埼玉工業大学

博士後期学位論文

テンソル分解算法及び画像処理応用に関する研究

陳啓鵬

埼玉工業大学大学院　　博士後期課程

工学研究科　　電子工学専攻

指導教員　　曹 建庭 教授

令和 4 年 2 月 07 日

SAITAMA INSTITUTE OF TECHNOLOGY

DOCTORAL THESIS

Study on Tensor Decomposition Algorithms and Its Application to Image Processing

---

Author:
Qipeng CHEN

Supervisor:
Jianting CAO

A thesis submitted in fulfillment of the requirements for the degree of

Doctor of Philosophy

in the

Graduate School of Engineering

February 24, 2022

論文概要

多視点センサーやデータストレージ技術の発展に伴い、取得されたデータの次元や複雑度が増している。これらのデータを伝統的な方法で処理すると、コンピューターへの負担が増え、データ処理の効率も悪くなる。これらのデータをいかに効率的に処理するかは重要な研究である。テンソルは、行列とベクトルを一般化したもので、データの高次の関係や内容を自然に表現することができる。近年、テンソル法は高次元データを処理するための強力なツールとなっている。テンソル法は、信号処理、機械学習、データマイニングなど多くの研究分野で応用されている。

テンソル法の中で、テンソル分解算法は最も重要で基本的な方法の一つであり、テンソルを低次元潜在因子のセットに分解することである。潜在因子は、データの潜在的な特徴を明らかにし、圧縮性の高い方法でデータを表現する強力なものである。CANDECOMP/PARAFAC(CP)分解とタッカー分解は、1 世紀以上にわたって研究されてきた最も有用なテンソル分解モデルである。近年、テンソルトレイン(TT)分解が提案された。TT 分解は、CP 分解やタッカー分解と比べて、計算の利便性が高い。

今、テンソル分解算法を用いて画像処理への応用を中心に研究している。主な貢献は、テンソル分解に基づいてデータ処理の効率と性能が高い様々なアルゴリズムを提案した。まず、データ復元の問題に着目し、データ復元に TT とトータルバリエーション(TV)制約を課すことによって、よい性能を発揮することができる。TT-TV モデルを解くため、新たなアプローチを提案した。提案手法は、TT ランクに核ノルム正則化を導入した。テンソルコアへの初期化・更新の必要はない。 次に、ブラックボックス攻撃に関する研究を行なった。機械学習(ML)モデルが日常生活でますます重要な役割を果たしているので、ブラックボックス攻撃をもう一つの研究対象として選んだ。提案手法は、原画像をテンソル特異値分解（t-SVD）で分解し、ノイズテンソルを特異値テンソルに加算または減算する。Google Cloud Vision API を含むいくつかのニューラルネットワークに攻撃を与えて、提案手法の有効性と効率性を実証した。本論文の研究は、テンソル分解法への研究とその応用を充実させ、テンソル法に貢献しており、研究や産業の分野に良い参考となっている。

Abstract

Study on Tensor Decomposition Algorithms and Its Application to Image Processing

by Qipeng CHEN

Doctor of Philosophy

SAITAMA INSTITUTE OF TECHNOLOGY

Graduate School of Engineering

With the development of multi-view sensors and data storage technology, the dimension and complexity of the acquired data is getting higher. Processing these data by traditional methods will not only increase the burden on the computer, but also reduce the efficiency of data processing.How to efficiently process these data is a vital problem to be solved. Tensor is the generalization of matrix and vector, which can naturally represent high-order relations and objects of the data. In recent years, tensor methods have become powerful tools to process high-dimensional data. Numerous applications of tensor methods have been applied in signal processing, machine learning, data mining,etc.

Among the tensor methods, tensor decomposition is one of the most important and fundamental tools, which is to decompose a tensor into a set of latent factors of low dimensionality. The latent factors are powerful to reveal the latent feature of the data and represent the data in a highly compressive way. CANDECOMP/PARAFAC decomposition (CPD) and Tucker decomposition (TKD) are the most classical tensor decomposition models which have been studied for over a century. In recent years, TT decomposition has been proposed. Compared with traditional CP decomposition and Tucker decomposition, TT decomposition has good calculation convenience.

The research is focusing on tensor decomposition algorithms and application to image processing. The the main contribution is to propose various algorithms to increase the efficiency and performance of data processing via tensor method. Firstly, aiming at the problem of data recovery, imposing tensor train (TT) and total variation (TV) constraint on data completion can produce impressive performance, we propose a new approach to solve TT-TV model. The nuclear norm regularization on TT-ranks is introduced in our method and our solution does not need to initialize and update tensor cores. Secondly, we choose black-box attack as another research object as machine learning(ML) models are playing an increasingly important role in daily life. The method decompose the original image by Tensor Singular Value decomposition(t-SVD), the noise tensor is either add or subtract it to the Singular value tensor. We demonstrate the efficacy and efficiency of the proposed method by fooling some widely used neural networks including Google Cloud Vision API. The work in the thesis has enriched the theoretical study and applications of tensor, which contribute to the tensor methodology and will be a good reference in the research and industry fields.

Acknowledgements

First and foremost, I would like to give my sincere gratitude to my supervisor, Professor Jianting Cao, who always has gave full support to my research career during my PhD study. He patiently guided me and gave me advice whenever I faced research difficulties. Before I started the PHD study, professor Cao gave me a lot of suggestions on how to involving the study of tensor. On the first-year of my PhD study, he encouraged me to work on Quasi-brain-death EEG Diagnosis by Tensor Train Decomposition, which is a new research direction for me. I followed his guidance and tried to apply the tensor method to classify the Quasi-brain-death patients and death patients. At the first-year of study, my work was accepted and I had an opportunity to attend my first international conference in Moscow, which had totally opened my eyes and intrigued my passion of research. Before the presentation, he taught me how to prepare the PowerPoint and how to give an academic presentation fluently. After that, I have the confidence to continue the research work of tensor.

I also would like to thank Dr. Qibin Zhao who is an excellent researcher and an impressive mentor for me. I am very appreciate for giving me the opportunity to study in his team. Under his supervision, I have made great progress in tensor knowledge and machining learning. The days in RIKEN AIP I improved my academic writing, and learned how to reply the question to reviewer and these are invaluable experiences for me.

Last but not least, I would like to thank all my friends and families, who have helped me a lot and gave me support during these years.

At last, I want to thank Professor Yamazaki, Professor Hashimoto, Professor Ohyama for their comments concerning my thesis. Those comments are valuable and very helpful.

Contents

Chapter 1

Introduction

1.1 Background

Both matrices and vectors can be considered as tensors. Vectors are one-dimensional tensors, matrices are tensors with 2 dimensions. When the number of dimension is more than 3, we call it the high-dimensional tensor. Color images, videos, and the multi-channel electroencephalogram are tensors. Grey image is a 2-dimensional data(height $\times$ width), colourful image is a 3-dimensional data(height $\times$ width$\times$ RGB channel), video is a 4-dimensional data(height $\times$ width $\times$ RGB channel $\times$ time) and electroencephalography (EEG) signals is a 3-dimensional data(magnitude $\times$ trails $\times$ time). How to process these high-dimensional data becomes a vital problem for us. The traditional methods usually transform tensor to matrices or vectors, but it will lead to spatial redundancy and less efficient factorization.

Tensor can keep the original high-order data form, and it can maintain more spatial information in data processing [1]. With the rapid development of computer communication and network technology, it is necessary to store, process, and analyze the data with a larger scale, higher-dimensional, and more complex structure. Among the various tensor methods, tensor decomposition is the most important tools of them. The purpose of tensor decomposition is to find the latent factors of the data (i.e. the generalization of multi-dimensional arrays), to represent a high-dimensional data by a series of low-dimensional data. The decomposed factors can also be considered as latent features of the original data. There are some types of tensor decomposition and they have different specific form and operations among latent low-dimensional tensors. Some of these decomposition models are widely applied in different fields such as machine learning [2–4] and signal processing [5,6]. Tucker decomposition (TKD) and CANDECOMP/PARAFAC decomposition (CPD) are classical tensor decomposition models, which have been studied for nearly half a century [1,7,8]. In recent years, Tensor Train(TT) decomposition has been proposed. Compared to traditional CP decomposition and Tucker decomposition, TT decomposition has good calculation convenience, and it scales linearly to the tensor order.

The thesis is studying on image processing by tensor method. With the development of internet and sensor technology, many industries enjoys the convenience of high-quality of images and videos. For example internet shopping, urban traffic management, social networks, and intelligent production. Our purpose is to process these high-quality and high-dimensional data efficiently. $Chapter$ 1 firstly introduces the contributions of this thesis. Then the background of tensor, some basic tensor decomposition models and the tensor completion method we utilized in our research. In $Chapter$ 2, We will introduce these methods in detail which is applied in our research, including tensor train decomposition model, tensor train(TT) rank, tensor singular value decomposition, and some representation of tensor calculation. in $Chapter$ 3, we present a new method to minimize TT rank with a total variation model, and visual data tensorization (VDT) is introduced in this paper to reshape the magnetic resonance imaging(MRI) data to enhance the performance of the proposed algorithm. In $Chapter$ 4, we propose a simple and effective black-box attack method. The original image is divided into two orthogonal tensors and one rectangular diagonal tensor by Tensor Singular Value decomposition(t-SVD). $Chapter$ 5 provides the overall conclusion of the thesis and our future work.

1.2 Summary of contributions

1.2.1 TT rank with TV for MRI data reconstruction

As a common medical diagnostic method, magnetic resonance imaging (MRI) is wildly applied to hospitals for patients. MRI utilizes magnetic resonance to obtain electromagnetic signals, thus forming the images of body's physiological process, and it is applicable to almost all kinds of diseases, including tumor, inflammation, and trauma. However, the drawback of applying MRI diagnosis is the long the scanning time, and the whole progress may last from more than ten minutes to even an hour. Therefore, patients have to stay completely still during the scan process, which is difficult to diagnose some patients who do not cooperate such as children or babies. If the whole process can be finished in a shorter time, the latency time for patients will be reduce, thereby improving the efficiency of hospitals. So it is necessary to propose a method to reconstruct MRI images in a shorter time [9, 10].

Data completion methods have been applied in MRI to decrease data acquisition time and remove the artifacts in the image. We can reconstruct the unsampled MRI data by observing the sampled experimental data through some mathematical tools such as parametric modeling and phase constraint [11]. By analyzing the relationship of acquired data through this method, the unknown data can be predicted and sampling time can be reduced as well. Phase constrained completion is a common data reconstruction method. Firstly, it transforms the data by Fourier transform and then reconstructs the data by Fourier symmetry of phase information. LORAKS [12] proposed a phase constraint based on single-channel MRI data completion and analyze the relationship between phase constraints in partial Fourier reconstruction by data reconstruction. These methods are based on the matrix structure. It is proved that the tensor is an attractive and promising tool for the representation and processing of MRI data [5]. In our method, we process MRI data by applying the tensor method which can capture more inner structure information.

In research, the model by imposing the low-rank minimization has been proved to be effective for magnetic resonance imaging (MRI) completion. Recent studies have also shown that imposing tensor train (TT) and total variation (TV) constraint on tensor completion can produce impressive performance, the lower TT-rank minimization constraint can be represented as the guarantee for global constraint, while the total variation as the guarantee for regional constraint. In our solution, a new approach is utilized to solve TT-TV model. In contrast with imposing the alternating linear scheme, nuclear norm regularization on TT-ranks is introduced in our method as it is an effective surrogate for rank optimization and our solution does not need to initialize and update tensor cores. By applying alternating direction method of multipliers (ADMM), the optimization model is disassembled into some sub-problems, singular value thresholding can be used as the solution to the first sub-problem and soft thresholding can be used as the solution to the second sub-problem. The new optimization algorithm ensures the effectiveness of data recovery. In addition, a new method is introduced to reshape the MRI data to a higher-dimensional tensor, so as to enhance the performance of data completion. Furthermore, the method is compared with some other methods including tensor reconstruction methods and a matrix reconstruction method. It is concluded that the proposed method has a better recovery accuracy than others in MRI data according to the experiment results.

1.2.2 Black-box adversarial attack by T-svd

Machine learning(ML) plays an essential role in our daily life and ML classifiers are used in many fields to do the work of classification. For instance, a credit card fraud detector is a classifier taking the user's credit card transactions as inputs and identify which transactions are performed by the user and which are not. However, the safety of the model becomes an important topic for consideration. Adversarial attacks is to add a small perturbation to the input to misclassify the result and it is proved that the output of neural networks can be affected by small perturbation [13] [14]. There are two kinds of adversarial attacks, the white-box technology requires the attacker to know complete information about the target model, but there is no such restriction on black-box technology and it modified the perturbation according to the output of the previous query [15].

It seems that the output of most image classification models can be changed by white-box attacks [16] and the result indicates that after learning by ML classifiers, these image data are going to be close to decision boundaries. The white-box attack is an effective method to attack the target model because the attacker possesses the model's information, including its parameter values setting and training methods, etc. The white-box attack can be guided effectively with gradient descent [13] [17] and tend to have high query efficiency than black-box attack(the search for successful ResNet/ImageNet attacks require on the order of $10^4 - 10^5$ queries). but in most scenarios, it is impossible to acquire the information of the model. Hence black-box attack is more applicable for attackers [18] [19]. The number of queries is a vital indicator of the efficiency of the attack algorithm. A low number of queries means less money and time costs for adversarial attacks. It is necessary to propose a query-efficient black-box attacks method.

Unlike the white-box attack, the black-box attack is practical to construct the adversarial images. In this research, the proposed method utilizes the following simple iterative principle: we decompose the original image by Tensor Singular Value decomposition(t-SVD), the noise tensor is randomly picked from pre-specified set and then either add or subtract it to the Singular value tensor which is a rectangular diagonal data and its size is same as the original image but with much fewer value, therefore our method significantly reduces the query cost. From the experiment result, We demonstrate the efficacy and efficiency of the proposed method by fooling some widely used neural networks including Google Cloud Vision API.

Chapter 2

Tensor decomposition models

2.1 Tensor preliminaries

2.1.1 Notations

Notations in [1] are adopted in this thesis. A scalar is denoted by a normal lowercase/uppercase letter, e.g., $x, X \in \mathbb{R}$, a vector is denoted by a boldface lowercase letter, e.g., $\mathbf{x} \in \mathbb{R}^{I}$, a matrix is denoted by a boldface capital letter, e.g., $\mathbf{X} \in \mathbb{R}^{I \times J}$, a tensor of order $N \geq 3$ is denoted by an Euler script letter, e.g., $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$.

In addition, the Frobenius norm of $\mathcal{X}$ can be represented by $\|\mathcal{X}\|_F = \sqrt{\langle \mathcal{X}, \mathcal{X} \rangle}$, and $\langle \mathcal{X}, \mathcal{X} \rangle$ represents inner product. The nuclear norm of $\mathbf{X}$ can be represented by $\|\mathbf{X}\|_*$ and it is the sum of singular values of $\mathbf{X}$. A tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ and its element can be represented by $\mathcal{X}_{(i_1, i_2, \ldots, i_N)}$ with index $(i_1, i_2, \ldots, i_N)$. Moreover, we are going to introduce two kinds of tensor unfolding methods in our paper. One is the standard mode-n unfolding [1], which is represented as $\mathbf{X}_{(n)} \in \mathbb{R}^{I_n \times I_1 \cdots I_{n-1} I_{n+1} \cdots I_N}$, and another mode-n unfolding is represented as $\mathbf{X}_{[n]} \in \mathbb{R}^{I_1 \cdots I_n \times I_{n+1} \cdots I_N}$.

2.1.2 CP decomposition and Tucker decomposition

CP decomposition. CPD decomposes a tensor into a sum of rank-one tensors. For a tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, it decomposes the tensor as follows:

$$\boldsymbol{\mathcal{X}} = \sum_{r=1}^{R} \vec{a}_r^{(1)} \circ \vec{a}_r^{(2)} \circ \cdots \vec{a}_r^{(N)}, \tag{2.1}$$

where $\circ$ is the out product, and $\mathbf{A}^{(n)} = [\vec{a}_1^{(n)}, \vec{a}_2^{(n)}, \ldots, \vec{a}_R^{(n)}]$ is the CP factors.

Tucker decomposition. Tucker decomposition approximates a tensor by a core tensor and several factor matrices as follow:

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{G}} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \times \cdots \times_n \mathbf{A}^{(n)}, \tag{2.2}$$

where $\boldsymbol{\mathcal{G}}$ is the core tensor, and $[\mathbf{A}]$ are the factor matrices.

2.1.3 Tensor train decomposition

Tensor train decomposition (TTD) is to decompose a tensor into a sequence of two matrices and $N-2$ order-three core tensors (factor tensors): $\mathbf{G}^{(1)}, \boldsymbol{\mathcal{G}}^{(2)}, \cdots, \mathbf{G}^{(N)}$. The relation between the approximated tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ and core tensors can be expressed as follow:

$$\boldsymbol{\mathcal{X}} = \ll \mathbf{G}^{(1)}, \boldsymbol{\mathcal{G}}^{(2)}, \cdots, \mathbf{G}^{(N)} \gg, \tag{2.3}$$

where for $n = 1, \cdots, N$, $\boldsymbol{\mathcal{G}}^{(n)} \in \mathbb{R}^{R_{n-1} \times I_n \times R_n}$, $R_0 = R_N = 1$, and the notation $\ll \cdot \gg$ is the operation to transform the core tensors to the approximated tensor. $\mathbf{G}^{(1)} \in \mathbb{R}^{I_1 \times R_1}$ and $\mathbf{G}^{(N)} \in \mathbb{R}^{R_{N-1} \times I_N}$ are two matrices in the first and the last positions. The sequence $R_0, R_1, \cdots, R_N$ is named TT-rank which limits the size of every core tensor. Furthermore, the $(i_1, i_2, \cdots, i_N)$th element of tensor $\boldsymbol{\mathcal{X}}$ can be represented by the multiple product of the corresponding mode-2 slices of the core tensors as:

$$x_{i_1 i_2 \cdots i_N} = \prod_{n=1}^{N} \mathbf{G}_{i_n}^{(n)}, \tag{2.4}$$

where $\mathbf{g}_{i_1}^{(1)}, \mathbf{G}_{i_1}^{(1)}, \cdots, \mathbf{g}_{i_N}^{(N)}$ is the sequence of slices from each core tensor. For $n = 1, 2, \cdots, N$, $\mathbf{G}_{i_n}^{(n)} \in \mathbb{R}^{R_{n-1} \times R_n}$ is the mode-2 slice extracted from $\boldsymbol{\mathcal{G}}^{(n)}$ according to each mode of the element index of $x_{i_1 i_2 \cdots i_N}$. $\mathbf{g}_{i_1}^{(1)} \in \mathbb{R}^{R_1}$ and $\mathbf{g}_{i_N}^{(N)} \in \mathbb{R}^{R_{N-1}}$ are extracted from first core tensor and last core tensor, they are considered as two order-one matrices for overall expression convenience.

2.1.4 Tensor Singular value decomposition

For a 3-dimensional tensor, in order to keep its adjacent structure information for data, we introduce the tensor method to process the image data [20] [21]. Tensor methods have been applied more and more widely in the field of image processing. In the paper, the t-product $*$ is introduced to tensor calculation. The t-product of $\mathcal{A} \in \mathbf{R}^{n_1 \times n_2 \times n_3}$, and $\mathcal{B} \in \mathbf{R}^{n_2 \times n_4 \times n_3}$ is a tensor $\mathcal{C} \in \mathbf{R}^{n_1 \times n_4 \times n_3}$ is given by:

$$\mathcal{C} = \mathcal{A} * \mathcal{B} = \mathbf{Fold}(\mathbf{Circ}(\mathcal{A}) \times \mathbf{Vec}(\mathcal{B})), \tag{2.5}$$

where $\mathbf{Fold}()$ is an operation that takes $\mathbf{Vec}(\mathcal{B})$ into tensor $\mathcal{B}$ and it can be described as:

$$\mathbf{Vec}(\mathcal{B}) = \begin{bmatrix} \mathcal{B}^{(1)} \\ \mathcal{B}^{(2)} \\ \cdots \\ \mathcal{B}^{(n_3)} \end{bmatrix} \tag{2.6}$$

and $\mathbf{Circ}()$ is described as:

$$\mathbf{Circ}(\mathcal{A}) = \begin{bmatrix} \mathcal{A}^{(1)} & \mathcal{A}^{(n_3)} & \cdots & \mathcal{A}^{(n_3-1)} & \mathcal{A}^{(2)} \\ \mathcal{A}^{(2)} & \mathcal{A}^{(1)} & \mathcal{A}^{(n_3)} & \cdots & \mathcal{A}^{(3)} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathcal{A}^{(n_3)} & \mathcal{A}^{(n_3-1)} & \cdots & \mathcal{A}^{(2)} & \mathcal{A}^{(1)} \end{bmatrix} \tag{2.7}$$

Chapter 3

TT rank with TV for MRI data reconstruction

3.1 Preliminaries

3.1.1 Proposed method introduction

In this research, we present a new method to minimize tensor train(TT) rank with total variation model. TT rank is a well-known tensor rank, and it constitutes of ranks of matrices formed by a well-balanced matricization method to reshapes the tensor to matrix along with each mode. TT rank appears in physical experiments [22], and it is applied to quantum dynamics simulation experiments [23, 24]. Low TT rank can also be applied to the compression of big data by singular value decomposition [25, 26]. The alternating least squares (ALS) is a satisfactory solution to tensor completion [27, 28]. The low TT rank tensor is applied in implementing the steepest descent iteration to solve large-scale least squares problems [29, 30]. Bengua et al. [31] proposed an approach to tensor completion by minimizing a nuclear norm on TT rank. Previous studies reveal that the method by imposing TT rank has good performance in processing tensor data. Total variation (TV) [32] is a guaranteed norm regularization to encourage piece-wise smoothness, and has been used to solve many visual data problems. A tensor completion model combined with Tucker rank and TV is proposed in [33, 34], and the result shows its performance in visual data completion and also analyzes the expansion under noisy observation. A low-rank smooth PARAFAC decomposition method that considers TV and quadratic variation (QV) is proposed in [35]. Another completion model, which combined TT rank with TV is proposed in [36] by assuming tensor train structures in the underlying regression model. This model is rephrased as a regression task and uses the alternating linear scheme to update tensor train cores, but the result also shows that ADMM-TV method performs better than this method in RSE and PSNR scores.

In our method, we introduce nuclear norm regularization on TT rank which is the most effective surrogate for rank optimization for the global data structure. Meanwhile, we choose anisotropic TV as another regularization term, as anisotropic TV performs well in our model according to experimental results. VDT is introduced in this paper to reshape the MRI data to enhance the performance of the proposed algorithm. Based on the proposed optimization method, we divide the optimization problem into a series of sub-problems and then solve each problem. The results show that our method takes advantages in relative standard error (RSE), peak signal-to-noise ratio (PSNR), and structural similarity index (SSIM). It also concludes that our method achieves better accuracy compared with other methods based on the low-rank constraint.

3.1.2 Previous work about total variation and tensor completion

The formulation of total variation [37] can be denoted by:

$$\|\mathbf{X}\|_{TV-A} = \|\nabla_h \mathbf{X}\|_1 + \|\nabla_v \mathbf{X}\|_1, \tag{3.1}$$

where $\|\mathbf{X}\|_{TV-A}$ is the representation of anisotropic total variation, $\nabla_h \mathbf{X}$ is the horizontal difference operator and $\nabla_v \mathbf{X}$ is the vertical difference operator, and they can be written as:

$$\nabla_h \mathbf{X} = vec(\mathbf{X}_h), \nabla_v \mathbf{X} = vec(\mathbf{X}_v), \tag{3.2}$$

where $\mathbf{X}_v = \mathbf{X}_{(i_1+1,i_2)} - \mathbf{X}_{(i_1,i_2)}$, $\mathbf{X}_h = \mathbf{X}_{(i_1,i_2+1)} - \mathbf{X}_{(i_1,i_2)}$. The tensor completion is basically evolved from the matrix completing. The goal is to complete its lost parts from partially known entries of an incomplete matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$. We can apply the matrix-rank optimization model to solve this problem:

$$\min_{\mathbf{X}} \ \text{Rank}(\mathbf{X}) \quad s.t. \ P_\Omega(\mathbf{X}) = P_\Omega(\mathbf{T}), \tag{3.3}$$

where $\mathbf{T}$ is observed data, and Ω is the subaggregate of partially known entries and $P_\Omega(\mathbf{T})$ represents the observed entries. The matrix $\mathbf{X}$ with missing data can be recovered by assuming that the matrix has the low-rank structure. For example, the vector $(\lambda_1, \lambda_2, \ldots, \lambda_{min(m,n)})$ of the singular values λ_i is as sparse as possible. The completion accuracy of $\mathbf{X}$ can be influenced by the sparsity of $(\lambda_1, \lambda_2, \ldots, \lambda_{min(m,n)})$, because of its nature of the function, moreover, function (3) is an NP-hard problem. It is proved that matrix nuclear norm is an effective convex surrogate to solve rank minimization function, and the matrix completion can also be reformulated as:

$$\min_{\mathbf{X}} \ \|\mathbf{X}\|_* \quad s.t.\ P_\Omega(\mathbf{X}) = P_\Omega(\mathbf{T}). \tag{3.4}$$

In addition, the total variation is a classical model for image restoration, it is a guarantee for regional data structure, which is the important information for image completion. So a new model based on low rank and the total variation is proposed [38], and its formulation is:

$$\min_{\mathbf{X}} (1-\varphi)\ \|\mathbf{X}\|_* + \varphi\ \|\mathbf{X}\|_{LTV} \quad s.t.\ P_\Omega(\mathbf{X}) = P_\Omega(\mathbf{T}), \tag{3.5}$$

where φ is a trade-off parameter, and its value is between 0 and 1, then we choose anisotropic TV as optimization norm, the optimization can be written as:

$$\|\mathbf{X}\|_{LTV} = \sum_{i_1,i_2} (\mathbf{X}_v(i_1,i_2)^2 + \mathbf{X}_h(i_1,i_2)^2). \tag{3.6}$$

Alternating direction method of multipliers (ADMM) [39] is introduced to solve the problem (5). In fact, in some practical experiments, the processed data is larger than 3 dimensions. Therefore it is necessary to reshape the data from high order tensor to the matrix. However, at the same time, it will lead to performance loss, because some high-order space information is lost during the data conversion.

Tensor completion is similar to matrix completion. Recover a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ from its known data with a subset Ω can be written as:

$$\min_{\mathcal{X}} \ \mathrm{Rank}(\mathcal{X}) \quad s.t.\ P_\Omega(\mathcal{X}) = P_\Omega(\mathcal{T}), \tag{3.7}$$

$\mathrm{Rank}(\mathcal{X})$ represents the rank of $\mathcal{X}$. $P_\Omega(\mathcal{X}) = P_\Omega(\mathcal{T})$ means $\mathcal{X}_{(i_1,\cdots,i_N)} = \mathcal{T}_{(i_1,\cdots,i_N)}$ and $(i_1,\cdots,i_N) \in \Omega$. CP ranks and Tucker ranks can also be applied to this optimization model [40]. Tucker rank minimization model can be written as :

$$\min_{\mathbf{X}_{(n)}} \sum_{n=1}^{N} \alpha_n \ \mathrm{Rank}(\mathbf{X}_{(n)}) \quad s.t.\ P_\Omega(\mathcal{X}) = P_\Omega(\mathcal{T}), \tag{3.8}$$

where α_n are the elements with $\sum_{n=1}^{N} \alpha_n = 1$. It can be reformulate as:

$$\min_{\mathbf{X}_{(n)}} \sum_{n=1}^{N} \alpha_n \ \|\mathbf{X}_{(n)}\|_* \quad s.t.\ P_\Omega(\mathcal{X}) = P_\Omega(\mathcal{T}), \tag{3.9}$$

$\mathbf{X}_{(\mathbf{n})}$ is denoted as the mode-n unfolding matrix of tensor $\mathcal{X}$. The high accuracy low-rank tensor completion (HaLRTC) is applied to solve model (9) by adding an equation constraint [40].

There is another method for LRTC problem (8), which is based on TT rank optimization [31]. It can be written as:

$$\min_{\mathbf{X}_{[n]}} \alpha_n \sum_{n=1}^{N-1} \text{Rank}\|\mathbf{X}_{[n]}\|_* \quad s.t.\ P_\Omega(\mathcal{X}) = P_\Omega(\mathcal{T}), \tag{3.10}$$

where α_n represents the parameter of matrix $\mathbf{X}_{[n]}$, which denoted as the mode-n unfolding matrix $\mathbf{X}$, and its condition is $\sum_{n=1}^{N-1} \alpha_n = 1$. The TT rank obtains the relationship between n modes and the other modes. Hence $(Rank(\mathbf{X}_{[1]}), Rank(\mathbf{X}_{[2]}), \cdots, Rank(\mathbf{X}_{[N]}))$ guarantees a satisfactory way to obtain the global structure of the data. However, it is difficult to find a solution to the problem (10). At last, the problem is based on TT nuclear norm, and it can be written as:

$$\min_{\mathcal{X}} \sum_{n=1}^{N-1} \alpha_n \|\mathbf{X}_{[n]}\|_* \quad s.t.\ P_\Omega(\mathcal{X}) = P_\Omega(\mathcal{T}). \tag{3.11}$$

3.1.3 Simple low rank tensor completion combined with TT rank

To address the problem (11) it can be converted to the following problem:

$$\begin{aligned} &\min_{\mathcal{X},\mathbf{M}_n} \sum_{n=1}^{N-1} \alpha_n \|\mathbf{M}_n\|_* + \beta_n/2 \|\mathbf{X}_{[n]} - \mathbf{M}_n\|_F^2 \\ &s.t.\ P_\Omega(\mathcal{X}) = P_\Omega(\mathcal{T}), \end{aligned} \tag{3.12}$$

where β_k are a series of positive parameters, and problem (12) is applied on block coordinate descent (BCD) which is a generalization of coordinate descent. The method decomposes the variables into two groups. The first one involves the unfolding matrices $M_1, M_2, \ldots, M_{(N-1)}$ and the other variable is $\mathcal{X}$, The $\mathbf{M}_n$ can be obtained by solving the following optimization problem:

$$\begin{aligned} &\min_{\mathbf{M}_n} \alpha_n \|\mathbf{M}_n\|_* + \beta_n/2 \|\mathbf{X}_{[n]} - \mathbf{M}_n\|_F^2 \\ &s.t.\ P_\Omega(\mathcal{X}) = P_\Omega(\mathcal{T}), \end{aligned} \tag{3.13}$$

the $\mathbf{X}_{[n]}$ is fixed and the optimal solution for (13) has another expression [41], it can be represented as:

$$\mathbf{M}_n = D_{\gamma n}(\mathbf{X}_{[n]}), \tag{3.14}$$

where $D_{\gamma n} = \frac{\alpha_n}{\beta_n}$ and $D_{\gamma n}(\mathbf{X}_{[n]})$ represents the thresholding SVD of $\mathbf{X}_{[n]}$. In addition, if the SVD of $\mathbf{X}_{[n]} = U\lambda V^T$, it can be written as:

$$D_{\gamma n}(\mathbf{X}_{[n]}) = U\lambda_{\gamma n} V^T, \tag{3.15}$$

where $\lambda_{\gamma n} = diag(max(\lambda_l - \gamma_n, 0))$. When the $\mathbf{M}_n$ is obtained, the tensor $\mathcal{X}$ can be computed by another equation, which can be written as:

$$\mathcal{X}_{i_1,\ldots,i_N} = \begin{cases} (\frac{\sum_{n=1}^N \beta_n fold(\mathbf{M_n})}{\sum_{n=1}^N \beta_n})_{i_1,\ldots,i_N}, (i_1,\ldots,i_N) \notin \Omega; \\ t_{i_1,\ldots,i_N}, (i_1,\ldots,i_N) \in \Omega. \end{cases} \tag{3.16}$$

This algorithm can be named as simple low-rank tensor completion based on tensor train (SiLRTC-TT) [40]. The convergence condition will be satisfied when the difference value between two consecutive recovered data is small than a given value.

3.2 Proposed method

In the third part, we are going to introduce the model, which combines TT rank and total variation, and it can be denoted as:

$$\min_{\boldsymbol{\mathcal{X}}} \ \alpha^T k(x) + \varphi TV(x) \quad s.t. \ P_\Omega(\boldsymbol{\mathcal{X}}) = P_\Omega(\boldsymbol{\mathcal{T}}), \tag{3.17}$$

where φ is the parameter, and $k(x) = [k_1, k_2, \ldots, k_N]$ is the TT ranks. $\alpha(x) = [\alpha_1, \alpha_2, \ldots, \alpha_N]$ is denoted as the TT rank which condition is $\sum_{n=1}^N \alpha_n = 1$, anisotropic TV is chosen as TV norm. $\mathcal{T}$ is the observed tensor, and Ω is the subaggregate of partially known data in $\mathcal{T}$. The equation $P_\Omega(\mathcal{X}) = P_\Omega(\mathcal{T})$ represents $\mathcal{X}_{i_1,\ldots,i_N} = \mathcal{T}_{i_1,\ldots,i_N}$ when $i_1, \ldots, i_N \in \Omega$. The TT ranks $k(x) = [k_1, k_2, \ldots, k_N]$ are nonconvex in the objective function, and matrix nuclear norms are applied to the optimization model as the convex surrogates, and the new convex model can be written as:

$$\begin{aligned} &\min_{\mathcal{X}} \sum_{n=1}^{N-1} \alpha_n \left\| \mathbf{X}_{[n]} \right\|_* + \varphi \left\| \mathcal{D}(\mathcal{X}) \right\|_p \\ &s.t. \ P_\Omega(\mathcal{X}) = P_\Omega(\mathcal{T}), \end{aligned} \tag{3.18}$$

where $\mathbf{X}_{[n]}$ an be obtained by matricization of $\mathcal{X}$. $\|\mathcal{D}(\mathcal{X})\|_p$ is total variation based on data $\mathcal{X}$. In this method we choose anisotropic TV. We apply additional tensor variables $\mathcal{Y}$ and

$\mathcal{M}$ which structure is similar to $\mathcal{X}$ to solve the optimization problem (18). Then the $\mathbf{L}$ is introduced to denote the difference of the data, and the new formulation can be written as:

$$\begin{aligned} &\min_{\boldsymbol{\mathcal{X}}} \sum\nolimits_{n=1}^{N-1} \alpha_n \left\| \mathbf{X}_{[n]} \right\|_* + \varphi \left\| \boldsymbol{\mathcal{L}} \right\|_p \\ &s.t.\ P_\Omega(\mathcal{X}) = P_\Omega(\mathcal{T}), \mathcal{M} = \mathcal{X}, \mathcal{Y} = \mathcal{M}, \mathbf{L} = \mathcal{D}(\mathcal{Y}), \end{aligned} \tag{3.19}$$

the problem (19) can be transformed to another form:

$$\begin{aligned} &\min_{\mathcal{X},\mathcal{Y},\mathcal{M},\mathbf{L},\Lambda_1,\Lambda_2,\Lambda_3} \sum\nolimits_{n=1}^{N-1} \alpha_n \left\| \mathbf{X}_{[n]} \right\|_* + \varphi \left\| \mathbf{L} \right\|_{TV} - \langle \Lambda_1, \mathcal{M} - \\ &\mathcal{X} \rangle + \frac{\beta_1}{2} \left\| \mathcal{M} - \mathcal{X} \right\|_F^2 - \langle \Lambda_2, \mathcal{Y} - \mathcal{M} \rangle + \frac{\beta_2}{2} \left\| \mathcal{Y} - \mathcal{M} \right\|_F^2 - \\ &\langle \Lambda_3, \mathbf{L} - \mathcal{D}(\mathcal{Y}) \rangle + \frac{\beta_3}{2} \left\| \mathbf{L} - \mathcal{D}(\mathcal{Y}) \right\|_F^2 \\ &s.t.\ P_\Omega(\mathcal{X}) = P_\Omega(\mathcal{T}), \end{aligned} \tag{3.20}$$

where $\Lambda_1, \Lambda_2, \Lambda_3$ are the dual variables and $\beta_1, \beta_2, \beta_3$ are positive parameters. We can achieve a global optimization solution because it is a convex problem. Alternating direction method of multiples (ADMM) is applied to solve this problem (20). By applying ADMM method, one of the variables can be minimized along with the other variables are fixed. The (20) can be split into some sub-problems:

The first one problem can be written as:

$$\begin{aligned} &\min_{\mathcal{M}} \sum\nolimits_{n=1}^{N-1} \alpha_n \left\| \mathbf{M}_{[n]} \right\|_* - \langle \Lambda_1, \mathcal{M} - \mathcal{X} \rangle + \\ &\frac{\beta_1}{2} \left\| \mathcal{M} - \mathcal{X} \right\|_F^2 - \langle \Lambda_2, \mathcal{Y} - \mathcal{M} \rangle + \frac{\beta_2}{2} \left\| \mathcal{Y} - \mathcal{M} \right\|_F^2. \end{aligned} \tag{3.21}$$

The problem (21) can be transformed into:

$$\begin{aligned} &\min_{\mathcal{M}} \sum\nolimits_{n=1}^{N-1} \alpha_n \left\| \mathbf{M}_{[n]} \right\|_* + \\ &\frac{\beta_1 + \beta_2}{2} \left\| \mathcal{M} - \frac{\Lambda_1 + \beta_1 \mathcal{X} + \beta_2 \mathcal{Y} - \Lambda_2}{\beta_1 + \beta_2} \right\|_F^2. \end{aligned} \tag{3.22}$$

Letting $\tau = \frac{\alpha_n}{\beta_1 + \beta_2}$, $\mathbf{S} = \frac{\Lambda_1 + \beta_1 \mathcal{X} + \beta_2 \mathcal{Y} - \Lambda_2}{\beta_1 + \beta_2}$,it can reformulated as:

$$\min_{\mathcal{M}} \sum\nolimits_{n=1}^{N-1} \tau \left\| \mathbf{M}_{[n]} \right\|_* + \frac{1}{2} \left\| \mathbf{M}_{[n]} - \mathbf{S}_n \right\|_F^2, \tag{3.23}$$

the $\mathbf{M}_{[n]}$ can be obtained by optimizing the (23), and problem (23) is similar to (13) that can be solved by the same method.

The second problem can be represented as:

$$\begin{aligned}&\min_{\mathcal{Y}} -\langle \Lambda_2, \mathcal{Y}-\mathcal{M}\rangle + \frac{\beta_2}{2}\,\|\mathcal{Y}-\mathcal{M}\|_F^2 \\ &-\langle \Lambda_3, \mathbf{L}-\mathcal{D}(\mathcal{Y})\rangle + \frac{\beta_3}{2}\,\|\mathbf{L}-\mathcal{D}(\mathcal{Y})\|,\end{aligned} \tag{3.24}$$

the second function with $\mathcal{Y}$ is different. This problem can be solved by the following equation:

$$(\beta_2\mathbf{I}+\beta_3\mathcal{D}^*\mathcal{D})\mathcal{Y} = \mathcal{D}(\beta_3 L-\Lambda_3)+\beta_3\mathcal{M}+\Lambda_2, \tag{3.25}$$

where $\mathcal{D}^*$ is the adjoint of $\mathcal{D}$. and $\mathcal{D}^*\mathcal{D}$ is changed into the Fourier domain and fast calculated. Moreover, the off-the-shelf conjugates gradient method [42] is applied to solve the equation, and the solution of $\mathcal{Y}$ can be denoted as:

$$\mathcal{Y} = ifftn(\frac{fftn(\mathcal{S})}{\beta_2\mathbf{I}+\beta_3(fftn(\mathcal{D}^*\mathcal{D}))}), \tag{3.26}$$

where $\mathcal{S} = \mathcal{D}(\beta_3 L-\Lambda_3)+\beta_3\mathcal{M}+\Lambda_2$. $fftn$ is fast 3D Fourier transform, and $ifftn$ is fast 3D in-verse Fourier transform. Moreover, the computational cost can be decreased by pre-computing the operator $\boldsymbol{\mathcal{D}}^*\boldsymbol{\mathcal{D}}$ that outside the main loop. The third problem can be written as:

$$\min_{\mathbf{L}} \varphi\,\|\mathbf{L}\|_{TV} - \langle \Lambda_3, \mathbf{L}-\mathcal{D}(\mathcal{Y})\rangle + \frac{\beta_3}{2}\,\|\mathbf{L}-\mathcal{D}(\mathcal{Y})\|_F^2, \tag{3.27}$$

the problem can be transformed to another form as it is the anisotropic total variation:

$$\min_{\mathbf{L}} \varphi\,\|\mathbf{L}\|_{TV} + \frac{\beta_3}{2}\,\|\mathbf{L}-(\mathcal{D}(\mathcal{Y})+\frac{\Lambda_3}{\beta_3})\|_F^2, \tag{3.28}$$

this problem can also be solved by:

$$\mathbf{L} = sth(\mathcal{D}(\mathcal{Y})+\frac{\Lambda_3}{\beta_3}, \frac{\varphi}{\beta_3}), \tag{3.29}$$

where sth is the soft thresholding, and it can be written as follows:

$$sth(x,\tau) = sgn(x)max(|x|)-\tau, 0). \tag{3.30}$$

The fourth problem is denoted as:

$$\min_{\mathcal{M}} \langle \Lambda_1, \mathcal{M} - \mathcal{X} \rangle + \frac{\beta_1}{2} \|\mathcal{M} - \mathcal{X}\|_F^2 \; s.t. \; P_\Omega(\mathcal{X}) = P_\Omega(\mathcal{T}). \tag{3.31}$$

The problem (20) is the convex problem, and the objective function is smooth and differentiable, and the tensor $\mathcal{X}$ is updated as:

$$\boldsymbol{\mathcal{X}}_{i_1,\ldots,i_N} = \begin{cases} (\boldsymbol{\mathcal{M}} - \frac{\Lambda_1}{\beta_1})_{i_1,\ldots,i_N}, (i_1, \ldots, i_N) \in \Omega; \\ t_{i_1,\ldots,i_N}, (i_1, \ldots, i_N) \notin \Omega. \end{cases} \tag{3.32}$$

The last problem can be written as:

$$\begin{aligned} &\min_{\Lambda_1,\Lambda_2,\Lambda_3} -\langle \Lambda_1, \mathcal{M} - \mathcal{X} \rangle + \frac{\beta_1}{2} \|\mathcal{M} - \mathcal{X}\|_F^2 - \langle \Lambda_2, \mathcal{Y} - \mathcal{M} \rangle \\ &+ \frac{\beta_2}{2} \|\mathcal{Y} - \mathcal{M}\|_F^2 - \langle \Lambda_3, \mathbf{L} - \mathcal{D}(\mathcal{Y}) \rangle + \frac{\beta_3}{2} \|\mathbf{L} - \mathcal{D}(\mathcal{Y})\|_F^2 \\ &s.t. \; P_\Omega(\mathcal{X}) = P_\Omega(\mathcal{T}). \end{aligned} \tag{3.33}$$

On the basis of ADMM, Λ_1, Λ_2 and Λ_3 can be solved through following equation:

$$\begin{aligned} \Lambda_1 &= \Lambda_1 - \beta_1(\boldsymbol{\mathcal{M}} - \boldsymbol{\mathcal{X}}) \\ \Lambda_2 &= \Lambda_2 - \beta_2(\boldsymbol{\mathcal{Y}} - \boldsymbol{\mathcal{M}}) \\ \Lambda_3 &= \Lambda_3 - \beta_3(\boldsymbol{\mathcal{L}} - \gamma(\boldsymbol{\mathcal{X}})), \end{aligned} \tag{3.34}$$

and the parameter $\beta = [\beta_1, \beta_2, \beta_3]$ is solved by the below equation:

$$\beta^t = \begin{cases} \eta_1 \beta^{(t-1)}, if \zeta^{(t)} > \eta_2 \zeta^{(t-1)} \\ \beta^{(t-1)}, otherwise, \end{cases} \tag{3.35}$$

where $\zeta^{(t)} = [\|(\mathcal{M}) - (\mathcal{X})\|, \|(\mathcal{Y}) - (\mathcal{M})\|, \|\mathbf{L} - \mathcal{D}(\mathcal{Y})\|]_T$ in t-th iteration, η_1 and η_2 are scale parameters. The missing ratio of data determines the value of η. The convergence condition will be satisfied, when the relative error between two consecutive recovered data is small than the given value, it can be denoted as $(\|\mathcal{X}^{(n)}\|_F - \|\mathcal{X}^{(n-1)}\|_F)/\|\mathcal{X}^{(n)}\|_F \leq \varepsilon$, $\mathcal{X}^{(n)}$ is the completed tensor in $t - th$ iteration and ε is a given value. This algorithm can ensure convergence of the global optimal solution.

TABLE 3.1: Algorithm

TT low-rank completion with total variation
Input: A tensor $\mathcal{X}$, which is going to be recovered, index Ω, vector β and ε is a small value for convergence condition and iteration number K.
Initialization: $\mathcal{P}_\Omega = \mathcal{T}_\Omega$, $\beta = \left[\frac{1}{\|\mathcal{T}_\Omega\|_F}, \frac{1}{\|\mathcal{T}_\Omega\|_F}, 0.01\right]^T$, $K = 300$, $\varepsilon = 10^-6$ other variables are set by experience.
Output: recovered tensor $\mathcal{X}$
1: update $\mathcal{M}$ by (23)
2: update $\mathcal{Y}$ by (26)
3: update $\mathcal{L}$ by (29)
4: update $\mathcal{X}$ by (32)
5: update $\Lambda_1, \Lambda_2, \Lambda_3$ via (34)
6: end while

3.3 Experiment and result

3.3.1 Experimental parameter selection

Three types of performance evaluation indicators on images are introduced to estimate the accuracy of different methods. They are relative standard error (RSE), peak signal-to-noise ratio ($PSNR$), and Structural similarity index measurement ($SSIM$). The RSE can be defined as: $RSE = \frac{\|\mathcal{X}-\mathcal{X}_0|_F}{|\mathcal{X}_0|_F}$, where $\mathcal{X}$ is the completed data and $\mathcal{X}_0$ is the original data. The $PSNR$ can be described as the error between two kinds data, and it can be written as $10\log_{10}(MAX^2/MSE)$, where MAX is the maximum value of the data. Mean squared error (MSE) can be written as $MSE = \sum_{i=0}^{m-1}\sum_{i=0}^{n-1} \|\boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{X}}_o\|_F^2/mn$.SSIM is an index which value is ranging from 0 to 1 to measure the similarity between two different images. It compares luminance and contrast [43] from the regional patterns of pixel intensities, the higher value of $SSIM$ represents better recovering performance. Our experiments are conducting on a computer with an Intel Core i7, 2.2 GHz CPU, and 16GB 1600 MHz DDR3 memory. The experiment is based on MRI images by applying the proposed method, and RSE, $SSIM$, and $PSNR$ are used to estimate its performance. We are going to compare our methods with some others: 1. ADMM-TV [44]; 2. T-mac [45]; 3. TMac-TT method [31]; 4. SiLRTC-TT method [31]; 5. PCLR method [46];6. TTC and TTC-TV method [36].

3.3.2 MRI image with size $256 \times 256 \times 30$

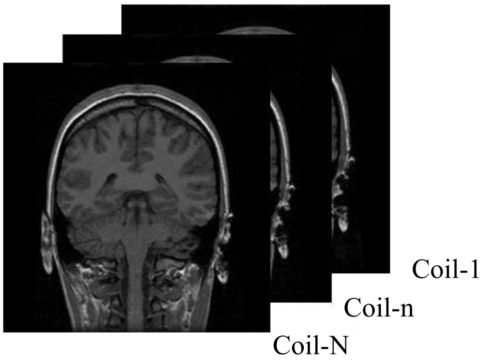

FIGURE 3.1: MRI image.

In the experiment, we applied the proposed method for MRI(Fig. 1) with size $256 \times 256 \times 30$ which can be downloaded from Figshare database. We randomly choose the missing ratio from 40% to 90%. Visual Data Tensorization (VDT) method [47] is applied in our experiment, and it is proved to be effective to improve the performance of tensor train method processing. The VDT method reshapes a matrix with size $2^l \times 2^l$ to a real ket of a Hilbert space, which is generalized from the visual data compression and entanglement method [41]. It is also can be described as developing from the KA augmentation [31]. It reshapes the original data to higher-dimensional data by a specific transformation with spatial structure information. The VDT operates as follows: there is a matrix with size $U \times V$ and the data can be reshaped to $u1 \times u2 \cdots \times ul \times v1 \times v2 \cdots \times vl$, then permute the data and represents it by another mode, which size is $u1v1 \times u2v2 \cdots \times ulvl$. The new tensor has the same elements as the original data, but the element is arranged in another way. There is a close correspondence between $u1 \times v1$ pixel block of the data and the first order of this reshaped tensor. Through adopting the VDT method, the proposed method can effectively use the structural information of data to obtain a better representation of low-rank tensor. The explanation of the VDT method procedure is shown in Fig. 2. The MRI data is reshaped from $256 \times 256 \times 30$ size to a 17-order tensor, which size is $2 \times 2 \times \cdots \times 30$, then reshape to a 9-order tensor by VDT method with size $4 \times 4 \times \cdots \times 30$.

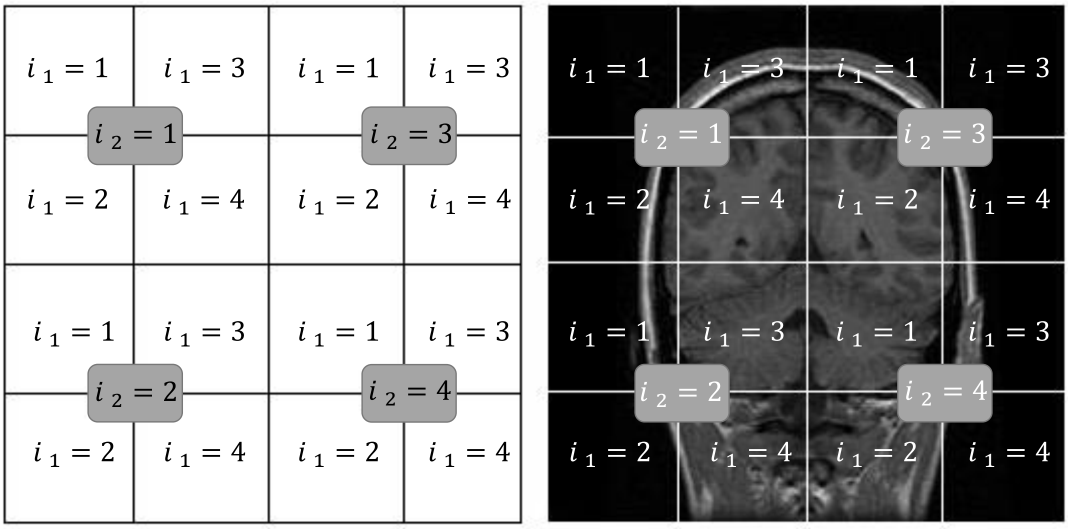

FIGURE 3.2: The left figure is the application of VDT on a matrix. The right figure is the operation on the MRI image.

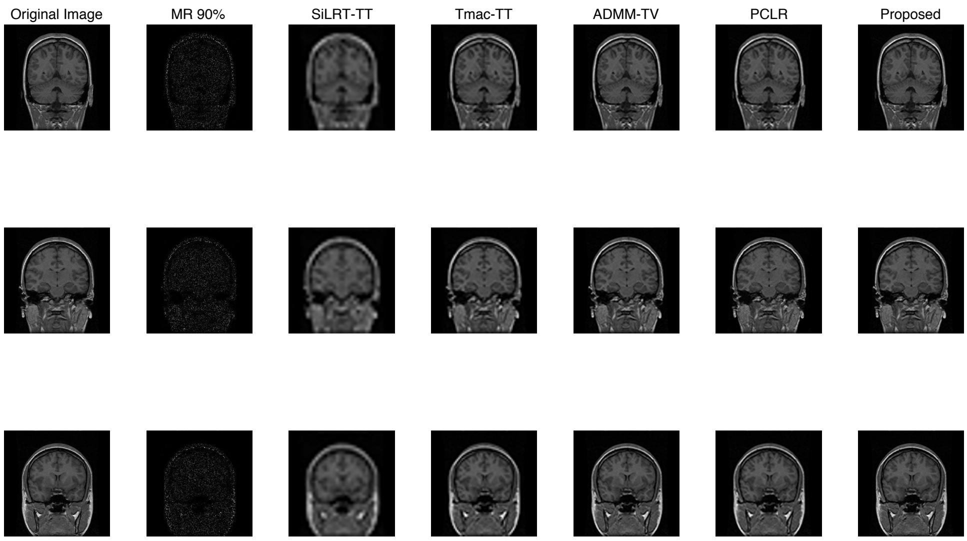

FIGURE 3.3: Figure from the first row to last row: the original MRI image and MRI image with 90% missing data and recovered image with different methods. The first row is based on a 6-coil image, the second and third row is based on 15-coil and 24-coil.

Table 1 shows the different models of completion efficiency in terms of RSE and PSNR. For the 6-coil, 15-coil, the 24-coil, and the whole coil of MRI data, the missing ratio ranges from 40% to 90% and the result indicates that our method consistently obtains better completion results over all other methods. Compared TTC-TV with our method, although the algorithmic complexity is reduced, the accuracy of data recovery cannot be guaranteed. Fig 3 and Fig 4 presents the performance of some well-known methods. There are two conclusions that can be drawn from the experiment. First, the observation demonstrates that TT low-rank completion is helpful and TT decomposition based on total variation works better than TT low-rank completion. Second, our proposed method produces better results than the other algorithms. The performance of our method is superior to ADMM-TV. Since TT ranks is well-balanced and capture the inner low-rank information efficiently. Compared our method with the SiLRTC-TT, our method has better performance, since incorporate total variation into SiLRTC guarantees regional piece-wise smooth structures. $PCLR$ method applies linear relation-ship and phase constraint to recover the missing data, and in this method, the original MRI data is reconstructed to the matrix which size is bigger than previous data, but it loses some structure information to recover the data. The regular low-rank completion performs well with observed data as the prediction. As the missing ratio improves, our model performs better than other models as the result shows that the $PSNR$ and $SSIM$ of SiLRTC decline faster than the proposed method. The proposed model describes the global and relative information of the MRI data, even with 90% missing ratio, it uses this constraint to recover the data with satisfactory accuracy.

TABLE 3.2: RSE and PSNR of different methods on MRI data

Method	The 6-coil		The 15-coil		The 24-coil		The whole coil	
	RSE	PSNR	RSE	PSNR	RSE	PSNR	RSE	PSNR
SiLRTC-TT	0.192	17.02	0.199	16.40	0.197	15.95	0.203	16.10
TTC	0.173	19.85	0.175	18.13	0.175	18.80	0.177	19.62
Tmac	0.177	17.71	0.181	16.64	0.179	16.58	0.184	16.49
Tmac-TT	0.169	17.79	0,168	17.63	0.168	17.18	0.169	17.63
TTC-TV	0.146	22.41	0.148	22.09	0.145	22.39	0.147	22.17
ADMM-TV	0.129	23.04	0.130	22.53	0.127	23.21	0.133	22.93
PCLR	0.126	22.52	0.128	21.86	0.126	22.09	0.129	21.36
Proposed	0.117	25.32	0.121	24.15	0.120	25.08	0.122	24.15

3.4 Conclusion

In this paper, a new solution to the low-rank tensor train combining with total variation model is proposed. The lower tensor train rank minimization is a guarantee for the global information regularization and the total variation encourages piece-wise smoothness for regional data constraint. By using the VDT method, we permuted the MRI images from 3-dimensional tensor to high-higher-dimensional tensor, and then apply ADMM method to solve the proposed low-rank model to reconstruct the MRI data. In numerical experiments, the result proves that our method achieves a better performance than other methods.

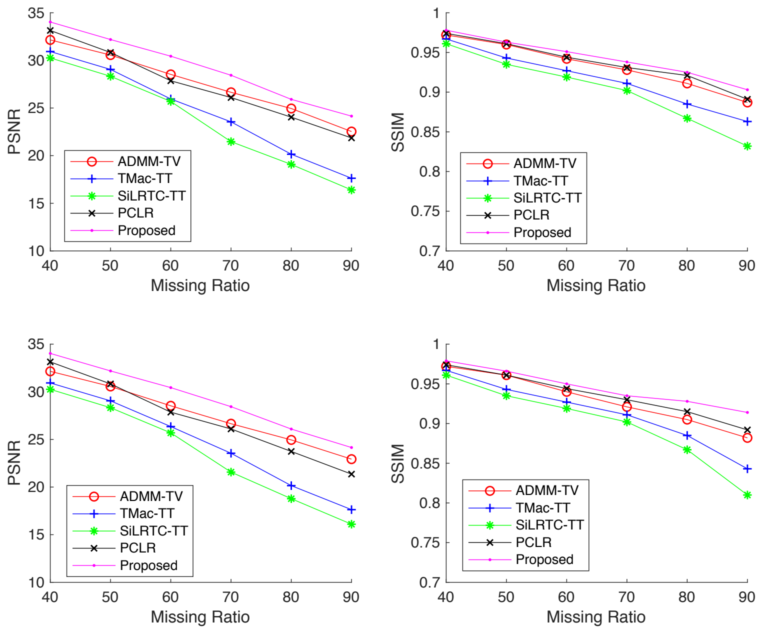

FIGURE 3.4: The left figure is the PSNR of different methods of the 15th coil and the whole image on different missing ratios. The right figure is the SSIM of different methods of the 15th coil and whole image on different missing ratios.

Chapter 4

Black-box adversarial attack by T-svd

4.1 Preliminaries

4.1.1 Proposed method introduction

In order to improve the query efficiency, we propose a method that changes the objective of the adversarial perturbation attacks from the original image pixel data to another form with a smaller amount of data. Preserve the original structure of high-dimensional tensor can obtain more spatial information from data processing by tensor method. Tensor singular value decomposition [48] is one of the essential tensor methods and it is utilized to decompose the image data and it is an important tool to analyze data [49] [50], we can obtain low-rank(high value) parts and high-rank parts of the image. Some attack methods have been confirmed that the perturbation is roughly concentrated in the high-rank part and these attack methods can be easily defended by low-rank assumptions [49] [51]. In the proposed method, the perturbation is added to both the high-rank part and the low-rank part.

In this paper, we propose a simple and effective black-box attack method. Firstly, the original image is divided into two orthogonal tensors and one rectangular diagonal tensor by Tensor Singular Value decomposition(t-SVD). The noise tensor is added into the rectangular diagonal tensor to construct image perturbation. In order to improve the efficiency of the proposed method, we don't have to pay too much attention to the optimal direction. Specifically, we randomly pick the noise tensor from specified sets and then attack the data by adding or subtracting the direction tensor into the singular value tensor. We utilize the confidence scores to check if the result is away from the decision boundary.

4.2 Adversarial attack

When constructing adversarial perturbation in image classification, the purpose is to change the output of the model predictions by adding imperceptible perturbation to original images. The perturbation should be restricted and they are imperceptible to humans. Generally, the same images should be classified into the same label and prediction, but the same images may have different outputs for machine learning classifiers. In this paper, we define the classifier model as h, and the image data as $\mathcal{X}$ with the model correctly predicts $y = h(\mathcal{X})$, the purpose of the adversary attack is going to find a perturbed image $\mathcal{X}'$ to change the output:

$$h(\mathcal{X}') = \mathcal{X}' \ \textbf{subject to} \ \forall \mathcal{X}' \in \{\delta(\mathcal{X}, \mathcal{X}')\} \leq \rho \tag{4.1}$$

the $\delta(\mathcal{X}, \mathcal{X}')$ is the perceptual difference between the original and perturbed images, and it can be defined by the L_0, L_2 and L_∞. Following [52] [53], we choose $\delta(\mathcal{X}, \mathcal{X}') = \|\mathcal{X} - \mathcal{X}'\|_2$ as perceptual difference. For a successful adversarial attack algorithm, the perceptual difference should be as small as possible to the extent that the perturbed image is imperceptibly different.

4.2.1 Untargeted and targeted attack

There are two different kinds of successful attack conditions. The simple one is the untargeted attack and it is defined as $h(\mathcal{X}') \neq y$, the objective of this attack is to change the output of original prediction. Another kind attack is targeted attack and it is represented as $h(\mathcal{X}') = y'$, y' is an incorrect pre-chosen prediction of the model.

Adding adversarial perturbation to original data to change the output is a discrete optimization problem. Therefore it is necessary to define a surrogate loss $\ell_y(\cdot)$ to measure the degree between model h and output y. The problem can be described as:

$$\min_{\delta} \ \ell_y(\mathcal{X} + \delta) \ \textbf{subject to} \ \|\delta\|_2 < \rho \tag{4.2}$$

4.2.2 Attack models

There are two kinds of attack models, they are white-box attacks and black-box attacks. If attackers are familiar with classifier model h, back-propagation can be utilized on the target model because the model structure and parameter settings are exposed to the attacker. Gradient descent can be performed on the loss function $l_y(\mathbf{x}')$, y represents correct class.

In fact, for most real-world scenarios, attackers do not have information about the target model, white-box attacks are restricted to be applied. For black-box attacks, the most valid operation is to input the data to the model and get the corresponding output. The black-box attack method is much more practical for the adversary. For example, when we choose to attack Google Cloud Vision, it will cost time and money in each query, therefore in addition to remaining the perturbed image is imperceptible, minimize the number of queries should also be considered. The new optimization problem can be represented as:

$$\min_{\delta} \;\; \ell_y(\mathbf{x}+\delta) \;\; \textbf{subject to} \;\; \|\delta\|_2 < \rho, \textbf{queries} \leq B \tag{4.3}$$

where B is the maximum of the queries we fix in the algorithm.

Theorem 1 There is a tensor $\mathcal{A}$ with size $\mathbf{R}^{n_1\times n_2\times n_3}$, a tensor $\mathcal{B}$ with size $\mathbf{R}^{n_2\times n_4\times n_3}$, and a tensor $\mathcal{C}$ with same size with tensor $\mathcal{B}$, and they satisfy the commutative law:

$$\mathcal{A} * (\mathcal{B} + \mathcal{C}) = \mathcal{A} * \mathcal{B} + \mathcal{A} * \mathcal{C} \tag{4.4}$$

Theorem 2 If a tensor $\mathcal{A}$ with size $\mathbf{R}^{n_1\times n_2\times n_3}$, then we define the $\mathcal{A}^T$ by conjugate transposing each of the frontal slice of $\mathcal{A}$ and then reversing the order of transposed frontal slices 2 through n_3.

Theorem 3 A tensor $\mathcal{A}$ with size $\mathbf{R}^{n_1\times n_1\times n_3}$is orthogonal, if it satisfies:

$$\mathcal{A}^T * \mathcal{A} = \mathcal{A} * \mathcal{A}^T = \mathcal{I} \tag{4.5}$$

where $\mathcal{I}$ is identity tensor with size $\mathbf{R}^{n_1\times n_1\times n_3}$ whose first frontal slice is identity matrix and other frontal slices are zero matrix.

Theorem 4 If $\mathcal{A}$ is an orthogonal tensor, the L_2 norm of $\mathcal{A} * \mathcal{B}$ can be denoted as:

$$\langle \mathcal{A} * \mathcal{B}, \mathcal{A} * \mathcal{B} \rangle = \langle \mathcal{B}, \mathcal{B} \rangle \tag{4.6}$$

For a color image data $\mathcal{X} \in \mathbf{R}^{n_1\times n_2\times n_3}$, the t-SVD of $\mathcal{X}$ can be represented as:

$$\mathcal{X} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T \tag{4.7}$$

where $\mathcal{U}$ and $\mathcal{V}$ are orthogonal tensors with size $n_1 \times n_1 \times n_3$ and $n_2 \times n_2 \times n_3$. $\mathcal{S}$ is the rectangular diagonal tensor with size $n_1 \times n_2 \times n_3$. Although tensor $\mathcal{X}$ and tensor $\mathcal{S}$ have same size, $\mathcal{S}$ is a diagonal tensor and $\mathcal{X}$ is a tensor with full data, hence adding perturbation on tensor $\mathcal{X}$ is more efficient. The perturbed image can be formulated as $\mathcal{X}' = \mathcal{U} * (\mathcal{S}') * \mathcal{V}^T$, and the equation(3) can be rewritten as:

$$\min_{\delta} \quad \ell_y(\mathcal{X}, \mathcal{X}') \ \ \textbf{subject to} \ \ \|\delta\|_2 < \rho, \textbf{queries} \leq B \tag{4.8}$$

Theorem 5 For a $\mathcal{X}$ with size $\mathbf{R}^{n_1 \times n_2 \times n_3}$, and the t-SVD of $\mathcal{X}$ is decomposed as $\mathcal{X} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T$. The L_2 norm of $\mathcal{X}$ can be written as:

$$\langle \mathcal{X}, \mathcal{X} \rangle = \langle \mathcal{S}, \mathcal{S} \rangle \tag{4.9}$$

4.3 proposed method

4.3.1 Algorithm

In this section, we are going to introduce our method. There are some original images, and we define them as $\mathcal{X}$. Through a neural network classifier model h, the output of label y is classified with predicted confidence or probability $p_h(y|\mathcal{X})$. The proposed algorithm is to add perturbation δ to change the output $h(\mathcal{X} + \delta) \neq y$. Because we are blind to the model h, the output of each query $h(\mathcal{X} + \delta)$ is valuable and exclusive information for us.

The algorithm is proposed in this section. Firstly, we decompose the original image $\mathcal{X}$ by t-SVD and the diagonal tensor S can be calculated, which is the objective to be attacked. In our method, we represent the noise tensor as $\mathcal{Q}$ and step size as ϵ, and the perturbation can be written as $\mathcal{U} * \alpha \mathcal{Q} * \mathcal{V}^T$. The perturbation will be added to the original image, if the output probabilities of image $p_h(y|\mathcal{X} + \delta)$ is decreasing, we consider the step of attack can be kept to the data $\mathcal{X}$ and next attack perturbation can be written as $\delta + \mathcal{U} * \alpha \mathcal{Q} * \mathcal{V}^T$, otherwise we subtract perturbation. If neither adding nor subtracting perturbation can reduce the probability

TABLE 4.1: Algorithm

Simple Black-box adversarial attacks by t-SVD
Input: Original image $\mathcal{X}$, query direction $\mathcal{Q}$ that belong to vectors **Q**, step size. ϵ
1: $\mathcal{X} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T$, δ=0
2: $\mathbf{p}$=$p_y(y\|\mathcal{X})$
3: **if** $\mathbf{p}_y = max_{y'}\mathbf{p}_{y'}$ **do**
4: **for** $\alpha \in (0, \epsilon)$ **do**
5: $\mathbf{p}'$=$p_h(y\|\mathbf{x} + \delta + \mathcal{U} * \alpha\mathcal{Q} * \mathcal{V}^T)$
6: **if** $\mathbf{p}'_y < \mathbf{p}_y$ **then**
7: δ=δ+$\alpha\mathcal{Q}$
8: $\mathbf{p} = \mathbf{p}'$
9: **break**
10: **return** δ

of the result, we consider the step as an invalid attack and the perturbation will be discarded. The noise tensor $\mathcal{Q}$ is randomly picked from the set W.

The candidate diagonal tensor W can be comprised of some different kinds of basis tensor, they are the standard basis, random orthogonal diagonal basis and some specified diagonal basis. The first choice for the attack direction is the standard basis $\mathcal{I}$. Recent work has discovered that orthogonal noise is more likely to be adversarial [54]. The random diagonal basis attack is effective, but we found that compared with standard basis and random orthogonal diagonal basis, adding specific orthogonal diagonal basis noise into W will increase the efficiency of the attack and natural suitability to images [54]. In this paper, we prescribe each direction $\mathcal{Q}_i$ have two characteristics, the one is $\langle \mathcal{Q}, \mathcal{Q} \rangle = 1$ and another is $\langle \mathcal{Q}_i, \mathcal{Q}_{\neq i} \rangle = 0$.

4.3.2 Budget considerations

Considering the sets of noise tensor W, we find that the L_2 norm of perturbation $||\delta|_2$ can be restricted. For each attack iteration, the noise tensor is either added or subtracted to the tensor $\mathcal{S}$. If neither adding nor subtracting can change the output probability, we discard the picked noise tensor in this iteration. In this paper, we define $\alpha \in (0, \epsilon)$ as the step size and after T iteration, the perturbation can be represented as:

$$\delta_T = \delta_t + \mathcal{U} * \alpha_t \mathcal{Q}_t * \mathcal{V}^T \tag{4.10}$$

the perturbation can also be rewritten as the sum of these each search directions:

$$\delta_T = \mathcal{U} * \sum_{t=1}^{T} \alpha_t \mathcal{Q}_t * \mathcal{V}^T \tag{4.11}$$

and the L_2 norm of the adversarial perturbation can be written as:

$$\begin{aligned} \|\delta_T\|_2^2 &= \langle \mathcal{U} * \sum_{t=1}^{T} \alpha_t \mathcal{Q}_t * \mathcal{V}^T, \mathcal{U} * \sum_{t=1}^{T} \alpha_t \mathcal{Q}_t * \mathcal{V}^T \rangle \\ &= \alpha_t^2 \langle \mathcal{U} * \sum_{t=1}^{T} \mathcal{Q}_t * \mathcal{V}^T, \mathcal{U} * \sum_{t=1}^{T} \mathcal{Q}_t * \mathcal{V}^T \rangle \end{aligned} \tag{4.12}$$

since t-product satisfy the Theorem 1, the right part $\langle , \rangle$ can be unfolded as:

$$\begin{aligned} &\langle \mathcal{U} * \sum_{t=1}^{T} \mathcal{Q}_t * \mathcal{V}^T, \mathcal{U} * \sum_{t=1}^{T} \mathcal{Q}_t * \mathcal{V}^T \rangle \\ &= \langle \mathcal{U} * \mathcal{Q}_1 * \mathcal{V}^T + \mathcal{U} * \mathcal{Q}_2 * \mathcal{V}^T + \ldots + \\ &\mathcal{U} * \mathcal{Q}_T * \mathcal{V}^T, \mathcal{U} * \mathcal{Q}_1 * \mathcal{V}^T + \ldots + \mathcal{U} * \mathcal{Q}_T * \mathcal{V}^T \rangle \end{aligned} \tag{4.13}$$

we assume the formula $\mathcal{U} * \mathcal{Q}_1 * \mathcal{V}^T$ into $\mathbf{a}_1$, ... and $\mathcal{U} * \mathcal{Q}_T * \mathcal{V}^T$ into $\mathbf{a}_T$, for any $i_1, i_2 \in [0, T]$, according to the matrix triple product operational rule, the equation can be transformed into:

$$\langle \mathbf{a}_1, \mathbf{a}_1 \rangle + \langle \mathbf{a}_1, \mathbf{a}_2 \rangle + \ldots + \langle \mathbf{a}_{i_1}, \mathbf{a}_{i_2} \rangle + \ldots + \langle \mathbf{a}_T, \mathbf{a}_T \rangle \tag{4.14}$$

according to theorem 5 and $\langle \mathcal{Q}_i, \mathcal{Q}_{\neq i} \rangle = 0$, for any $i_1 \neq i_2$ we have:

$$\langle \mathbf{a}_{i_1}, \mathbf{a}_{i_2} \rangle = \langle \mathcal{U} * \mathcal{Q}_{i_1} * \mathcal{V}^T, \mathcal{U} * \mathcal{Q}_{i_2} * \mathcal{V}^T \rangle = 0 \tag{4.15}$$

hence the equation(15) can be rewritten as:

$$\begin{aligned} \|\delta_T\|_2^2 &= \alpha_t^2 \sum_{t=1}^{T} \langle \mathcal{U} * \mathcal{Q}_t * \mathcal{V}^T, \mathcal{U} * \mathcal{Q}_t * \mathcal{V}^T \rangle \\ &= \alpha_t^2 \sum_{t=1}^{T} \langle \mathcal{Q}_t, \mathcal{Q}_t \rangle \leq T\epsilon^2 \end{aligned} \tag{4.16}$$

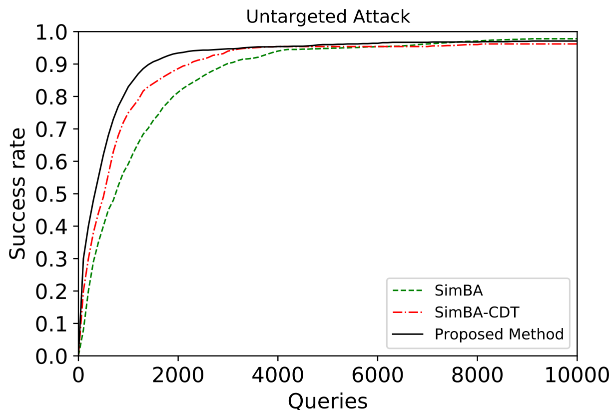

FIGURE 4.1: The success rate and the number of cost queries compared with SimBA, SimBA-DCT and our proposed method by untargeted attacks. The success rate of proposed method increases faster than SimBA and SimBA-DCT methods.

Since $\mathcal{U}$ and $\mathcal{V}$ are constant tensors. From equation (15), we can find that ϵ is a vital parameter to restrict the perturbation. Meanwhile, We found that if the query is restricted, we can set ϵ higher to reduce the number of iterations, thereby obtaining a higher disturbance L2-norm. Otherwise, if small-norm solutions are proposed, restrict ϵ will require more queries in the same L_2 norm.

4.4 experiment and results

In this section, we are going to demonstrate the efficiency of the method by fooling the convolutional neural network (CNN) models with three types of performance evaluation: the cost of queries(B), the L_2 norm of perturbation(P), and the rate of the optimization problem to find a feasible point($success\ rate$). Meanwhile, we compare the proposed method with other black-box algorithms: the QL attack [55], the SimBA and the SimBA-DCT [54]. We use standard dataset: ImageNet [56]. Firstly, We randomly choose 1000 images from the ImageNet and then classify them with the correct label. In the experiment, we try to minimize the probability of the correct label in untargeted attacks and maximize the probability of the target label in targeted attacks, we limit the maximal $T = 10000$.

4.4.1 untargeted attack on google Cloud Vision

For the untargeted attack, the purpose is to change the correctly labeled image into the incorrect label. In this experiment, we test our proposed method by attacking the Google Cloud Vision API, and Fig 1 shows its efficiency, we also compare our method with SimBA and SimBA-DCT. The result shows that our method ultimately achieves a relatively high success rate and our method increases dramatically faster in success rate than SimBA and SimBA-DCT.

4.4.2 untargeted and targeted attack on ResNet-50

TABLE 4.2: Untargeted and targeted attack on ResNet-50

Attack Method	Avg queries		Avg L_2 norm		Success rate	
	Untargeted	Targeted	Untargeted	Targeted	Untargeted	Targeted
QL-attack	28185	20857	8.54	11.48	85.7%	98.9%
SimBA	1957	7902	4.31	9.48	98.7%	100%
SimBA-DCT	1539	8759	3.89	7.08	97.4%	96.4%
Proposed	1207	5783	4.76	8.76	96.8%	97.8%

Four attack methods are performed on ImageNet by the untargeted and targeted attack, and we choose three different metrics to evaluate the methods: the number of cost queries (lower is better), average L2-norm of average perturbation (lower is better), and success rate(higher is better). The proposed method achieves close to 98% success rate slightly lower than other methods but requires significantly fewer model queries. In this experiment, we test the performance of our method by attacking the ResNet-50 network [57] and compare it with QL-attack, SimBA and SimBA-DCT. Furthermore, untargeted attack and targeted attack are performed and the number of cost queries, success rate and average L_2 norm of perturbation is utilized to evaluate the performance of our method.

Ideally, we ensure that the success rate of each algorithm attack is as high as possible. We believe that the successful method constructs the perturbation with lower L_2 norm and the lower queries. From Table 2, we can find that our method has significantly lower queries than other methods. In the untargeted attack experiment, QL-attack only gets 85% but costs 28000 queries. Although compared to SimBA and SimBA-DCT, we do not achieve a higher success rate, but our method costs fewer queries. In the targeted attack experiment, the test methods are much more comparable, but our method still requires fewer queries than other methods.

4.4.3 The qualitative comparison of different methods

In this part, we randomly selected several images to verify the qualitative results of different methods. In this experiment, we choose SimBA and SimBA-DCT for comparison. Figure 2 shows the original images and the attacked images, as well as the L_2 norm of adversarial perturbation of each image and the number of cost queries. All methods have successfully

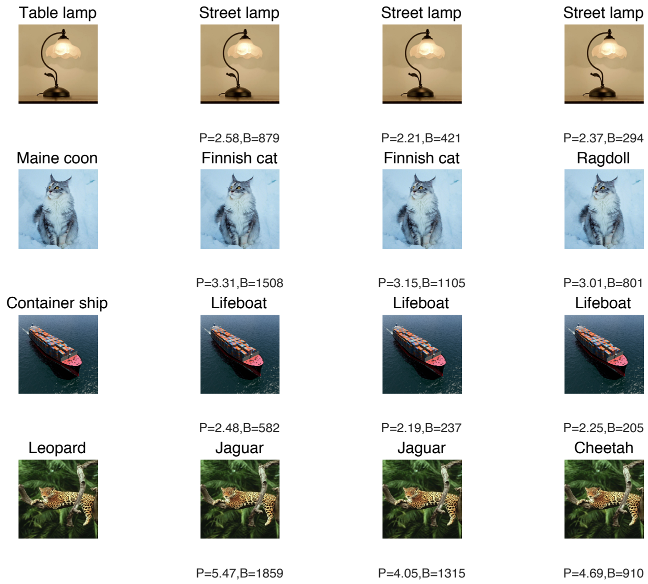

FIGURE 4.2: The first row of the figure is original image, the other rows are the result attacked by SimBA, SimBA-DCT, and the proposed method. P means the L_2 norm of adversarial perturbation and B means the cost number of queries. Comparing SimBA and SimBA-DCT, our method cannot guarantee the lowest L_2 norm of perturbation, but the number of queries is significantly less than the other two methods.

attacked the original image. Although our method cannot always achieve the smallest L2 norm, the number of queries consumed by our method is significantly less than other methods.

4.4.4 Evaluating different networks

In order to verify that our proposed method is also effective for other convolutional neural networks models, we choose DenseNet-121 [58] as our objective model for the untargeted attack. The result shows the success rate and the number of model queries with DenseNet-121 and ResNet-50 models. From Fig 3, we find that whether DenseNet-121 or ResNet-50 model are both vulnerable to our attack method, and DenseNet-121 model is trended to be fooled easier. From the experimental results, our method successfully attacks different CNN models with high probability.

4.5 Conclusion

In this paper, We are the first to utilize the tensor method to construct adversarial perturbations. A simple and effective black-box algorithm is proposed. We use tensor singular value decomposition to process the image and add specific perturbation into the singular value tensor to create perturbation. Our attack method is not only effective for different CNN models, but

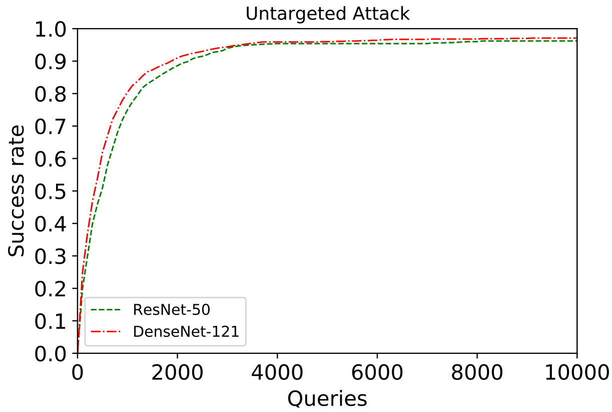

FIGURE 4.3: The success rate and the number of queries through ResNet-50 and DenseNet-121 models for untargeted attacks. Our method can fool both ResNet-50 and DenseNet-121 successfully within 10000 queries with high probability. Compared with ResNet-50 model, DenseNet is more vulnerable against untargeted attacks.

also more efficient than other methods (our method has a higher success rate in the first 1000 queries).

4.6 APPENDIX

4.6.1 PROOF OF THEOREM 1

For tensor $\mathcal{A}$ with size $\mathcal{A} \in \mathbf{R}^{n_1 \times n_2 \times n_3}$, a tensor $\mathcal{B}$ with size $\mathcal{B} \in \mathbf{R}^{n_2 \times n_4 \times n_3}$, and a tensor $\mathcal{C}$ with same size with tensor $\mathcal{B}$, the t-product of $\mathcal{A}$ and $\mathcal{B} + \mathcal{C}$ can be written as:

$$\begin{aligned} &\mathcal{A} * \mathcal{B} + \mathcal{A} * \mathcal{C} \\ &= \mathbf{Fold}(\mathbf{Circ}(\mathcal{A}) \times \mathbf{Vec}(\mathcal{B})) + \mathbf{Fold}(\mathbf{Circ}(\mathcal{A}) \times \mathbf{Vec}(\mathcal{C})) \\ &= \mathbf{Fold}(\mathbf{Circ}(\mathcal{A}) \times \mathbf{Vec}(\mathcal{B}) + \mathbf{Circ}(\mathcal{A}) \times \mathbf{Vec}(\mathcal{C})) \end{aligned} \tag{4.17}$$

Since the matrix standard multiplication satisfy the commutative law and it can be rewritten:

$$\begin{aligned} &\mathbf{Fold}(\mathbf{Circ}(\mathcal{A}) \times (\mathbf{Vec}(\mathcal{B}) + \mathbf{Vec}(\mathcal{C}))) \\ &= \mathbf{Fold}(\mathbf{Circ}(\mathcal{A}) \times (\mathbf{Vec}(\mathcal{B} + \mathcal{C}))) \\ &= \mathcal{A} * (\mathcal{B} + \mathcal{C}) \end{aligned} \tag{4.18}$$

4.6.2 PROOF OF THEOREM 4

For a tensor $\mathcal{A}$ and its L_2 norm is described as:

$$\begin{aligned}\langle \mathcal{A}, \mathcal{A} \rangle = \|\mathcal{A}\|_F^2 &= \mathbf{trace}((\mathcal{A} * \mathcal{A}^T)_{(:,:,1)}) \\ &= \mathbf{trace}((\mathcal{A}^T * \mathcal{A})_{(:,:,1)})\end{aligned} \tag{4.19}$$

where $(\mathcal{A}^T * \mathcal{A})_{(:,:,1)}$ is the frontal slice of $\mathcal{A}^T * \mathcal{A}$ and $(\mathcal{A} * \mathcal{A}^T)_{(:,:,1)}$ is the frontal slice of $\mathcal{A} * \mathcal{A}^T$. If $\mathcal{A}$ is an orthogonal tensor, the L_2 norm of $\mathcal{A} * \mathcal{B}$ can be denoted as:

$$\begin{aligned}\langle \mathcal{A} * \mathcal{B}, \mathcal{A} * \mathcal{B} \rangle &= \mathbf{trace}((\mathcal{A} * \mathcal{B})^T * (\mathcal{A} * \mathcal{B})) \\ &= \mathbf{trace}(\mathcal{B} * \mathcal{A}^T * \mathcal{A} * \mathcal{B}) \\ &= \langle \mathcal{B}, \mathcal{B} \rangle\end{aligned} \tag{4.20}$$

4.6.3 PROOF OF THEOREM 5

For a $\mathcal{X}$ with size $\mathbf{R}^{n_1 \times n_2 \times n_3}$, and it can be decomposed as $\mathcal{X} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T$. The L_2 norm of $\mathcal{X}$ can be written as:

$$\begin{aligned}\langle \mathcal{X}, \mathcal{X} \rangle &= \langle \mathcal{U} * \mathcal{S} * \mathcal{V}^T, \mathcal{U} * \mathcal{S} * \mathcal{V}^T \rangle \\ &= \mathbf{trace}([(\mathcal{U} * (\mathcal{S} * \mathcal{V}^T))^T * (\mathcal{U} * (\mathcal{S} * \mathcal{V}^T))]_{(:,:,1)}) \\ &= \langle \mathcal{S} * \mathcal{V}^T, \mathcal{S} * \mathcal{V}^T \rangle \\ &= \mathbf{trace}([(\mathcal{S} * \mathcal{V}^T) * (\mathcal{S} * \mathcal{V}^T)^T]_{(:,:,1)}) \\ &= \langle \mathcal{S}, \mathcal{S} \rangle\end{aligned} \tag{4.21}$$

Chapter 5

Conclusion and Future Work

5.1 Conclusion

In this thesis, we try to improve the efficiency of algorithm in image processing. We applied tensor method to data completion and adversarial attack technology. The contributions in the thesis prove that keep the original data in high-dimensional form and process these data by tensor method will increase the efficiency of data processing. The main conclusion of the thesis are summarized as follows:

- TT-TV model for data completion (Chapter 2): In this research, we present a new method to minimize TT rank with total variation model. In our method, we introduce nuclear norm regularization on TT rank which is the most effective surrogate for rank optimization for the global data structure. Meanwhile, we choose anisotropic TV as another regularization term, as anisotropic TV performs well in our model according to experimental results. In contrast with imposing the alternating linear scheme, nuclear norm regularization on TT-ranks is introduced in our method as it is an effective surrogate for rank optimization and our solution does not need to initialize and update tensor cores. VDT is introduced in this paper to reshape the MRI data to enhance the performance of the proposed algorithm. Based on the proposed optimization method, we divide the optimization problem into a series of sub-problems and then solve each problem. The results show that our method takes advantages in relative standard error (RSE), peak signal-to-noise ratio (PSNR), and structural similarity index (SSIM). It also concludes that our method achieves better accuracy compared with other methods based on the low-rank constraint

- Black-box adversarial attack(Chapter 3): 1. In this research, we first try the tensor method in adversarial attack technology. The attacked image is processed by tensor singular value decomposition, and we add the noise tensor in singular value diagonal tensor to create perturbation instead of changing the pixel of original image with the same size. We also impose restrictions on noise tensor to generate less L_2 norm of the image. We design a simple and fast algorithm to attack the targeted ML model by adding perturbation to images effectively. The noise tensor is randomly picked from pre-specified sets and then add or subtract it to the pre-acquired diagonal tensor. We show that without adding the perturbation to the original image, our method achieves better query efficiency compared with the state-of-the-art method. We also attack different CNN models to demonstrate the robustness of our method.

5.2 Future work

Though we have proposed several algorithms based on tensor method in the ML field, there are still remained problems to be explored in the future:

- We are going to combined other tensor model rank minimization such as tensor ring rank with total variation.
- We are going to design another kind of experiments to illustrate the performance of different methods and we try to complete another kind of data to estimate the performance of our proposed method.
- In the experiment, we found that attacks on different positions of the singular value tensor, the perturbation had different characteristics. In the next research, we will conduct research on this characteristic to improve the efficiency of the algorithm.
- We are going to try to design a defense technology to improve the robust of ML model.
- As we discussed in chapter 4, ϵ is a vital parameter to balance the query and L_2 norm of perturbation. In the next research, we will try if it is possible to find the optimal parameter.

Bibliography

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

[2] Alexander Novikov, Dmitrii Podoprikhin, Anton Osokin, and Dmitry P Vetrov. Tensorizing neural networks. In *Advances in Neural Information Processing Systems*, pages 442–450, 2015.

[3] Animashree Anandkumar, Rong Ge, Daniel Hsu, Sham M Kakade, and Matus Telgarsky. Tensor decompositions for learning latent variable models. *The Journal of Machine Learning Research*, 15(1):2773–2832, 2014.

[4] Heishiro Kanagawa, Taiji Suzuki, Hayato Kobayashi, Nobuyuki Shimizu, and Yukihiro Tagami. Gaussian process nonparametric tensor estimator and its minimax optimality. In *International Conference on Machine Learning*, pages 1632–1641, 2016.

[5] Guoxu Zhou, Qibin Zhao, Yu Zhang, Tülay Adalı, Shengli Xie, and Andrzej Cichocki. Linked component analysis from matrices to high-order tensors: Applications to biomedical data. *Proceedings of the IEEE*, 104(2):310–331, 2016.

[6] Fengyu Cong, Qiu-Hua Lin, Li-Dan Kuang, Xiao-Feng Gong, Piia Astikainen, and Tapani Ristaniemi. Tensor decomposition of eeg signals: a brief review. *Journal of neuroscience methods*, 248:59–69, 2015.

[7] Ledyard R Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3):279–311, 1966.

[8] RA Harshman. Foundations of the parafac procedure: Models and conditions for an" explanatory" multi-mode factor analysis. *UCLA Working Papers in Phonetics*, 16:1–84, 1970.

[9] JW Carlson and T Minemura. Imaging time reduction through multiple receiver coil data acquisition and image reconstruction. *Magnetic resonance in medicine*, 29(5):681–687, 1993.

[10] Yudong Zhang, Bo Peng, Shuihua Wang, Yu-Xiang Liang, Jiquan Yang, Kwok-Fai So, and Ti-Fei Yuan. Image processing methods to elucidate spatial characteristics of retinal microglia after optic nerve transection. *Scientific reports*, 6:21816, 2016.

[11] Zhi-Pei Liang, FE Boada, RT Constable, EM Haacke, PC Lauterbur, and MR Smith. Constrained reconstruction methods in mr imaging. *Rev Magn Reson Med*, 4(2):67–185, 1992.

[12] Justin P Haldar. Low-rank modeling of local k-space neighborhoods (loraks) for constrained mri. *IEEE transactions on medical imaging*, 33(3):668–681, 2013.

[13] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. Intriguing properties of neural networks. *arXiv preprint arXiv:1312.6199*, 2013.

[14] Battista Biggio, Igino Corona, Davide Maiorca, Blaine Nelson, Nedim Šrndić, Pavel Laskov, Giorgio Giacinto, and Fabio Roli. Evasion attacks against machine learning at test time. In *Joint European conference on machine learning and knowledge discovery in databases*, pages 387–402. Springer, 2013.

[15] Minhao Cheng, Thong Le, Pin-Yu Chen, Jinfeng Yi, Huan Zhang, and Cho-Jui Hsieh. Query-efficient hard-label black-box attack: An optimization-based approach. *arXiv preprint arXiv:1807.04457*, 2018.

[16] Anish Athalye and Nicholas Carlini. On the robustness of the cvpr 2018 white-box adversarial example defenses. *arXiv preprint arXiv:1804.03286*, 2018.

[17] Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. Towards deep learning models resistant to adversarial attacks. *arXiv preprint arXiv:1706.06083*, 2017.

[18] Chuan Guo, Jared S Frank, and Kilian Q Weinberger. Low frequency adversarial perturbation. *arXiv preprint arXiv:1809.08758*, 2018.

[19] Wieland Brendel, Jonas Rauber, and Matthias Bethge. Decision-based adversarial attacks: Reliable attacks against black-box machine learning models. *arXiv preprint arXiv:1712.04248*, 2017.

[20] Andrzej Cichocki, Rafal Zdunek, Anh Huy Phan, and Shun-ichi Amari. *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*. John Wiley & Sons, 2009.

[21] Andrzej Cichocki, Danilo Mandic, Lieven De Lathauwer, Guoxu Zhou, Qibin Zhao, Cesar Caiafa, and Huy Anh Phan. Tensor decompositions for signal processing applications: From two-way to multiway component analysis. *IEEE signal processing magazine*, 32(2):145–163, 2015.

[22] Ivan V Oseledets. Tensor-train decomposition. *SIAM Journal on Scientific Computing*, 33(5):2295–2317, 2011.

[23] Guifré Vidal. Efficient classical simulation of slightly entangled quantum computations. *Physical review letters*, 91(14):147902, 2003.

[24] Guifré Vidal. Efficient simulation of one-dimensional quantum many-body systems. *Physical review letters*, 93(4):040502, 2004.

[25] Thomas Mach. Computing inner eigenvalues of matrices in tensor train matrix format. In *Numerical Mathematics and Advanced Applications 2011*, pages 781–788. Springer, 2013.

[26] Namgil Lee and Andrzej Cichocki. Estimating a few extreme singular values and vectors for large-scale matrices in tensor train format. *SIAM Journal on Matrix Analysis and Applications*, 36(3):994–1014, 2015.

[27] Sebastian Holtz, Thorsten Rohwedder, and Reinhold Schneider. The alternating linear scheme for tensor optimization in the tensor train format. *SIAM Journal on Scientific Computing*, 34(2):A683–A713, 2012.

[28] Lars Grasedyck, Melanie Kluge, and Sebastian Kramer. Variants of alternating least squares tensor completion in the tensor train format. *SIAM Journal on Scientific Computing*, 37(5):A2424–A2450, 2015.

[29] Curt Da Silva and Felix J Herrmann. Optimization on the hierarchical tucker manifold–applications to tensor completion. *Linear Algebra and its Applications*, 481:131–173, 2015.

[30] Holger Rauhut, Reinhold Schneider, and Željka Stojanac. Tensor completion in hierarchical tensor representations. In *Compressed sensing and its applications*, pages 419–450. Springer, 2015.

[31] Johann A Bengua, Ho N Phien, Hoang Duong Tuan, and Minh N Do. Efficient tensor completion for color image and video recovery: Low-rank tensor train. *IEEE Transactions on Image Processing*, 26(5):2466–2479, 2017.

[32] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992.

[33] Tatsuya Yokota and Hidekata Hontani. Simultaneous visual data completion and denoising based on tensor rank and total variation minimization and its primal-dual splitting algorithm. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3732–3740, 2017.

[34] Tatsuya Yokota and Hidekata Hontani. Simultaneous tensor completion and denoising by noise inequality constrained convex optimization. *IEEE Access*, 7:15669–15682, 2019.

[35] Tatsuya Yokota, Qibin Zhao, and Andrzej Cichocki. Smooth PARAFAC decomposition for tensor completion. *IEEE Transactions on Signal Processing*, 64(20):5423–5436, 2016.

[36] Ching-Yun Ko, Kim Batselier, Lucas Daniel, Wenjian Yu, and Ngai Wong. Fast and accurate tensor completion with total variation regularized tensor trains. *IEEE Transactions on Image Processing*, 29:6918–6931, 2020.

[37] Antonin Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical imaging and vision*, 20(1):89–97, 2004.

[38] Wenfei Cao, Yao Wang, Jian Sun, Deyu Meng, Can Yang, Andrzej Cichocki, and Zongben Xu. Total variation regularized tensor rpca for background subtraction from compressive measurements. *IEEE Transactions on Image Processing*, 25(9):4075–4090, 2016.

[39] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122, 2011.

[40] Ji Liu, Przemyslaw Musialski, Peter Wonka, and Jieping Ye. Tensor completion for estimating missing values in visual data. *IEEE transactions on pattern analysis and machine intelligence*, 35(1):208–220, 2012.

[41] Jose I Latorre. Image compression and entanglement. *arXiv preprint quant-ph/0510031*, 2005.

[42] Yipeng Liu, Zhen Long, and Ce Zhu. Image completion using low tensor tree rank and total variation minimization. *IEEE Transactions on Multimedia*, 2018.

[43] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[44] Xutao Li, Yunming Ye, and Xiaofei Xu. Low-rank tensor completion with total variation for visual data inpainting. In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[45] Yangyang Xu, Ruru Hao, Wotao Yin, and Zhixun Su. Parallel matrix factorization for low-rank tensor completion. *arXiv preprint arXiv:1312.1254*, 2013.

[46] Longyu Jiang, Runguo He, Jie Liu, Yang Chen, Jiasong Wu, Huazhong Shu, and Jean-Louis Coatrieux. Phase-constrained parallel magnetic resonance imaging reconstruction based on low-rank matrix completion. *IEEE Access*, 6:4941–4954, 2017.

[47] Longhao Yuan, Qibin Zhao, Lihua Gui, and Jianting Cao. High-order tensor completion via gradient-based optimization under tensor train format. *Signal Processing: Image Communication*, 73:53–61, 2019.

[48] Zemin Zhang, Gregory Ely, Shuchin Aeron, Ning Hao, and Misha Kilmer. Novel methods for multilinear data completion and de-noising based on tensor-SVD. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3842–3849, 2014.

[49] Gregory Ely, Shuchin Aeron, and Eric L Miller. Exploiting structural complexity for robust and rapid hyperspectral imaging. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2193–2197. IEEE, 2013.

[50] John Wright, Arvind Ganesh, Kerui Min, and Yi Ma. Compressive principal component pursuit. *Information and Inference: A Journal of the IMA*, 2(1):32–68, 2013.

[51] Negin Entezari and Evangelos E Papalexakis. Tensorshield: Tensor-based defense against adversarial attacks on images. *arXiv preprint arXiv:2002.10252*, 2020.

[52] SM Moosavi-Dezfooli, A Fawzi, P Frossard, and Deepfool. A simple and accurate method to fool deep neural networks. In *Proceedings of the CVPR*, pages 2574–2582.

[53] Seyed-Mohsen Moosavi-Dezfooli, Alhussein Fawzi, Omar Fawzi, and Pascal Frossard. Universal adversarial perturbations. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1765–1773, 2017.

[54] Chuan Guo, Jacob Gardner, Yurong You, Andrew Gordon Wilson, and Kilian Weinberger. Simple black-box adversarial attacks. In *International Conference on Machine Learning*, pages 2484–2493. PMLR, 2019.

[55] Andrew Ilyas, Logan Engstrom, Anish Athalye, and Jessy Lin. Black-box adversarial attacks with limited queries and information. In *International Conference on Machine Learning*, pages 2137–2146. PMLR, 2018.

[56] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[57] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[58] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.