Tensor Network for Image Classification

Guoming Chen,Qiang Chen,Xiongyong Zhu,Yiqun Chen,Zeduo Yuan
Department of Computer Science
Guangdong University of Education
Guangdong 510303,China
*Corresponding Author: chenguoming@gdei.edu.cn

2020 8th International Conference on Digital Home (ICDH) | 978-1-7281-9234-5/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICDH51081.2020.00031

***Abstract*—In this paper we propose a feature representation and dimensionality reduction method in tensorflow and tensor network framework for binary image classification. The chaology image data is properly down scaled with some techniques: restricted rank tensor approximations, pulling through with QR decomposition and multi-stage tensor decomposition. In a tensor network classifier framework, performance among the three tensor internal features are compared with each other. The visualization of two-dimensional electronic cloud about chaotic scale classification results on chaology image datasets show that the restricted rank tensor method and multi-stage tensor decompositions method have a similar high accuracy and excellent dimensionality reduction ability performance, their separability performance is better than QR decomposition method.**

Keywords*—*Tensor Network; Chaology image; Binary Classification;

I. INTRODUCTION

Deep sparse tensor filtering network[1] can be used to extract discriminative features from synthetic aperture radar(SAR) images for classification. SAR image is organized into a data tensor then it is developed by dimension-inseparable geometric filters from a least squares support vector machine, finally the constructed sparse tensor filters are cascaded to a deep network to extract the features for classification.

Principal component analysis network is good at 2-D images classification. However, feature extraction is difficult when dealing with multi-dimensional images, for the spatial relationships inside the structures of images may be ignored. A multilinear principal component analysis network[2] is to extract the high-level semantic features from high dimensional images. The extracted features minimize the intraclass invariance of tensor by spatial relationships within images.

Tensor networks[3] is a dimension reduction and data optimization method. Low-rank tensor decomposition is to obtain super-compressed higher-order pattern data or parameters. It is a hot research issue in machine learning and data mining. It can approximate low-rank tensor, contract core tensors through graphic illustrations and thus to alleviate the curse of dimensionality in regression, classification, eigenvalue decomposition and deep neural networks optimization. Tensor train and hierarchical tucker decompositions are effective tools in tensor networks.

Tensor network is also good at approximating many-body quantum states with 1D chain of tensors. It is efficient in capturing the local correlations between neighboring subsystems and has the similar structure with artificial neural networks. Long chain of tensors suffer from vanishing gradients. The long-chain tensor network[4] is decomposed into short chains with a stable stochastic gradient descent convergence. The short-chain has the advantages of being robust to network initializations. It can improve classification accuracy on mnist dataset with less network parameters.

Higher order data and high dimensionality[5] which is referred to tensor has a widely appliance in chemometrics, hyperspectral imaging, neuroimaging and biometrics. It reformats tensor data as vectors or matrices and then resorts to dimensionality reduction methods such as principal component analysis (PCA). PCA dimensionality reduction, low-rank tensor approximation and supervised learning has a perfect harmonious integration to make it easier to discover hidden components in feature extraction.

Topology of tensor networks is used for multimodal data[6] on multi-label classification problems assume that the data is in the form of nonnegatively constrained multi-way arrays. A tensor-train model is trained for extracting low-rank and nonnegative 2D features for classification problems.

Extracting features from incomplete tensors is also a challeng problem in machine learning and pattern recognition. One can recover the missing entries by tensor completion only on estimation rather than feature extraction. Low-rank tensor decomposition with feature variance maximization (TDVM)[7] is proposed for unsupervised learning. By orthogonal Tucker and CP decompositions, low-dimensional features are obtained from core tensors of the Tucker model and the weight vectors of the CP model. TDVM can maximize feature variance and estimate the missing entries by integrating regularization into low-rank tensor approximation and make better use of the alternating direction method of multipliers with the block coordinate descent tactics.

A deep belief network(DBN)[8] and tensor dimensionality reduction is design for semi-supervised polarimetric SAR image classification which uses multilinear principle component analysis to reduce the dimension of tensor form data and then use DBN for classification. From neighborhood information of each pixel of polarimetric SAR data, multiple features of abundant information and spatial structure can also obtained to get appropriate classification accuracy.

Tensor Networks can represent sets of correlated data in some common applications which include: quantum many-body systems, where tensor networks encode the coefficients of the state wave function and ensembles of microstates. It can also be used in data mining, where tensor networks give a nice expression to multi-dimensional data especially in signal processing, pattern recognition,etc. It has advantages about

978-1-7281-9234-5/20/$31.00 ©2020 IEEE
DOI 10.1109/ICDH51081.2020.00031

data structure representation with a greate compression ratio.

There are many tensor network methods include MPS, TTN, MERA and PEPS in many-body problem, we start with a Hamiltonian where the subsystems interact with each other inside the model, and it tries to find the lowest energy eigenstate and ground state of the system. Tensor networks acts as initiation for quantum states. In order to build a bridge between machine learning and quantum physics, find an appropriate tensor network for practical application and utilize available geometry shape and physical properties to benefit machine learning. This may require further study to initialize the data in the network, through iterations, the parameters in the tensors are varied so as to better approximate to the ground state of the quantum system with energy minimization and Euclidean time evolution tactics.

II. Feature Representation Schemes in Tensor Network

Tensor networks potentially and naturally demand a characterized structure, especially with correlations. The diagrammatic notation with a visual representation of networks structure gives an easily understanding of data distribution, so that manipulations can also be possibly performed in parallel data processing. In these schemes, sometimes a small parts of tensor network, has capability of analyzing data without specific knowledge about the underlying data set or what it does means, even with noisy or missing data.

A. Restricted rank tensor approximations

Given some tensor A and its indices, decomposition rank r is the minimal internal dimension of A which is product of tensors, A =$B \cdot C$. The decomposition rank r of tensor A corresponds to the number of non-zero singular values in the corresponding SVD.

It is possible to find the optimal restricted rank approximation to a tensor by the singular value decomposition. Given a tensor A with decomposition rank r correspond to some specified divide indices, there exist an equivalent sized tensor B of reduced rank, $\chi < r$, that can approximates A with the difference ϵ minimized by applying SVD of A to trim the smallest singular values and truncate the U and V matrices. Fig.1 is the approximation order-5 tensor A with an order-5 tensor B that has rank χ =8 by truncated SVD.

The accuracy on truncation error ϵ of the optimal restricted rank approximation is defined as the number of singular values greater than or equal to some value in the decompose process. If the spectrum of singular values is sharply reduce by truncated. We observe the effect of feature extraction after truncation.

B. Setting a Orthogonality center by Pulling through with QR

Fig.2 is an example of a tree tensor network T:A,B,C,$\cdots$. The center root node is A, it has a set of distinct branches from tensor A. There are four branches from the order-4 tensor A, the branch forms an isometry between its open indices and the index connected to tensor A. Tensor network is to fix the gauge to work correctly. The gauge degree of freedom is fixed

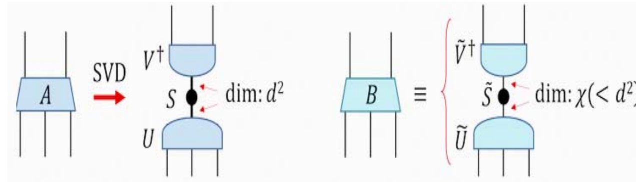

Fig. 1: Restricted rank tensor approximations

when a center of orthogonality is created. If we transform every tensor inside a branch into aisometry, then the entire branch becomes an isometry. It starts with guiding each index with an arrow pointing to the center tensor A. QR decomposition on the tensor is between incoming and outgoing arrows. Compute the orthogonal Q part of the QR decomposition and absorb the R matrix into the tensor with the outgoing arrow. Repeat the process when all tensors are isometric.

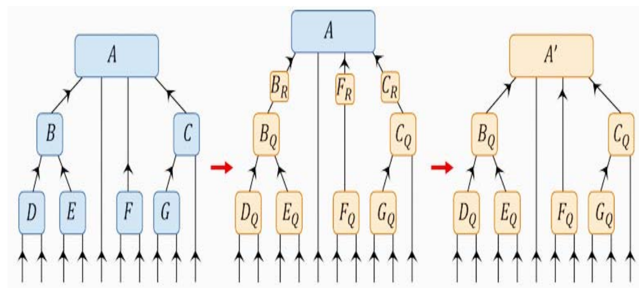

Fig. 2: Pulling through with QR

C. Multi-stage tensor decompositions

Multiple tensor decompositions is to fix the gauge degrees of freedom with a canonical form for tensor networks. In order to obtain the network approximation to a tensor we shall employ a multi-stage decomposition: a sequence of single tensor decompositions via the SVD. This is the opposite procedure to the contraction. The tensor to be decomposed is a center of orthogonality, which guarantees the truncation error to be minimized.Given a many-index tensor H, it is to decompose it into a network T by some geometry with tensors that minimizes the difference, for some fixed dimension χ of the internal indices in T.

Fig.3 shows that a single tensor H_0 is decomposited into the network T. A tensor H_k is split by a truncated SVD with rank χ into a product of three tensors U_k, S_k, V_k. The matrix of singular S_k is absorbed into the tensor that is to be decomposed at the next step, so it is a center of orthogonality when all other tensors in the network are isometric. After several iterations we get the network geometry and check the accuracy of the decomposition. When contracting a tensor network, we choose the sequence of decompositions in a multi-stage decomposition by index truncations to obtain performance advantages. This procedure continues absorbing the singular weights into the new desired orthogonality center. The truncation error at each individual decomposition step of multi-stage decompositions is minimized.

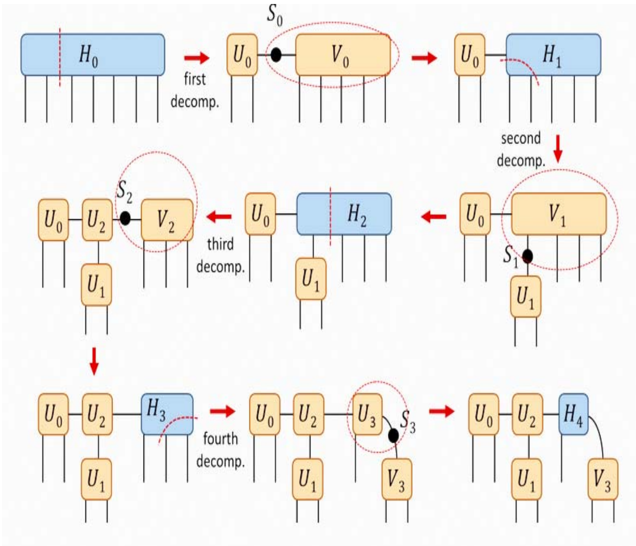

Fig. 3: Multi-stage tensor decompositions

III. Experiment and Result

In this section, we will apply the above-mentioned three schemes inside tensor networks on a fully connected neural network and tensor network layer instead of the normal dense weight matrix. The network has nearly ten times fewer parameters when compared with traditional ones. We try to see which type of scheme is appropriate for image classification pattern. The experiments were carried out in an environment of tensorflow and open-source tensor network library, on a chaology image datasets for binary image classification which is to measure the scale of chaos to a certain extent. The images which is shown in Fig.4 are chaology images on different scale levels of chaos. The original chaology image dataset contains 12000 images. But only part of them can be used in experiment due to the limitation of memory capacity.

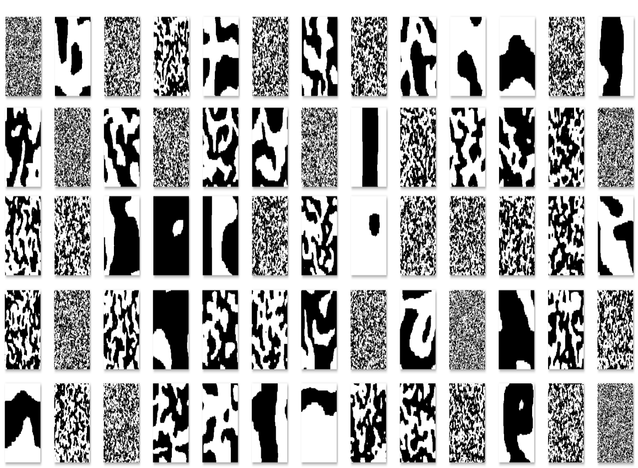

Fig. 4: Chaology Image

Fig.3 is the multi-stage tensor decompositions procedure where d is the local dimension and χ is the max internal dimension. We reshape the image $Horizontal \times vertical \times 3$ as H_0 where $H_0 \in R^{d_1 \times d_2 \times d_3 \times d_4, \times d_5, \times d_6 \times d_7}$. Then SVD of H_0 is denoted as:

$$H_0 = U * S * V \tag{1}$$

after the fourth decomposition, H_0 recovered = ncon(U0,U1,U2,V3,H4,[-1,-2,1],[-3,-4,2],[1,2,3],[-6,-7,4],[3,-5,4]); then we get $S_{trunated}$ after the fourth decomposition. We arrange the diagonal elements of $S_{trunated}$ in order and take out the first two as salient features after that we project salient features and feed them into a fully connected neural network and tensor network framework and then put the result into a color plot and also scatter the training points, we get the visualization of binary image recognition two dimensional electronic cloud image of output results which can be shown in Fig.5.

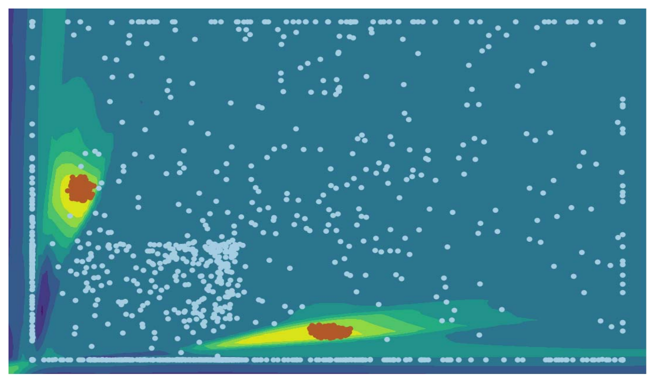

Fig. 5: Multi-stage Tensor Decompositions Visualization

The Restricted rank tensor approximations scheme can be shown in Fig.1, where χ is the rank. We reshape the image $Horizontal \times vertical \times 3$ as A where $A \in R^{d_1 \times d_2 \times d_3 \times d_4, \times d_5}$. Then trunated SVD of A which is denoted as:

$$A = U * S * V \tag{2}$$

after the trunated SVD, A recovered = ncon($U_{trunated}$, $S_{trunated}$,$V_{trunated}$,[-1,-2,-3,1],[1,2],[-4,-5,2]);then we take the diagonal elements $S_{trunated}$ as features. We arrange the diagonal elements of $S_{trunated}$ in order and take out the first two as salient features after that we project salient features and feed them into a fully connected neural network and tensor network framework and then put the result into a color plot and also scatter the training points, we get the visualization of binary image recognition two dimensional electronic cloud image of output results which can be shown in Fig.6.

Pulling through with QR decomposition method is shown in Fig.2. Its main process is : 1. define tensors, 2.through iterative use of QR decomposition,3.check if initial and final networks contract to the same tensor, connectlist = [3,-5,4,5],[1,2,3],[6,-10,5],[-1,-2,1],[-3,-4,2],[-6,-7,4],[-8,-9,6]; H_0 = ncon(A,B,C,D,E,F,G,connectlist); From the QR decomposition, we can get BR,CR,DR,ER,FR,GR salient features from each of the images, after that we project salient features and feed them into a fully connected neural network and tensor

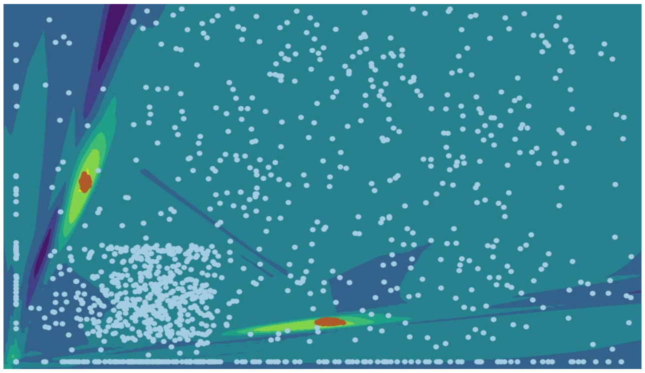

Fig. 6: Restricted Rank Tensor Visualization

network framework and then put the result into a color plot and also scatter the training points, we get the visualization of binary image recognition two dimensional electronic cloud image of output results which can be shown in Fig.7.

Fig.5 is the multi-stage tensor decompositions visual dimension reduction electronic cloud. Fig.6 is the restricted rank tensor approximations visual dimension reduction electronic cloud. And Fig.7 is the pulling through with QR visual dimension reduction electronic cloud. In a tensor network classifier framework, performance among the three tensor internal features are compared with each other. The visualization of two-dimensional electronic cloud chaotic scale classification results on chaology image datasets show that the restricted rank tensor contraction scheme and multi-stage tensor decompositions contraction scheme have a similar high accuracy and excellent dimensionality reduction ability performance, their separability performance is better than pulling through with QR decomposition contraction scheme.

In addition, in the historical development process of tensor network contraction, many new technologies have emerged, such as tree decomposition method, which is aim to minimize the highest cost tree node. The hypergraph formulation method, which is to simplify diagonal gates. Dynamic slicing method is to implement index slicing by parallelization. Furthermore, stem optimization method is to reduce parallelization overhead on stem.

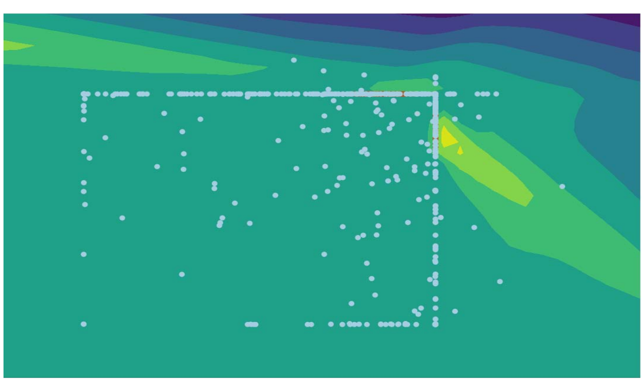

Fig. 7: QR Decomposition Visualization

IV. To find a good contraction scheme

The image data is properly down scaled with three basic contraction scheme : restricted rank tensor approximations, pulling through with QR decomposition and multi-stage tensor decomposition before it is fed and projected into a tensor network and neural network framework. Among these contraction schemes, one of the main tasks is to decide which kind of contraction can play a leading role and make better advantages of contract that we call these nodes the dominating nodes after the contraction. Then we can analyze and compress data with multi-way relationship in machine learning for feature extraction and dimensionality reduction aim to help to solve the curse of dimensionality in decomposing data and is able to perform dimensionality reduction, e.g. in classification tasks which is scalable and possible.

The truncation error ϵ in the optimal restricted rank approximation has something to do with the discarded singular values. The decomposition rank r is the minimal internal dimension that some tensor can be expressed as a product of tensors. We want to find equivalent sized tensor of reduced rank $\chi < r$, that best approximates tensor. The factorization minimizes the truncation error ϵ from the SVD of tensor. Fig.8 demonstrate that with the rank χ increasing, truncation error ϵ first increase and then decrease. Truncation error ϵ gets the maximum value when χ=2. Increasing

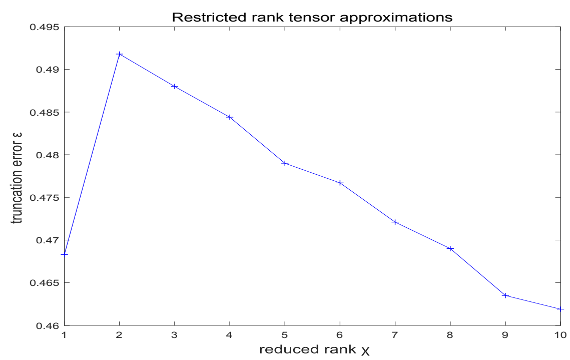

Fig. 8: Truncation Error in Restricted rank tensor approximations

Fig. 9 is the error of the decomposition in pulling through with QR.Through iterative of the QR decomposition, each tensor in the network except the center is given as the ’Q’ part of a QR decomposition is isometric. Fig.9 illustrates that error of the decomposition will rise first and then decrease with the increasing of dimension d, and the error gets the maximum value when d=6, the initial and final networks will contract to the same tensor.

Next we will discuss the truncation error in multi-stage tensor decompositions in Fig.10. Any many-index tensor H can be accurately decomposed into a network of tensors which minimizes the difference from the origin ones, given some fixed dimension χ of the internal indices, the tensor at each decomposed step is a center of orthogonality that the global truncation error is minimized, a tensor is split by a truncated SVD with rank χ to a product of three tensors, the truncation

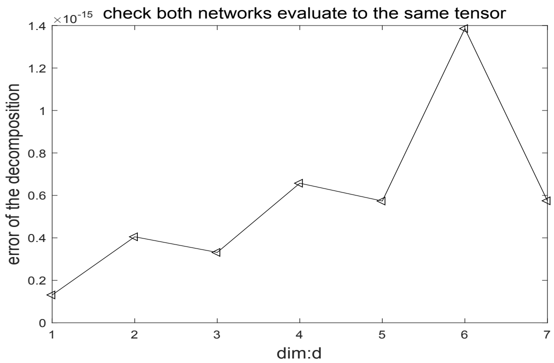

Fig. 9: Error of the Decomposition in Pulling through with QR

error at each decomposition step is minimized. And cumulative error from the sequence of decompositions is small. In Fig.10, if we fix local dimension d to 5, and we observe that as the max internal dimension χ increases, the truncation error is decreasing.

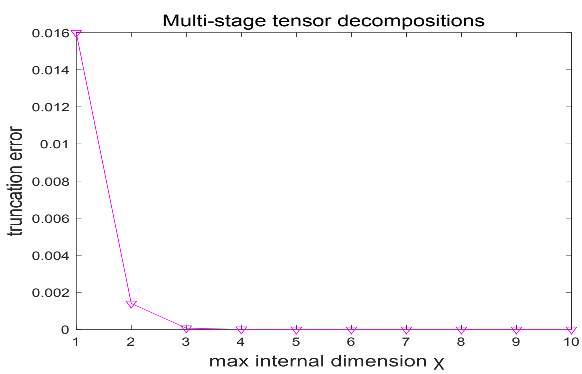

Fig. 10: Truncation Error in Multi-stage tensor decompositions

From three modes of contraction scheme, we find that in pulling through with QR the error is the smallest, the second is multi-stage tensor decompositions, the maximum error is the restricted rank tensor approximations. So except for the ways to represent the image structure, it has the least error from the perspective that it should be considered in the simulation to fit the experimental data better after contraction,pulling through with QR scheme is apt to find the dominating nodes after the contraction in the future.

V. CONCLUSIONS

In this paper we propose a feature representation and dimensionality reduction method in tensorflow and tensor network framework for binary image classification. The chaology image data is properly down scaled with some techniques: restricted rank tensor approximations, pulling through with QR decomposition and multi-stage tensor decomposition before it is fed and projected into a tensor network classifier framework. Performance among the three tensor internal features are compared with each other. The visualization of electronic cloud chaotic scale classification results on chaology image datasets show that the restricted rank tensor method and multi-stage tensor decompositions method have similar high accuracy and excellent dimensionality reduction ability performance, their separability performance is better than QR decomposition method.

ACKNOWLEDGMENT

This work is partially supported by National Natural Science Foundation of China (No.61772140),Natural Science Foundation of Guangdong Province(No.2018A0303130169), the Opening Project of Guangdong Province Key Laboratory of Big Data Analysis and Processing at the Sun Yat-sen University(No.201902).

REFERENCES

[1] S.Yang, M. Wang, Z. Feng, Z. Liu and R. Li, ”Deep Sparse Tensor Filtering Network for Synthetic Aperture Radar Images Classification,” in IEEE Transactions on Neural Networks and Learning Systems, vol. 29, no. 8, pp. 3919-3924, Aug. 2018.

[2] J. Wu, S. Qiu, R. Zeng, V. Kong, L. Senhadji and H. Shu, ”Multilinear Principal Component Analysis Network for Tensor Object Classification,” in IEEE Access, vol. 5, pp. 3322-3331, 2017.

[3] Andrzej Cichocki; Namgil Lee; Ivan Oseledets; Anh-Huy Phan; Qibin Zhao; Danilo P. Mandic, ”Tensor Networks for Dimensionality Reduction and Large-scale Optimization: Part 2 Applications and Future Perspectives,2017.

[4] Y. W. Chen, K. Guo and Y. Pan, ”Robust supervised learning based on tensor network method,” 2018 33rd Youth Academic Annual Conference of Chinese Association of Automation (YAC), Nanjing, 2018, pp. 311-315.

[5] C. Roberts et al. TensorNetwork: A Library for Physics and Machine Learning, arXiv:1905.01330

[6] A. Zare, A. Ozdemir, M. A. Iwen and S. Aviyente, ”Extension of PCA to Higher Order Data Structures: An Introduction to Tensors, Tensor Decompositions, and Tensor PCA,” in Proceedings of the IEEE, vol. 106, no. 8, pp. 1341-1358, Aug. 2018.

[7] R. Zdunek and K. Fona, ”Nonnegatively Constrained Tensor Network for Classification Problems,” 2019 Eleventh International Conference on Ubiquitous and Future Networks (ICUFN), Zagreb, Croatia, 2019, pp. 218-223.

[8] Q. Shi, Y. Cheung, Q. Zhao and H. Lu, ”Feature Extraction for Incomplete Data Via Low-Rank Tensor Decomposition With Feature Regularization,” in IEEE Transactions on Neural Networks and Learning Systems, vol. 30, no. 6, pp. 1803-1817, June 2019.

[9] B. Hou, X. Guo, W. Hou, S. Wang, X. Zhang and L. Jiao, ”PolSAR Image Classification Based on DBN and Tensor Dimensionality Reduction,” IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, 2018, pp. 8448-8450.

[10] E. M.Stoudenmire, D. J .Schwab. Supervised Learning with Quantum-Inspired Tensor Networks. 2016.

[11] V. Murg, F. Verstraete, O. Legeza, and R. M. Noack Simulating strongly correlated quantum systems with tree tensor networks, Phys. Rev. B 82, 205105 (2010), arXiv:1006.3095

[12] G. Evenbly, Gauge fixing, canonical forms and optimal truncations in tensor networks with closed loops, Physical Review B, 98(8), 085155 (2018).

[13] Z. Sun et al., Generative Tensor Network Classification Model for Supervised Machine Learning, arXiv: 1903.10742

[14] Y. Levine, O. Sharir, N. Cohen, A. Shashua, Quantum entanglement in deep learning architectures, Physical Review Letters, 122(6), 065301 (2019), arXiv:1803.09780

[15] W. Huggins, P. Patel, K. B. Whaley, E. M. Stoudenmire, Towards Quantum Machine Learning with Tensor Networks,Quantum Science and Technology, Volume 4, 024001 (2019), arXiv:1803.11537

[16] E. M. Stoudenmire Learning Relevant Features of Data with Multi-scale Tensor Networks, Quantum Science and Technology, Volume 3, 034003 (2018), arXiv: 1801.00315

[17] A. Cichocki, N. Lee, I. V. Oseledets, A.-H. Phan, Q. Zhao, D. Mandic, Low-Rank Tensor Networks for Dimensionality Reduction and Large-Scale Optimization Problems: Perspectives and Challenges PART 1, Foundations and Trends in Machine Learning, 9(4-5), 249-429. arXiv: 1609.00893

[18] A. Cichocki Tensor Networks for Big Data Analytics and Large-Scale Optimization Problems, arXiv: 1407.3124

[19] Z. Zhang, F. Yin, J. Liu, D. W. K. Wong, N. M. Tan, B. H. Lee, J. Cheng, and T. Y. Wong,ORIGA(-light): An Online Retinal Fundus Image Database for Glaucoma Analysis and Research, IEEE, 2010.

[20] Luo, Z., Feng, E. , Zhang, J. Numerical realization of the conditions of Max Nthers residual intersection theorem. Appl. Math. J. Chin. Univ. 29, 481502 (2014)

[21] Liu Li,Wang Ruo-Mei,Luo Xiao-Nan.Realistic Cloth Simulation Based on Geometrical Measurement and Deformation.Journal of Software,2015,26(7):1785-1799

[22] Ri-Sheng Liu, Cai-Sheng Mao, Zhi-Hui Wang, Hao-Jie Li. Blind Image Deblurring via Adaptive Optimization with Flexible Sparse Structure Control.Journal of Computer Science and Technology, 2019, 34(3): 609-621.

[23] Liang Zhao,Zhikui Chen,Yi Yang,ICFS Clustering With Multiple Representatives for Large Data,IEEE Transactions on Neural Networks and Learning Systems PP(99):1-11

[24] Shixia Liu, Changjian Chen, Yafeng Lu, Fangxin Ouyang, Bin Wang IEEE Transactions on Visualization and Computer Graphics. 25(1): 235-245, 2019.