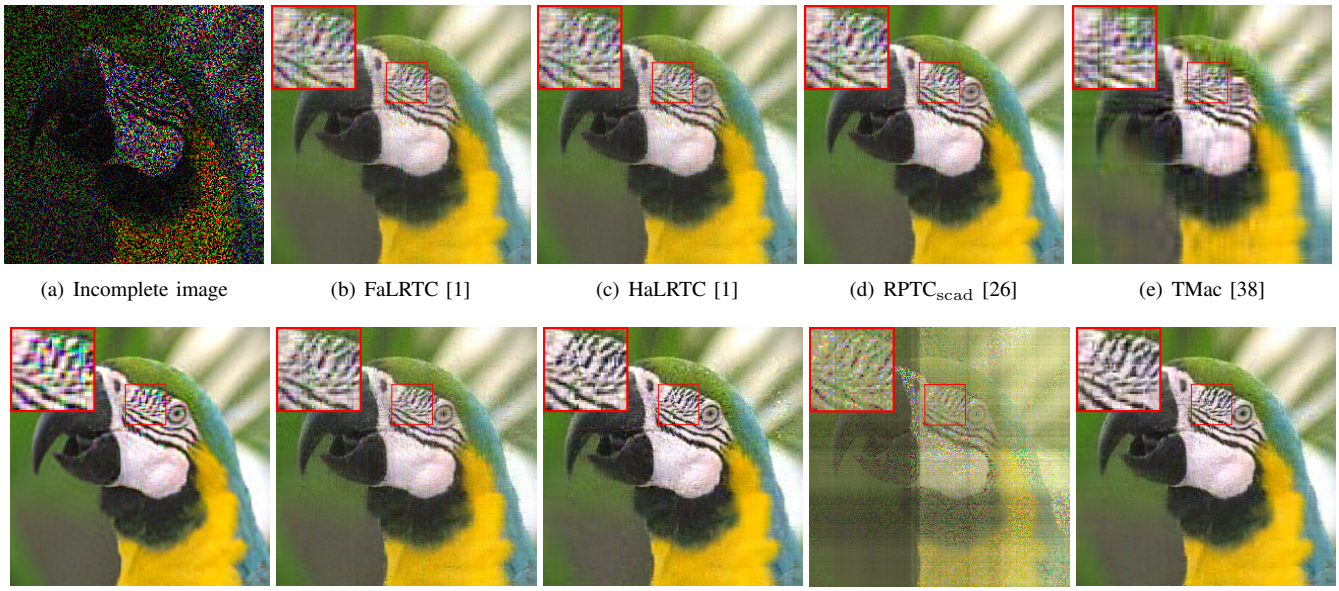
Method	60%			70%			80%			90%		
	RRE	PSNR	SSIM	RRE	PSNR	SSIM	RRE	PSNR	SSIM	RRE	PSNR	SSIM
FaLRTC [1]	0.1003	25.6337	0.7911	0.1289	23.4559	0.7024	0.1704	21.0291	0.5797	0.2445	17.8378	0.4068
HaLRTC [1]	0.0995	25.7348	0.7916	0.1281	23.5329	0.7010	0.1695	21.0939	0.5757	0.2430	17.9096	0.3997
$\text{RPTC}_{\text{scad}}$ [26]	0.0859	27.0793	0.8232	0.1336	24.1529	0.7106	0.1582	21.9985	0.6022	0.2210	18.9062	0.4198
TMac [38]	0.1408	22.7919	0.6465	0.1517	22.1323	0.6036	0.1700	21.1231	0.5375	0.2940	16.1311	0.2627
STDC [21]	0.0893	26.5850	0.8089	0.1078	25.0486	0.7579	0.1367	22.9630	**0.6801**	0.2326	17.9273	**0.4913**
t-SVD [25]	0.0915	26.4223	0.7832	0.1219	23.9397	0.6826	0.1649	21.3169	0.5472	0.2391	18.0432	0.3580
FBCP [22]	0.0916	26.5030	0.7601	0.1137	24.5926	0.6866	0.1508	22.1263	0.5727	0.2192	18.8112	0.3878
BRTF [27]	0.2685	16.7725	0.4421	0.2892	16.1276	0.4018	0.3089	15.5556	0.3611	0.3291	15.0054	0.3108
Ours	**0.0755**	**28.0223**	**0.8381**	**0.0966**	**25.9144**	**0.7676**	**0.1302**	**23.3615**	0.6549	**0.1970**	**19.8088**	0.4549



(a) Incomplete image (b) FaLRTC [1] (c) HaLRTC [1] (d) $\text{RPTC}_{\text{scad}}$ [26] (e) TMac [38]

(f) STDC [21] (g) t-SVD [25] (h) FBCP [22] (i) BRTF [27] (j) Ours

Fig. 7. Visual results of the 'facade' image, when missing ratio is 90%.



(a) Incomplete image (b) FaLRTC [1] (c) HaLRTC [1] (d) $\text{RPTC}_{\text{scad}}$ [26] (e) TMac [38]

(f) STDC [21] (g) t-SVD [25] (h) FBCP [22] (i) BRTF [27] (j) Ours

Fig. 8. Visual results for the 'parrot' image from the top four methods, when missing ratio is 70%.

(a) Incomplete frame (b) $\text{RPTC}_{\text{scad}}$ [26] (c) STDC [21] (d) FBCP [22] (e) Ours

Fig. 9. Visual results for the 11-th frame in the 'suzie' video [5].

(a) Incomplete frame (b) $\text{RPTC}_{\text{scad}}$ [26] (c) STDC [21] (d) FBCP [22] (e) Ours

Fig. 10. Visual results for the 19-th frame in the 'foreman' video [5].

TABLE VI
RRE, PSNR AND SSIM ON TWO VIDEOS.

Method	suzie			foreman		
	RRE	PSNR	SSIM	RRE	PSNR	SSIM
FaLRTC [1]	0.0746	29.7584	0.7108	0.0537	29.1385	0.7484
HaLRTC [1]	0.0746	29.7585	0.7108	0.0537	29.1385	0.7484
$\text{RPTC}_{\text{scad}}$ [26]	0.0726	29.9943	0.7158	0.0493	29.8815	0.7635
TMac [38]	0.0615	31.4253	0.8588	0.0702	26.8153	0.8103
STDC [21]	0.0747	29.7401	0.7091	0.0543	29.0494	0.7431
FBCP [22]	0.0426	34.6221	0.9103	0.0570	28.6269	0.8205
BRTF [27]	0.0951	27.6432	0.6967	0.0975	23.9687	0.7117
Ours	**0.0381**	**35.6013**	**0.9182**	**0.0422**	**31.2435**	**0.8512**

TABLE VII
RRE ON CMU-PIE DATASET WITH DIFFERENT MISSING RATIOS.

Method	60%	70%	80%	90%
FaLRTC [1]	0.3741	0.5058	0.7392	0.9103
HaLRTC [1]	0.3694	0.5021	0.7388	0.9086
$\text{RPTC}_{\text{scad}}$ [26]	0.2228	0.2599	0.3284	0.4166
TMac [38]	0.2629	0.4226	0.6574	0.9385
STDC [21]	0.2602	0.2606	0.3316	0.5154
FBCP [22]	0.1410	0.1911	0.2590	0.3746
Ours	**0.1172**	**0.1560**	**0.2274**	**0.3372**

B. Image inpainting

We test the proposed model on 10 benchmark images (see Figure 6) for image inpainting. Each image is of size $256 \times 256 \times 3$ and rescaled to $[0, 1]$. We generate the incomplete observation by randomly selecting a certain percentage of missing entries in each image. Given the incomplete observation, all methods are utilized to recover the latent image. In addition to RRE, peak signal-to-noise ratio (PSNR) and structural similarity (SSIM), two conventional image quality indices are also adopted to measure the recovery accuracy. Average RRE, PSNR and SSIM on 10 benchmark images are given in Table V. As can be seen, the proposed model outperforms other methods in most cases. To clarify these results, visual comparison of the recovery results for the 'facade' image (e.g., the missing ratio is 90%) as well as the 'parrot' image (e.g., the missing ratio is 70%) are shown in Figure 7 and Figure 8, respectively. In each image, an area of interest is enlarged to facilitate a detailed comparison. Because the 'facade' image is highly structured, most methods can recover the major structure of the image with even 90% missing entries. However, compared with the proposed method, most of the competing methods fail to well recover the fine details, e.g., the zoomed areas in Figure 8 (b)-(i). In contrast, these details are well recovered in the results of the proposed model, as shown in the enlarged area in Figure 8 (j). The superior performance of the proposed model can be attributed to the following. First, we propose an appropriate sparsity-induced low-rank model for tensor data, which differs completely from the matrix rank norm induced low-rank model in FaLRTC, HaLRTC and TMac. Second, we adopt a flexible MOG to model the complex non-low-rank structure, which is not considered in other competing methods. These can be further validated by the results for the 'Parrot' image shown in Figure 8. As can be seen, the 'Parrot' image contains a complex non-low-rank structure, and the proposed model provides clearer and more natural results than other methods, particularly for enlarged details.

C. Video completion

We also evaluate the proposed model on two famous videos, 'suzie' and 'foreman'[3] for video completion. For each video, 20 consecutive frames are extracted as experimental data of size $144 \times 176 \times 3 \times 20$, and then rescaled to $[0, 1]$. The incomplete observation for video 'susie' is a scrabbled version[4], as shown in Figure 9 (b); and the incomplete observation for video 'foreman' is corrupted by superimposed texts[4], as shown in Figure 10 (b). The observations are also corrupted by noise

[3]https://media.xiph.org/video/derf/

[4]For every two consecutive frames in the video, there is one corrupted with the same scrabbles or superimposed texts.

(a) Ground truth (b) STDC [21] (c) FBCP [22] (d) Ours

Fig. 11. Visual results for the 49 selected missing faces, when missing ratio is 70% [5].

sampled from $\mathcal{N}(0, 0.001)$. Since t-SVD cannot be applied to 4-mode tensor [25], here we only employ other 8 methods to recover the latent video. The RRE, PSNR and SSIM for these two videos are give in Table VI. The proposed model returns higher PSNR, SSIM and lower RRE than other methods, which demonstrates the superiority of the proposed method over others for video completion. To further clarify these results, visual comparison results of two corrupted frames are shown in Figure 9, 10. Compared with other methods, the proposed model produces sharper and clearer results.

D. Facial image synthesis

Motivated by the fact that a complete training set is often not available in actual facial recognition applications [51], [52], facial image synthesis is studied to recover missing faces in an incomplete dataset. In this subsection, we use the CMU-PIE face dataset [53] for facial image synthesis. Specifically, we select images from the first 3 subjects with 11 positions and 21 illumination changes to construct a 4-mode tensor $\mathcal{L} \in \mathbb{R}^{3\times 11\times 21\times 1024}$, which is then rescaled to $[0, 1]$. For the incomplete observation, a certain percentage of missing faces are randomly selected. Given the observation, all methods except t-SVD and BRTF [6] are employed to recover $\mathcal{L}$. Table VII gives the RRE for each method. As can be seen, the proposed model yields the lowest RRE with different missing ratios. Visual comparison results for 49 selected missing faces (e.g., the missing ratio is 70%) are shown in Figure 11. Compared with other competing methods, the proposed model recovers more details and produces less artifacts.

VI. Discussion

In this section, we will consider two remaining problems with the proposed model, including noise estimation and spatial similarity constraint in visual tensors.

A. Noise estimation

In the above discussion, we assume the noise precision τ_0 is predefined; however, this will limit the proposed model in the cases corrupted by noise with unknown precision. To address this problem, we assume precision τ_0 to follow a Gamma distribution as follows:

$$\tau_0 \sim \mathrm{Ga}(\tau_0 | a_0, b_0). \tag{52}$$

Then, it can be inferred with Gibbs sampling as Section IV and drawn from a Gamma distribution as

$$\tau_0 \sim \mathrm{Ga}(\tilde{a}_{\tau_0}, \tilde{b}_{\tau_0}) \tag{53}$$

with parameters

$$\begin{aligned} \tilde{a}_{\tau_0} &= a_0 + (\textstyle\sum_{\boldsymbol{i}} \mathcal{O}_{\boldsymbol{i}})/2; \\ \tilde{b}_{\tau_0} &= b_0 + \frac{1}{2}\textstyle\sum_{\boldsymbol{i}} \mathcal{O}_{\boldsymbol{i}}(y_{\boldsymbol{i}} - x_{\boldsymbol{i}} - e_{\boldsymbol{i}})^2; \end{aligned} \tag{54}$$

To demonstrate the effectiveness of the noise estimation discussed above, we compare the proposed model (i.e., given the true noise precision τ_0) with its variant termed 'Ours_est' which estimates the noise with the sampler in Eq. (53). The experiment is conducted on the synthetic tensor which consists of low rank structure of rank = 5 and a non-zero-mean mixture non-low-rank structure as Section V-A. Gaussian white noise with different precisions are added to the observation. With such noisy observations, the recovery results of the proposed model and 'Ours_est' are given in Table VIII. As can been seen, the proposed model only slightly outperforms 'Ours_est'. Therefore, we can conclude that the proposed model armed with the sampler in Eq. (53) can well mitigate the effect of noise corruption on tensor completion.

TABLE VIII
RRE on synthetic tensor with different noise corruptions.

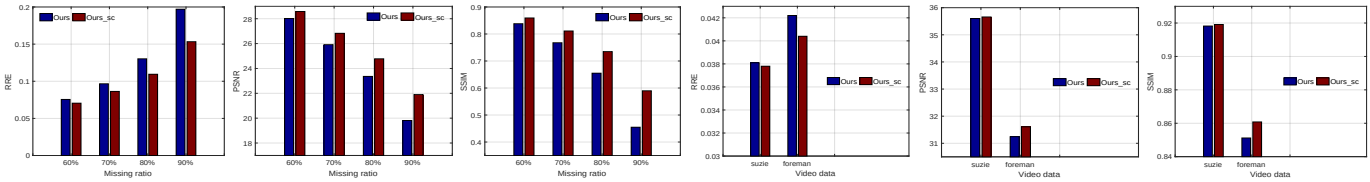
τ_0	1000	500	200	100	10	1
Ours	**0.2038**	**0.2068**	**0.2066**	**0.2068**	**0.2199**	**0.3293**
Ours_est	0.2068	0.2086	0.2093	0.2097	0.3072	0.3403

B. Spatial similarity constraint

Here we consider the spatial similarity constraint in visual tensors. Specifically, visual tensor data (e.g., image or video) often shows similarity in the spatial domain. Such similarity can provide extra prior information for the ill-posed completion task, thereby improving the recovery accuracy especially when the missing ratio is high. In general, such similarity can be modeled by exploiting the correlation among different rows

[5] Complete results for all methods can be found in the supplementary material.

[6] BRTF always fails on this dataset as rank becomes 0.



(a) RRE (b) PSNR (c) SSIM (d) RRE (e) PSNR (f) SSIM

Fig. 12. Performance comparison between the proposed method and its variant that considers the spatial similarity in two applications. (a)-(c) image inpainting results; (d)-(f) video completion results.

of the factor matrices $\boldsymbol{U}^{(k)}$. For example, if we consider a color image as 3-mode tensor, the correlation among different rows in each $\boldsymbol{U}^{(k)}$ represents the correlation among rows, columns and channels of the image. Considering the obviously local similarity in visual data, we assume neighboring rows in each $\boldsymbol{U}^{(k)}$ to be similar and introduce the following prior distribution

$$p\left(u_{ir}^{(k)}\right) \sim \mathcal{N}\left(u_{ir}^{(k)}|\mu^{(k)}, \tau^{(k)-1}\right)\mathcal{N}\left(u_{ir}^{(k)}|\boldsymbol{w}_i^T\boldsymbol{u}_r^{(k)}, \eta_0^{-1}\right) \tag{55}$$

where $\boldsymbol{w}_i = [w_{1i}, \cdots, w_{n_k i}]^T$ is a weight vector with $\sum_{j=1}^{n_k} w_{ji} = 1$ and $w_{ii} = 0$. This prior suggests that each row of $\boldsymbol{U}^{(k)}$ can be reconstructed by other rows, and also complies with the ℓ_2 norm constraint on each entry. To this end, a sufficiently large η_0 is adopted, e.g., 10^3. In addition, we set $w_{ji} = \exp\left(-2 * \rho * |i-j|^2\right)$ to imply the major contribution of neighboring rows in the reconstruction. The parameter $\rho = Nz/N$, where N_z denotes the number of observed entries in $\mathcal{Y}$. Thus, $1-\rho$ equals the missing ratio. With this prior, $u_{ir}^{(k)}$ thus can be drawn from a Gaussian distribution $\mathcal{N}\left(\tilde{\mu}_{u_{ir}^{(k)}}, \tilde{\tau}_{u_{ir}^{(k)}}\right)$ as Eq. (30) with parameters

$$\begin{aligned}\tilde{\tau}_{u_{ir}^{(k)}} &= \sum\nolimits_{\boldsymbol{i}:i_k=i} \tau_0 \mathcal{O}_{\boldsymbol{i}} \tilde{c}_{\boldsymbol{i}}^{rk2} + \tau^{(k)} + \eta_0, \\ \tilde{\mu}_{u_{ir}^{(k)}} &= \tau_{u_{ir}^{(k)}}^{-1}\left(\sum\nolimits_{\boldsymbol{i}:i_k=i} \tau_0 \mathcal{O}_{\boldsymbol{i}} \tilde{y}_{\boldsymbol{i}}^r \tilde{c}_{\boldsymbol{i}}^{rk} + \tau^{(k)}\mu^{(k)} + \eta_0 \boldsymbol{w}_i^T \boldsymbol{u}_r^{(k)}\right).\end{aligned} \tag{56}$$

To demonstrate the effectiveness of the prior in Eq. (55), we compare the proposed model with its variant 'Ours_sc' where $u_{ir}^{(k)}$ is sampled with Eq. (56), in above image inpainting and video completion applications. The numerical results are compared in Figure 12. As can be seen, 'Ours_sc' outperforms the proposed model in both applications. Moreover, its superior performance is more obvious when the missing ratio is high. Intuitively, this is because a high missing ratio results in a worsened ill-posed problem that requires more prior information to regularize the infinite solution space.

VII. Conclusion

We have presented an adaptive low-rank representation model for tensor completion. The model explicitly decomposes the latent tensor into the low-rank structure and the non-low-rank structure in a Bayesian way. The low-rank prior relies on a new formulation of CP rank, which forms the basis of automatic tensor rank determination by exploiting sparsity in CP factorization from an incomplete set of observations. The prior for the non-low-rank structure is a mixture of Gaussians which has shown to be sufficiently flexible to reflect a variety of real tensor data. These two priors allow the development of an MMSE method to estimate the posterior mean of missing entries and their uncertainty using Gibbs sampling. In addition, the proposed model has been shown to outperform other existing models in terms of tensor recovery.

References

[1] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 208–220, 2013.

[2] W. Wei, L. Zhang, Y. Jiao, C. Tian, C. Wang, and Y. Zhang, "Intracluster structured low-rank matrix analysis method for hyperspectral denoising," *IEEE Transactions on Geoscience and Remote Sensing*, no. 99, pp. 1–15, 2018.

[3] R. Dian, S. Li, and L. Fang, "Learning a low tensor-train rank representation for hyperspectral image super-resolution," *IEEE transactions on neural networks and learning systems*, 2019.

[4] S.-J. Wang, J. Yang, M.-F. Sun, X.-J. Peng, M.-M. Sun, and C.-G. Zhou, "Sparse tensor discriminant color space for face verification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 6, pp. 876–888, 2012.

[5] G.-S. Xie, X.-Y. Zhang, S. Yan, and C.-L. Liu, "Hybrid cnn and dictionary-based models for scene recognition and domain adaptation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 6, pp. 1263–1274, 2015.

[6] Y. Fu, J. Gao, D. Tien, Z. Lin, and X. Hong, "Tensor lrr and sparse coding-based subspace clustering," *IEEE transactions on neural networks and learning systems*, vol. 27, no. 10, pp. 2120–2133, 2016.

[7] L. Zhang, P. Wang, C. Shen, L. Liu, W. Wei, Y. Zhang, and A. v. d. Hengel, "Adaptive importance learning for improving lightweight image super-resolution network," *arXiv preprint arXiv:1806.01576*, 2018.

[8] S. Yang, M. Wang, Z. Feng, Z. Liu, and R. Li, "Deep sparse tensor filtering network for synthetic aperture radar images classification," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 8, pp. 3919–3924, 2018.

[9] G.-S. Xie, X.-Y. Zhang, S. Yan, and C.-L. Liu, "Sde: A novel selective, discriminative and equalizing feature representation for visual recognition," *International Journal of Computer Vision*, vol. 124, no. 2, pp. 145–168, 2017.

[10] Y. Gao, X. Wang, Y. Cheng, and Z. J. Wang, "Dimensionality reduction for hyperspectral data based on class-aware tensor neighborhood graph and patch alignment," *IEEE transactions on neural networks and learning systems*, vol. 26, no. 8, pp. 1582–1593, 2014.

[11] X. Li, M. K. Ng, G. Cong, Y. Ye, and Q. Wu, "Mr-ntd: Manifold regularization nonnegative tucker decomposition for tensor data dimension reduction and representation," *IEEE transactions on neural networks and learning systems*, vol. 28, no. 8, pp. 1787–1800, 2016.

[12] C.-Y. Ko, C. Chen, Z. He, Y. Zhang, K. Batselier, and N. Wong, "Deep model compression and inference speedup of sum-product networks on tensor trains," *IEEE transactions on neural networks and learning systems*, 2019.

[13] H. Huang and H. Yu, "Ltnn: A layerwise tensorized compression of multilayer neural network," *IEEE transactions on neural networks and learning systems*, 2019.

[14] E. Papalexakis, K. Pelechrinis, and C. Faloutsos, "Spotting misbehaviors in location-based social networks using tensors," in *Proceedings of the 23rd International Conference on World Wide Web*. ACM, 2014, pp. 551–552.

[15] L. Yao, Q. Z. Sheng, Y. Qin, X. Wang, A. Shemshadi, and Q. He, "Context-aware point-of-interest recommendation using tensor factorization with social regularization," in *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 2015, pp. 1007–1010.

[16] X. Guo and Y. Ma, "Generalized tensor total variation minimization for visual data recovery?" in *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.* IEEE, 2015, pp. 3603–3611.

[17] L. Zhang, W. Wei, C. Bai, Y. Gao, and Y. Zhang, "Exploiting clustering manifold structure for hyperspectral imagery super-resolution," *IEEE Transactions on Image Processing*, vol. 27, no. 12, pp. 5969–5982, 2018.

[18] N. Wang, X. Gao, L. Sun, and J. Li, "Anchored neighborhood index for face sketch synthesis," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 9, pp. 2154–2163, 2018.

[19] N. Wang, X. Gao, and J. Li, "Random sampling for fast face sketch synthesis," *Pattern Recognition*, vol. 76, pp. 215–227, 2018.

[20] M. Zhang, N. Wang, Y. Li, and X. Gao, "Deep latent low-rank representation for face sketch synthesis," *IEEE transactions on neural networks and learning systems*, 2019.

[21] Y.-L. Chen, C.-T. Hsu, and H.-Y. M. Liao, "Simultaneous tensor decomposition and completion using factor priors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 3, pp. 577–591, 2014.

[22] Q. Zhao, L. Zhang, and A. Cichocki, "Bayesian cp factorization of incomplete tensors with automatic rank determination," *PAMI*, vol. 37, no. 9, pp. 1751–1763, 2015.

[23] B. W. Bader, T. G. Kolda *et al.*, "Matlab tensor toolbox version 2.6," Available online, February 2015. [Online]. Available: http://www.sandia.gov/ tgkolda/TensorToolbox/

[24] Q. Gu, H. Gui, and J. Han, "Robust tensor decomposition with gross corruption," in *Proc. Advances in Neural Inf. Process. Syst.*, 2014, pp. 1422–1430.

[25] Z. Zhang, G. Ely, S. Aeron, N. Hao, and M. Kilmer, "Novel methods for multilinear data completion and de-noising based on tensor-svd," in *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, 2014, pp. 3842–3849.

[26] Q. Zhao, D. Meng, X. Kong, Q. Xie, W. Cao, Y. Wang, and Z. Xu, "A novel sparsity measure for tensor recovery," in *Proc. IEEE Int. Conf. Comp. Vis.*, 2015, pp. 271–279.

[27] Q. Zhao, G. Zhou, L. Zhang, A. Cichocki, and S.-I. Amari, "Bayesian robust tensor factorization for incomplete multiway data," *IEEE Trans. Neural Netw. & Learn. Syst.*, vol. 27, no. 4, pp. 736–748, 2016.

[28] D. Goldfarb and Z. Qin, "Robust low-rank tensor recovery: Models and algorithms," *SIAM Journal on Matrix Analysis and Applications*, vol. 35, no. 1, pp. 225–253, 2014.

[29] L. Zhang, W. Wei, Y. Zhang, C. Shen, A. van den Hengel, and Q. Shi, "Cluster sparsity field: An internal hyperspectral imagery prior for reconstruction," *International Journal of Computer Vision*, vol. 126, no. 8, pp. 797–821, 2018.

[30] Q. Xie, Q. Zhao, D. Meng, and Z. Xu, "Kronecker-basis-representation based tensor sparsity and its applications to tensor recovery," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 8, pp. 1888–1902, 2018.

[31] Y. Liu, F. Shang, W. Fan, J. Cheng, and H. Cheng, "Generalized higher order orthogonal iteration for tensor learning and decomposition," *IEEE transactions on neural networks and learning systems*, vol. 27, no. 12, pp. 2551–2563, 2015.

[32] Y. Yang, Y. Feng, and J. A. Suykens, "Robust low-rank tensor recovery with regularized redescending m-estimator," *IEEE transactions on neural networks and learning systems*, vol. 27, no. 9, pp. 1933–1946, 2015.

[33] Q. Zhao, D. Meng, Z. Xu, W. Zuo, and L. Zhang, "Robust principal component analysis with complex noise." in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 55–63.

[34] X. Chen, Z. Han, Y. Wang, Q. Zhao, D. Meng, and Y. Tang, "Robust tensor factorization with unknown noise," in *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, 2016, pp. 5213–5221.

[35] U. Schmidt and S. Roth, "Shrinkage fields for effective image restoration," in *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, 2014, pp. 2774–2781.

[36] L. Liu, C. Shen, L. Wang, A. Van Den Hengel, and C. Wang, "Encoding high dimensional local features by sparse coding based fisher vectors," in *Advances in Neural Information Processing Systems*, 2014, pp. 1143–1151.

[37] X. Cao, Y. Chen, Q. Zhao, D. Meng, Y. Wang, D. Wang, and Z. Xu, "Low-rank matrix factorization under general mixture noise distributions," in *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, 2015, pp. 1493–1501.

[38] Y. Xu, R. Hao, W. Yin, and Z. Su, "Parallel matrix factorization for low-rank tensor completion," *arXiv preprint arXiv:1312.1254*, 2013.

[39] W. Hu, D. Tao, W. Zhang, Y. Xie, and Y. Yang, "The twist tensor nuclear norm for video completion," *IEEE transactions on neural networks and learning systems*, vol. 28, no. 12, pp. 2961–2973, 2016.

[40] X. Zhang, "A nonconvex relaxation approach to low-rank tensor completion," *IEEE transactions on neural networks and learning systems*, vol. 30, no. 6, pp. 1659–1671, 2018.

[41] Z. Xu, F. Yan, and A. Qi, "Infinite tucker decomposition: Nonparametric bayesian models for multiway data analysis," in *Proc. Int. Conf. Mach. Learn.*, 2012, pp. 1023–1030.

[42] H. Yong, D. Meng, W. Zuo, and L. Zhang, "Robust online matrix factorization for dynamic background subtraction," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 7, pp. 1726–1740, 2018.

[43] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.

[44] L. Zhang, W. Wei, C. Tian, F. Li, and Y. Zhang, "Exploring structured sparsity by a reweighted laplace prior for hyperspectral compressive sensing," *IEEE Transactions on Image Processing*, vol. 25, no. 10, pp. 4974–4988, 2016.

[45] E. J. Candes, M. B. Wakin, and S. Boyd, "Enhancing sparsity by reweighted l1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, 2007.

[46] L. Zhang, W. Wei, Y. Zhang, C. Tian, and F. Li, "Reweighted laplace prior based hyperspectral compressive sensing for unknown sparsity," in *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, 2015, pp. 2274–2281.

[47] C. M. Bishop *et al.*, "Pattern recognition and machine learning, vol. 1springer," *New York*, no. 4, p. 12, 2006.

[48] H. Ishwaran and J. S. Rao, "Spike and slab variable selection: frequentist and bayesian strategies," *Annals of Statistics*, pp. 730–773, 2005.

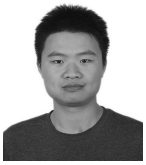
[49] U. Schmidt, Q. Gao, and S. Roth, "A generative perspective on mrfs in low-level vision," in *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.* IEEE, 2010, pp. 1751–1758.

[50] M. Kreh, "Bessel functions," *Lecture Notes, Penn State-Göttingen Summer School on Number Theory*, vol. 82, 2012.

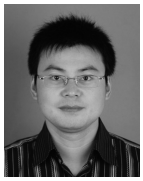
[51] X. Geng, K. Smith-Miles, Z.-H. Zhou, and L. Wang, "Face image modeling by multilinear subspace analysis with missing values," *IEEE Trans. Syst., Man, Cybern. B*, vol. 41, no. 3, pp. 881–892, 2011.

[52] N. Wang, X. Gao, L. Sun, and J. Li, "Bayesian face sketch synthesis," *IEEE Transactions on Image Processing*, vol. 26, no. 3, pp. 1264–1274, 2017.

[53] T. Sim, S. Baker, and M. Bsat, "The cmu pose, illumination, and expression database," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1615–1618, 2003.



Lei Zhang (M18) received the Ph.D. degree in computer science and technology from Northwestern Polytechnical University, Xian, in 2018. He was a research staff in the school of computer science,the University of Adelaide, Australia. He is currently a research scientist in the Inception Institute of Artificial Intelligence (IIAI), Abu Dhabi, United Arab Emirates. His research interests include image processing, machine learning and video analysis.



Wei Wei (SM'18-) received the Ph.D. degree from Northwestern Polytechnical University, Xian, China, in 2012. He is currently an Associate Professor with the School of Computer Science and Engineering, Northwestern Polytechnical University. His research interests include image processing, machine learning and pattern recognition. He has been authored more than 40 papers including IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE TRANSACTIONS ON IMAGE PROCESSING, PATTEN RECOGNITION, CVPR, ICCV, ECCV, AAAI, IJCAI, etc. He is a Reviewer for IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE GEOSCIENCE REMOTE SENSING LETTERS, and IEEE TRANSACTIONS ON NEURAL NETWORK AND LEARNING SYSTEM, etc. He served as the PC member for around 10 major international conference including CVPR, ICME, etc.

Qinfeng Shi is a Senior Lecturer at the School of Computer Science, The University of Adelaide. His interest includes Machine Learning, Computer Vision, and Compressive Sensing, particularly Structured Learning and Probabilistic Graphical Models. He was an ARC Discovery Early Career Researcher Award (DECRA) Fellow between 2012-2014. He received a PhD in computer science in 2011 at The Australian National University (ANU) in Machine Learning. Before that, he completed Bachelor and Master study in computer science and Technology in 2003 and 2006 at The Northwestern Polytechnical University (NPU).

Chunhua Shen is a Professor at School of Computer Science, The University of Adelaide, a Project Leader and Chief Investigator at the Australian Research Council Centre of Excellence for Robotic Vision (ACRV). His research interests are in the intersection of computer vision and statistical machine learning. Recent work has been on large-scale image retrieval and classification, object detection and pixel labelling using deep learning. He received the Australian Research Council Future Fellowship in 2012. He is serving as Associate Editor of IEEE Transactions on Neural Networks and Learning Systems.

Anton van den Hengel is the Director of the Australian Centre for Visual Technologies, the Program Lead for the Analytics and Decision Support Program of the Data 2 Decisions CRC, and a Professor of Computer Science at the University of Adelaide. Prof. van den Hengel has published over 200 papers, been a CI on over $50m in research funding, and leads a group of over 60 researchers working in Computer Vision and Machine Learning. Prof. van den Hengel has received a number of awards, including the Pearcey Award for Innovation and the CVPR Best Paper Award in 2010.

Yanning Zhang (SM12-) received the B.S. degree from the Dalian University of Technology, Dalian, China, in 1988, the M.S. degree from the School of Electronic Engineering, Northwestern Polytechnical University, Xian, China, in 1993, and the Ph.D. degree from the School of Marine Engineering, Northwestern Polytechnical University, Xian, China, in 1996. She is currently a Professor with the School of Computer Science, Northwestern Polytechnical University. She is also a Cheung Kong Professor of Ministry of Education, China. She has authored more than 200 papers. Her current research interests include remote sensing image analysis, computer vision and pattern recognition and etc. She is the Associate Editor of IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.