Supplemental Material of "Unsupervised Recognition of Informative Features via Tensor Network Machine Learning and Quantum Entanglement Variations"

Sheng-Chen Bai,[1,*] Yi-Cheng Tang,[1,*] and Shi-Ju Ran[1,†]

[1]*Department of Physics, Capital Normal University, Beijing 100048, China*

(Dated: September 16, 2022)

In this supplemental material, we provide more details about our recognition scheme based on the variations of entanglement entropy. We show the variations of entanglement entropy and the average given several specific samples in the MNIST and fashion-MNIST datasets. How the bond dimensions affect the variations of entanglement entropy is also demonstrated from an concrete example. The classification accuracies using just the selected features are compared between our scheme and the discriminative matrix product state method. By generalizing to the multi-qubit measurement, we show that our method can recognize single-pixel noises unsupervisedly. The feature selection by training the MPS with the samples from multiple classes are also discussed.

I. VARIATIONS OF ENTANGLEMENT ENTROPY BY MEASUREMENT

In Fig. 1, we show how the entanglement entropy (EE) varies after measuring on a specific qubit. The first column shows four images taken from the MNIST and fashion-MNIST datasets as examples. We evaluate $\langle \delta S\rangle_{m'}$ ($m' = 0,\ldots,M$) by Eq. (A2) according to these images, respectively, and find the pixels with the algebraically smallest and largest variations $m_{\min} = \arg\min_{m'}\langle \delta S\rangle_{m'}$ and $m_{\max} = \arg\max_{m'}\langle \delta S\rangle_{m'}$. The corresponding variations of EE ($dS_m = S'_m - S_m$) are demonstrated in the second and third columns, respectively. The position of the measured qubit is marked by a green symbol with dash line.

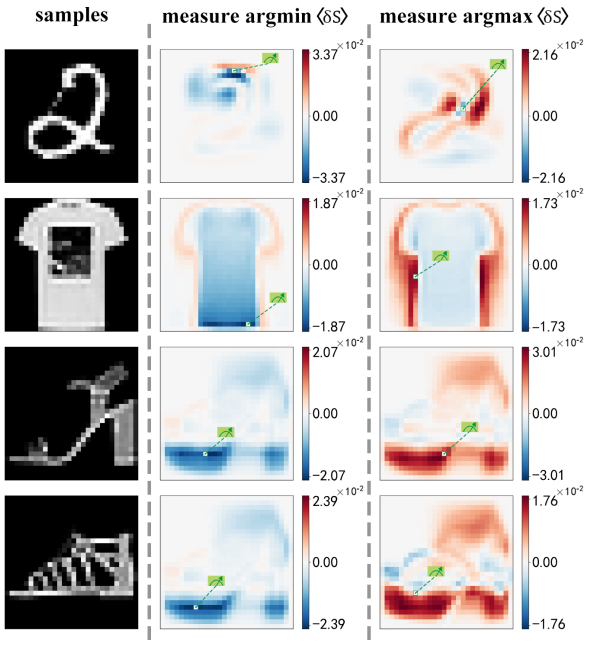

FIG. 1. (Color online) Four images from the MNIST and fashion-MNIST datasets (first column), and the variations of the entanglement entropy $dS_m = S'_m - S_m$ by taking $m = \arg\min_{m'}\langle \delta S\rangle_{m'}$ (second column) and $\arg\max_{m'}\langle \delta S\rangle_{m'}$ (third column). The green symbols with dash line show the positions of the measure qubits.

For the considered datasets, we have $\langle \delta S\rangle_{m_{\min}} < 0$. Comparing with the EE of the MPS (see, e.g., Fig.4 in the main text), the measurement on this qubit according to the value of the corresponding pixel will reduce the EE in the informative area. Remind that the informative area is defined by the pixels with large EE from the unmeasured MPS. Meanwhile, we also obverse certain positive variations that tend to locate at the edges of the informative area. By measuring on the $m_{\max}$-th qubit, the qubits whose EE increases in general tend to locate at the edges of the informative area.

II. BENCHMARK ON TESTING SET

Our scheme can be generalized to recognize the informative features and critical minorities of the samples that the generative TN has not learnt, e.g., those in the testing set. Fig. 2 show six testing images from the MNIST and fashion-MNIST datasets as the examples to show the $\langle \delta S\rangle_{m'}$ by implementing measurements. Be aware that we do not have any prior information on labeling the pixels even for the training set. The $\langle \delta S\rangle_{m'}$ can be considered as the labels of the pixels that characterize importance to the given sample. By eyes one could recognize the distinct shapes in the testing images from $\langle \delta S\rangle_{m'}$.

III. ROBUSTNESS WITH DIFFERENT HYPER-PARAMETERS

The recognition of the informative features by the EE variations is robust to the changes the values of hyper-parameters. We generalize the feature map given in Eq. (A1) to the following form, where the s-th element of the vector $\boldsymbol{v}$ from a given feature x satisfies

$$x \to v_s = \sqrt{\binom{d-1}{s-1}} \cos\left(\frac{\theta\pi}{2}x\right)^{d-s} \sin\left(\frac{\theta\pi}{2}x\right)^{s-1}, \quad (1)$$

* These two authors have contributed equally to this work.
† Corresponding author. Email: sjran@cnu.edu.cn

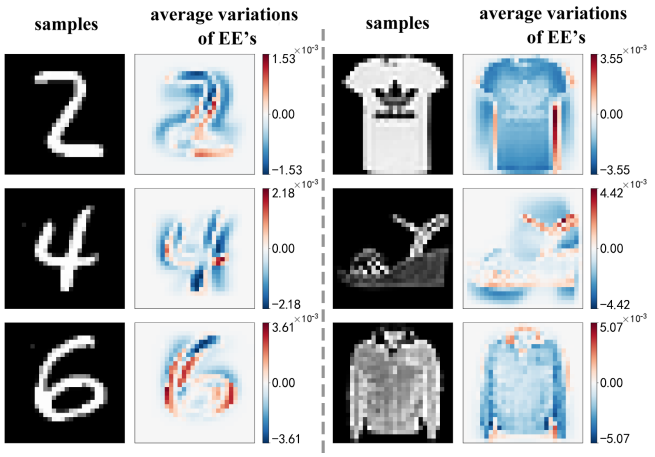

FIG. 2. (Color online) Six samples from the testing set of MNIST and fashion-MNIST datasets and the average variations of EE.

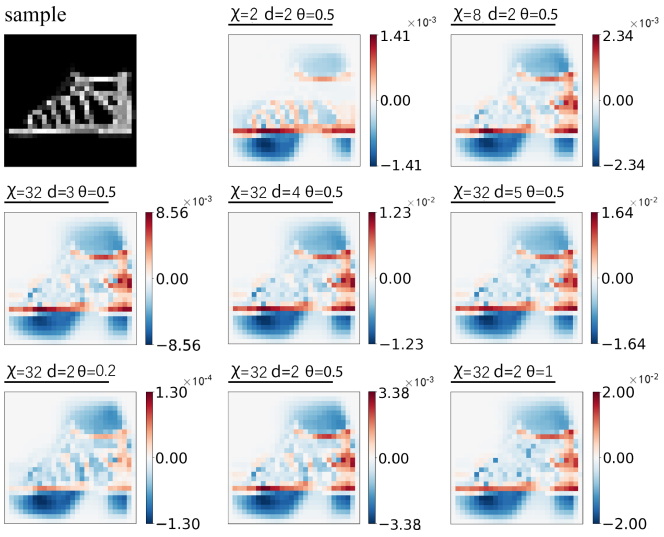

FIG. 3. (Color online) We take one specific image of shoe as an example (top-left) and show the variations of EE with different values of χ, d, and θ.

with $\binom{d-1}{s-1}$ the combination number and $\dim(\boldsymbol{v}) = d$ that is also the dimension of the physical indexes of the MPS. By taking $d = 2$ and $\theta = 0.5$, Eq. (1) is reduced to the feature map in Eq. (A1). We choose $\theta = 0.5$ in the main text since the classifier formed by the generative MPS's gives the highest classification accuracy [1].

In Fig. 3, we show the average variations of EE ($\langle \delta S \rangle_{m'}$) from a same image of shoe by taking different values of χ (the dimension of the virtual indexes), d and θ. In all cases, the distinct shapes at different parts of the shoe is well presented by $\langle \delta S \rangle_{m'}$ with slight differences. For instance , the strips of the shoe can be seen clearly with $\chi = 2$, $d = 2$, and $\theta = 0.5$ (top-middle of Fig. 3), and the back counter is clearly captured with $\chi = 32$, $d = 2$, and $\theta = 1$. In all cases, the sole and the topline are well presented. Ba aware that such distinct shapes of this specific image cannot be seen by the EE of the MPS, which is sample independent.

IV. ENTANGLEMENT ENTROPY FOR FEATURE SELECTION

The EE of the MPS can be utilized for feature selection. Following the idea of the generative MPS classification scheme proposed in Ref. [1], we take the MNIST as example and train a generative MPS for each of the classes. The classification is implemented by comparing the fidelity (a measure of similarity between two quantum states) between the product state [Eq. (A1)] from a target sample and the MPS's. For each MPS, we retain $M_f < M$ (note M is the total number of features in one sample) features whose qubits possess the largest EE. Fig. 4 shows the testing accuracy versus M_f with $\chi = 4$, 8, 16, and 32. By repeating ten times of simulations with a randomly initialized MPS, the thickness of the lines shows the variances that are insignificant.

Our accuracy surpasses that of the feature selection method proposed in Ref. [2] that is based on the EE of the discriminative MPS [3]. In the discriminative MPS approach, one trains an MPS that contains M physical indexes and an additional

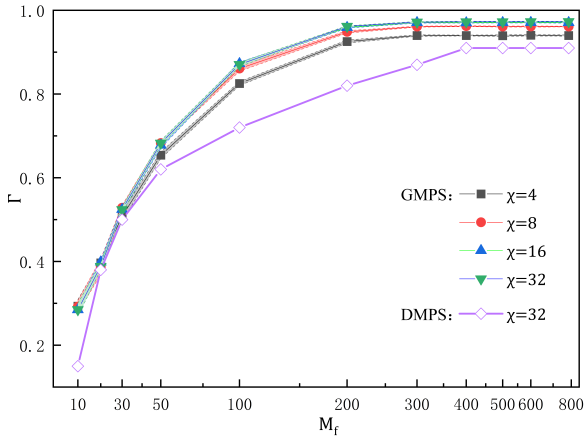

FIG. 4. (Color online) Testing accuracy Γ versus the number of features M_f selected according to the EE. The solid symbols with lines show the accuracy of the generative MPS's with $\chi = 4$, 8, 16, and 32. The thickness of the lines illustrates the variance evaluated by repeating the simulations for ten times. The purple hollow diamonds show the accuracy of the feature selection approach based on the discriminative MPS approach [2].

D-dimensional label index (with D the number of classes). The M_f features with the largest EE evaluated from the discriminative MPS are retained. Our method with the generative MPS classification scheme demonstrated obvious advantage for about $100 < M_f < 300$ and for the few-shot cases around $M_f \simeq 10$.

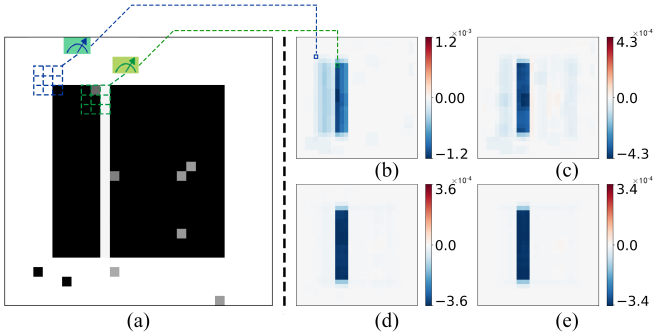

FIG. 5. (Color online) (a) An example in the noisy strip dataset. (b)-(e) The average variation of EE $\langle \delta S \rangle_{m'}$ with different numbers (6, 18, 180, and 540, respectively) of noisy training samples. The noises can be better excluded from the critical minority with more training samples.

V. GENERALIZING TO MULTI-QUBIT MEASUREMENTS TO DEAL WITH SINGLE-PIXEL NOISES

From the definition of the EE variations, the EE of each qubit is solely determined by the value of the corresponding pixel and the MPS, thus mathematically cannot distinguish the noises. In Fig. 5, we show that by applying the multi-qubit measurements, the single-pixel noises can be excluded from the critical minority. Fig. 5 (a) shows an example of the noisy strip dataset to test to the recognition of noises. By choosing $K > 1$ qubits numbered as $\{m_k\}$ $(k = 1, \ldots, K)$, the K-qubit measurement on these qubits are defined as

$$\Phi_{s_{1\ldots}} = \frac{1}{Z} \sum_{s_{m_1} \ldots s_{m_K}} \Psi_{s_{1\ldots}} \prod_{m' \in \{m_k\}} v^{[m']}_{s_{m'}}, \qquad (2)$$

where the state after the measurement $\mathbf{\Phi}$ contains all the indexes of $\mathbf{\Psi}$ but the summed ones, and $Z = |\mathbf{\Phi}|$ the normalization factor. One can see that the K-qubit measurement is a natural generalization of the single-qubit measurement given in Eq. (2). As an example, we for each time measure(3×3) neighboring qubits demonstrated by the blue dash squares. The average variations of EE $\langle \delta S \rangle_{m'}$ after the measurement are shown in Fig. 5 (b)-(e), where m' denotes the numbering of the (3×3) square for the measurement. The MPS is trained by 6, 18, 180, and 540 noisy samples, respectively. When the number of training samples is small, the noisy pixels in the background can be well excluded. But the $\langle \delta S \rangle_{m'}$ in the informative area show certain "fluctuations". When the number of training samples increases, our method can better select the informative features in all areas.

VI. RANDOMLY INITIALIZED MPS CANNOT SELECT IMPORT FEATURES

We calculate the variations of EE obtained from a randomly initialized MPS, where each tensor element is generated by the Gaussian distribution $\mathcal{N}(0, 1)$. Fig. 6 shows the variations of EE obtained from this MPS. It is obvious that the important features cannot be successfully selected by the variations of EE.

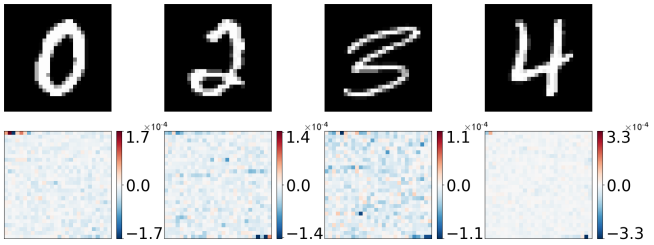

FIG. 6. (Color online) The first row presents four images from the MNIST dataset. The second row shows the variations of EE, which fail to select the important features.

VII. FEATURE SELECTION BY THE MPS TRAINED BY MULTI-CLASS DATA

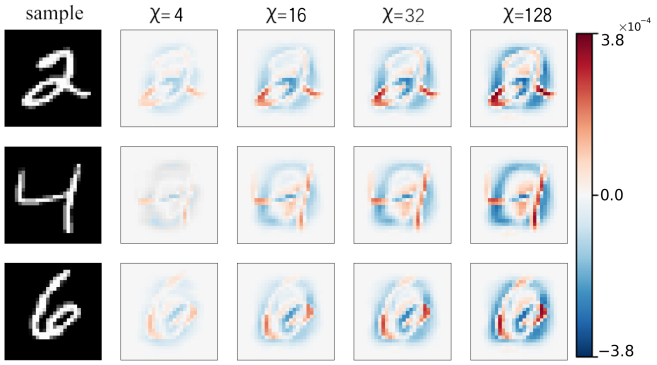

FIG. 7. (Color online) We take three images as examples (the left-most column) and show the variations of EE $\langle \delta S \rangle_{m'}$ for different bond dimensions $\chi = 4, 16, 32, 128$ by fixing $d = 2$ and $\theta = 0.5$.

In the main text, each MPS is trained by the samples from a same class. As a complement, we apply our method to the multi-class case, where we train an MPS by the samples from multiple classes. Specifically, we select 100 images from each class. The trained MPS is used for the feature selection by following the standard process proposed in the main text. The results of the EE variations with different virtual bond dimensions χ are shown in Fig. 7.

It can be observed by the naked eyes that one MPS manages to select the important features for the samples from all the classes. As χ increases, the importance of features is demonstrated more clearly. This relates to how well the probability distribution of the samples is captured by the MPS, which can be characterized by the negative logarithmic likelihood [4].

Appendix A: Formulas in the main text used in the supplemental material

$$\boldsymbol{v}^{[n]} = \prod_{\otimes m=1}^{M} \left[\cos\left(\frac{x_m^{[n]}\pi}{4}\right), \sin\left(\frac{x_m^{[n]}\pi}{4}\right) \right]^T, \quad \text{(A1)}$$

$$\langle \delta S \rangle_{m'} = \frac{\sum_{m \neq m'} (S'_m - S_m)}{M-1}, \quad \text{(A2)}$$

[1] Z.-Z. Sun, C. Peng, D. Liu, S.-J. Ran, and G. Su, Generative tensor network classification model for supervised machine learning, Phys. Rev. B **101**, 075135 (2020).

[2] Y. Liu, W.-J. Li, X. Zhang, M. Lewenstein, G. Su, and S.-J. Ran, Entanglement-based feature extraction by tensor network machine learning, Frontiers in Applied Mathematics and Statistics **7**, 10.3389/fams.2021.716044 (2021).

[3] E. Stoudenmire and D. J. Schwab, Supervised learning with tensor networks, in *Advances in Neural Information Processing Systems 29*, edited by D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett (Curran Associates, Inc., 2016) pp. 4799–4807.

[4] Z.-Y. Han, J. Wang, H. Fan, L. Wang, and P. Zhang, Unsupervised generative modeling using matrix product states, Phys. Rev. X **8**, 031012 (2018).