



Heuristic machinery for thermodynamic studies of $SU(N)$ fermions with neural networks

Entong Zhao¹, Jeongwon Lee², Chengdong He¹, Zejian Ren¹, Elnur Hajiyev¹, Junwei Liu¹  & Gyu-Boong Jo¹  [✉]

The power of machine learning (ML) provides the possibility of analyzing experimental measurements with a high sensitivity. However, it still remains challenging to probe the subtle effects directly related to physical observables and to understand physics behind from ordinary experimental data using ML. Here, we introduce a heuristic machinery by using machine learning analysis. We use our machinery to guide the thermodynamic studies in the density profile of ultracold fermions interacting within $SU(N)$ spin symmetry prepared in a quantum simulator. Although such spin symmetry should manifest itself in a many-body wavefunction, it is elusive how the momentum distribution of fermions, the most ordinary measurement, reveals the effect of spin symmetry. Using a fully trained convolutional neural network (NN) with a remarkably high accuracy of ~94% for detection of the spin multiplicity, we investigate how the accuracy depends on various less-pronounced effects with filtered experimental images. Guided by our machinery, we directly measure a thermodynamic compressibility from density fluctuations within the single image. Our machine learning framework shows a potential to validate theoretical descriptions of $SU(N)$ Fermi liquids, and to identify less-pronounced effects even for highly complex quantum matter with minimal prior understanding.

¹Department of Physics, The Hong Kong University of Science and Technology, Kowloon, Hong Kong, China. ²HKUST Jockey Club Institute of Advanced Study, The Hong Kong University of Science and Technology, Kowloon, Hong Kong, China. ✉email: gbo@ust.hk

Multi-component fermions with $SU(N)$ -symmetric interactions hold a singular position as a prototype system for understanding quantum many-body phenomena in condensed matter physics, high-energy physics, and atomic physics¹. In condensed matter, e.g., interacting electrons usually possess $SU(2)$ symmetry, whereas there are emergent higher spin symmetries for the low-energy property of systems as the $SU(4)$ symmetry in graphene due to the combination of spin and valley degrees of freedom². In quantum chromodynamics, nuclear interactions are represented by $SU(3)$ symmetry^{3,4}. In the past decades, developments in cooling and trapping of alkaline-earth-like fermions⁵ have opened possibilities to achieve even higher spin symmetries, owing to their distinctive inter-particle interactions, and thus provided ideal platforms to study various $SU(N)$ fermionic systems^{1,6,7}. Although the role of $SU(N)$ symmetry has been probed in optical lattices^{8–15}, the comprehensive characterization of interacting $SU(N)$ fermions in bulk, wherein the $SU(N)$ Fermi liquid description is valid, has still remained challenging^{16–19}. One of the bottlenecks is that the interaction-induced effect enhanced by enlarged $SU(N)$ symmetry is sufficiently pronounced by the tight confinement only in one-dimensional (1D)¹⁶ or two-dimensional (2D) cases¹⁸. It is only recently that thermodynamics and contact interactions are investigated in three-dimensional 3D $SU(N)$ fermions^{17,19}, but a comprehensive experimental study of $SU(N)$ fermions still remains to be done. Developing experimental techniques or designing approaches to uncover the subtle connection of various spin multiplicity-dependent properties with the available

experimental measurements in $SU(N)$ interacting fermions is crucial to advance our understanding of $SU(N)$ symmetry and the corresponding many-body phenomena.

Here we propose a framework to use machine learning (ML) as a guidance for the image analysis in quantum gas experiments and demonstrate the thermodynamic study of $SU(N)$ fermions. The main idea of this heuristic approach can be summarized into a three-step process as follows: (1) manually control the amount of information within each of the images we feed to the neural networks (NNs) during the training or testing processes; (2) determine the relative importance of the given (or removed) information based on the changes in the accuracy of the training or testing processes; and (3) identify the connection between the information and specific physical observables, which we can further focus our analytical efforts on.

To demonstrate the proposed machinery concretely, we take a density profile of $SU(N)$ Fermi gases as an example and show how it can guide the analytical studies. Besides the pronounced effects such as atom number and fugacity, we explore the connection between the spin multiplicity and less-pronounced features, such as compressibility and Tan's contact in the density profile. Based on this machinery, we demonstrate that one can extract less-pronounced effects even in the most ordinary density profiles and we successfully reveal thermodynamic features, which depend on the spin multiplicity, from density fluctuations and high-momentum distributions. This allows one to detect the spin multiplicity with a high accuracy $\sim 94\%$ in a single snapshot classification of $SU(N)$ density profiles. To further verify the

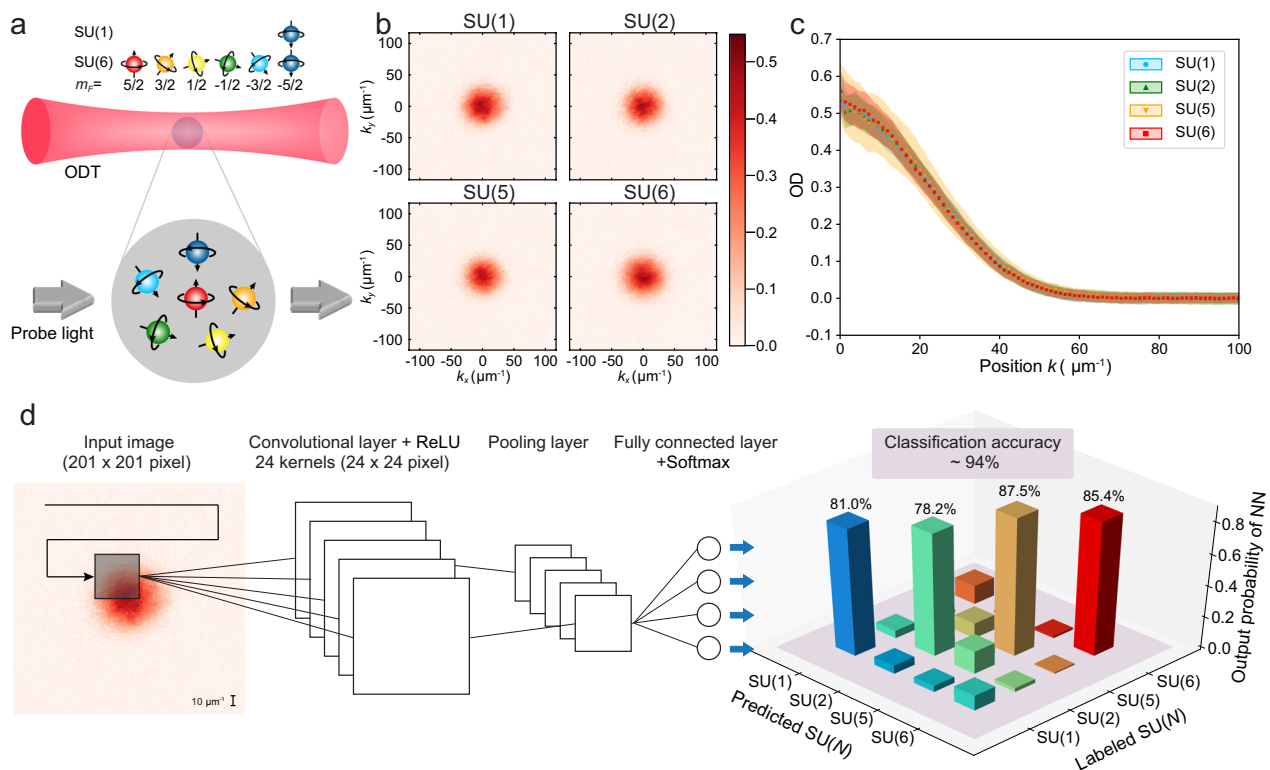


Fig. 1 Distinguishing $SU(N)$ fermions based on spin multiplicity by machine learning. **a** Schematic of preparing $SU(N)$ gases in the optical dipole trap (ODT). The momentum distribution of the $SU(N)$ Fermi liquid of ^{173}Yb atoms with tunable spin configuration is recorded. The collected datasets are then fed into the NN as the input images for classification. **b** Examples of single experimental images of $SU(N)$ gases. **c** Radially averaged optical density (OD) profiles in different $SU(N)$ gases. The shaded region represents the fluctuation of the density profiles. **d** Experimental images of $SU(N)$ gases are loaded into the neural network with one single convolutional layer. The black line and window represent how the kernel slides across the image. The output layer classifies the image into one of the class (i.e., $SU(1)$, $SU(2)$, $SU(5)$, $SU(6)$) resulting in a classification accuracy around 94%. For each input image, NN outputs probabilities of different $SU(N)$ with the highest value of the correct class. The output probabilities of NN are averaged over the test dataset.

validity of the connection between the less-pronounced effects and physical observable, we further measure the thermodynamic compressibility from density fluctuations within a single image benchmarking ML processes, which turns out to be in very good agreement with $SU(N)$ Fermi liquid descriptions. Besides providing general-purpose methods to extract various less-pronounced effects and consolidate our understanding of $SU(N)$ fermions, our approaches also complement recent ML studies of quantum many-body physics to explore the underlying physics^{20–26}.

Results

Train NNs to classify $SU(N)$ fermions. We begin by preparing the experimental measurements with appropriate labels. Here we choose one of the most ordinary experimental measurements for studying $SU(N)$ Fermi gases, the density profile, and the spin multiplicity as the labels. In our experiment, a degenerate $SU(N)$ Fermi gas with $N = 1, 2, 5, 6$ is prepared in an optical trap and the density profile is recorded by taking spin-insensitive absorption images after time-of-flight expansion, yielding the momentum distribution. The spin multiplicity is confirmed by optical Stern Gerlach measurements (see “Methods”). In principle, the density profile contains the momentum-space information of $SU(N)$ -interacting fermions, which reflects various thermodynamic observables, such as Tan’s contact or the compressibility, which is the underlying reason for the success of using ML techniques to detect the spin multiplicity. However, the effect of spin multiplicity on the momentum distribution is small compared to other features such as the fugacity and atom number because of small interaction strength. Therefore, the dataset should be prepared in such a way that images are indistinguishable based on the pronounced features (i.e., atom number or temperature), which forces the NN to seek for less-pronounced features. We post-select datasets and minimize possible correlations between spin multiplicity and atom number or temperature.

In detail, we focus on the density profiles with the interaction parameters $k_F a_s \simeq 0.3$ where k_F is the Fermi wave vector and a_s the scattering wavelength, and we only select the profiles based on similarities in widths of Gaussian fitting of the density profiles to result in indistinguishable momentum profiles as shown in Fig. 1b, c (see “Methods”). We collect 200 density profiles for each class of $SU(1)$, $SU(2)$, $SU(5)$, and $SU(6)$ (Fig. 1b). We randomly feed 150 of them to train the NNs by implementing the supervised ML techniques with spin multiplicity as labels and use the remaining 50 profiles to evaluate the classification accuracy that is defined as the ratio of number of samples with predictions matching true labels to the total sample number. To maximize the accuracy of NNs, we choose the architecture of convolutional NNs (CNNs) that is suited to explore the less-pronounced effects in an image (more details in “Methods”), as shown in Fig. 1d. By choosing the suitable structures and parameters in the CNNs, we can realize a very high accuracy ~94%, which is much better than the random guess (25%). We also test various unsupervised learning techniques such as the typical principal component analysis^{27,28} and only get a low classification accuracy of only ~43%. Moreover, it is worth to emphasize that the remarkably high accuracy ~94% of NNs is achieved by using only a single snapshot of the density profiles. All these results indicate that there are detectable, less-pronounced features in a single snapshot of density profile.

Extract less-pronounced effects in low- and high-momentum parts. We now analyze the attributes processed by the well-trained NNs and extract less-pronounced effects determined by

the spin multiplicity step by step. Due to the limited interpretability of NNs, it is usually difficult to identify what kinds of features the NNs use for classification. In our proposed machinery, we examine which parts of the density profile are related to the spin multiplicity as described in Fig. 2a. Usually, it is more efficient to use some prior knowledge, which can be obtained in our limited understanding of the current system or the well-established understanding of the similar system. In our example of studying the interacting $SU(N)$ fermions, we use the prior knowledge of non-interacting fermions and the associated energy (length) scale in choosing various filters in the momentum space. It is conceivable that our heuristic machinery can be applied to other systems.

To do this, we manually manipulate the experimental images and subsequently check the classification accuracy of the manipulated images. As different information is removed in different types of manipulated images, the classification accuracy will decrease with different amount, which will unfold what kind of information is more important for classification. As shown in Fig. 2b, we first replace the whole image with the Gaussian and Fermi–Dirac fitting profile to do the test based on the prior knowledge of non-interacting fermions. It turns out that the classification accuracy significantly decreases for both cases and the accuracy drop of the Gaussian fitting profile is even more, which implies the Fermi–Dirac type preserves the characteristics of the original profile better than the Gaussian fit. We further test the variations in accuracies by replacing profiles with radially averaged profiles, which results in test accuracies higher than the Fermi–Dirac fitting cases. However, the differences in accuracies between the radially averaged and Fermi–Dirac are much smaller compared to the differences between Fermi–Dirac and Gaussian, suggesting the $SU(N)$ -dependent modifications of the Fermi–Dirac distribution to be small.

Now we examine the contributions of low- and high-momentum parts by classifying the masked images with the well-trained NN as shown in Fig. 3a, b. This is motivated by the observation that the classification accuracy significantly decreases with filters, with various fitting functions that remove the $SU(N)$ -dependent effect in the high-momentum tail. We choose two different types of masks for the region of the replacements, which will be referred as background and central masks, respectively. Background mask covers from the edge of the image to some atomic momentum k_c , whereas the central mask covers from the center to k_c . Then, we replace the masked region with a fake image generated by averaging the corresponding region of all the images in the dataset and re-evaluate the test accuracies of the pre-trained NN.

First, we set the cutoff momentum of $k_c = 70 \mu\text{m}^{-1}$ such that >99% of atoms are contained within the low-momentum region (Fig. 3a), which still allows us to classify the spin multiplicity with the accuracy of 88%. Although this confirms that a NN perceives spin multiplicity-dependent information in low-momentum parts, questions remain on why the accuracies are not fully recovered beyond 94%. Such observation confirms the importance of the high-momentum information. In Fig. 3b, we prepare a dataset with a high-momentum part only ($k > k_c = 70 \mu\text{m}^{-1}$), in which a low-momentum region is deliberately replaced by the same fake image. Surprisingly, the test classification accuracy is still >50% and the overall classification accuracy increases to 65% when a NN is re-trained. Such a high accuracy based on the few information of only <1% of atoms strongly implies that the high-momentum tail is crucial for determining the spin multiplicity. This $SU(N)$ -dependent feature is not due to the finite resolution of the imaging system²⁹, as the NNs can classify the binned image with high classification accuracy (see “Methods” for more information).

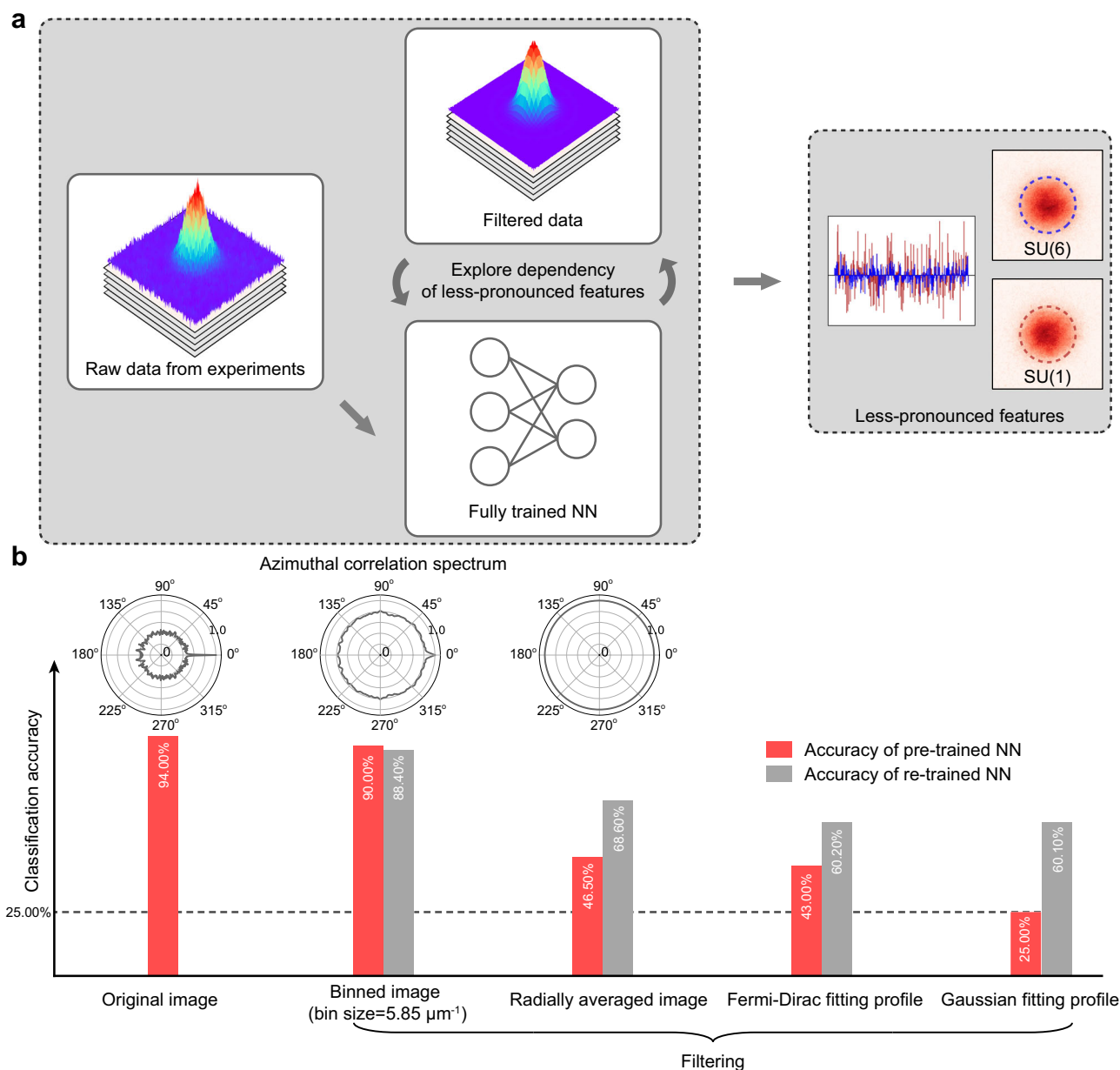


Fig. 2 Exploring less-pronounced effects with NNs using deliberately filtered experimental images. **a** Our heuristic machinery to uncover less-pronounced effects. **b** Effect of image filtering on the classification accuracy by neural networks trained with original (red) and filtered (gray) experimental images. The retraining process follows the same procedure of original training with fixed neural network parameters and hyper-parameters (e.g., the number of epochs and learning rate). In the inset, the azimuthal correlation spectrums calculated at $k = 58.5 \mu\text{m}^{-1}$ are shown.

High-momentum tails. In light of these results, we speculate that NNs utilize less-pronounced effects in the high-momentum part. To confirm this, we check the dependence of classification accuracy with the fully trained NNs on each $SU(N = 1, 2, 5, 6)$ class and we find the classification accuracy increases with N in Fig. 3b. In addition, the output probability of the correct spin multiplicity increases with N (see Fig. 3b). These results indicate that the less-pronounced feature being used in NNs becomes more prominent with increasing spin multiplicity N , which is consistent with the fact that atom–atom interactions are absent in the case of $SU(1)$ due to the Pauli principle in the ultracold regime, while they are significantly enhanced in $SU(6)$ fermions. Indeed, the amount of short-range interactions should be revealed in the high-momentum distribution in which the weight of such high-momentum tail is determined by the central quantity, so-called the contact^{30–32}, in a dilute quantum gas. The contact governs

many other physical observables^{33,34} and has been probed in strongly interacting gases^{33–39}, and even in a weakly interacting gas with $SU(N)$ symmetric interactions¹⁷. It is conceivable that the NNs detect the high-momentum distribution within a single image in contrast to the previous work where hundreds of momentum-space images are statistically averaged in a ^{173}Yb Fermi gas¹⁷. To be noted, our observation is consistent with the direct measurement of the high-momentum tail in the region of $k/k_F > 3$ ¹⁷, which is corresponding to $k > 100 \mu\text{m}^{-1}$ for an $SU(1)$ gas in this work.

Evaluating detection accuracy with tunable masks. To examine the less-pronounced effects in both the low-momentum and high-momentum regions, and build up the concrete connections between these less-pronounced effects and the spin multiplicity, we now quantitatively analyze the changes inflicted on the test

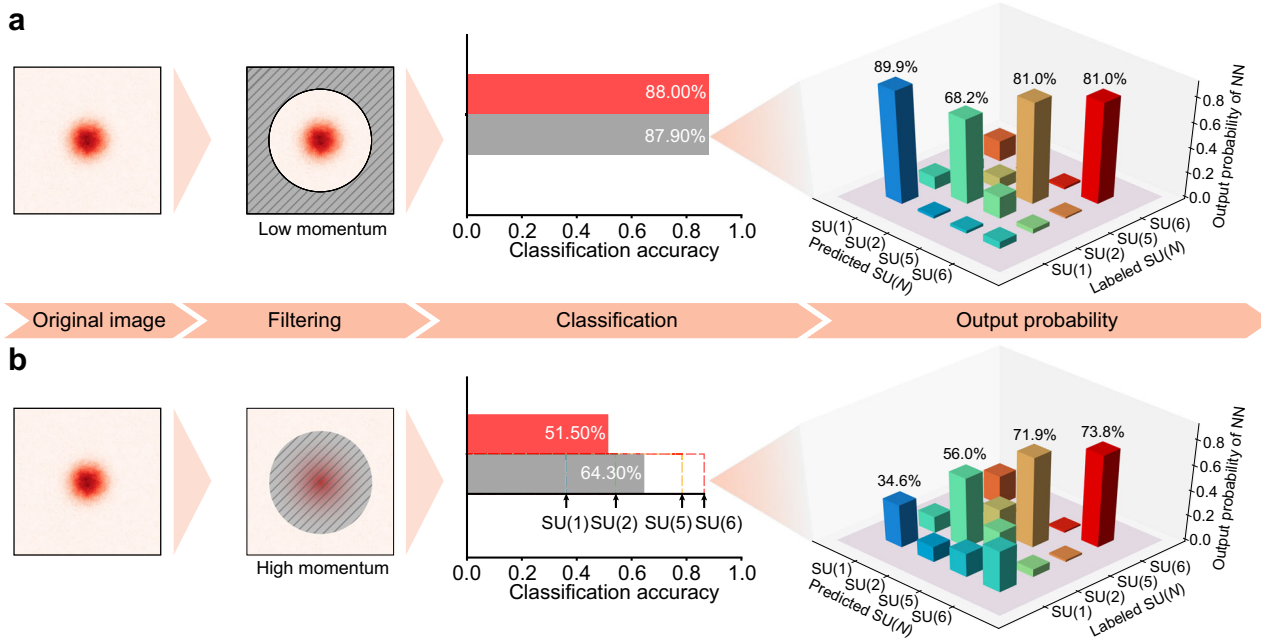


Fig. 3 Influence of low- and high-momentum parts on classification. Classification accuracy with neural networks trained only with low **a** and high **b** momentum parts. Red and gray bars represent the accuracy of pre-trained and re-trained NNs, respectively. Retraining of the neural networks with only high- or low-momentum parts enhances the accuracy as shown in gray bars. For high-momentum parts, classification accuracy increases with N as shown in black ticks in **b**, where the average of the four investigated $SU(N)$ cases is 64.30%. For the re-trained NN, the output probabilities of the NN are shown for input images.

accuracies when the cutoff momentum k_c is tuned over. It is clearly shown in Fig. 4a that the test accuracy decreases to $\sim 25\%$ by complete replacements of the images by the same average image. This accuracy is gradually recovered up to almost 90% when k_c is increased to $k_c \simeq 70 \mu\text{m}^{-1}$ being consistent with the result in Fig. 2, as less information from the low-momentum regime is removed. Based on replacement analysis, it is conceivable that $SU(N)$ -dependent features must exist in the low-momentum part, as we will discuss the details below. In comparison, the classification accuracy of the binned images (bin size = $5.85 \mu\text{m}^{-1}$) does not decrease too much, which is only a partial removal of fluctuations. As a complementary study, we utilize the central mask and replace the low-momentum information up to a variable momenta of k_c in Fig. 4b. The classification accuracy gradually decreases with increasing k_c from 0 to $\sim 50 \mu\text{m}^{-1}$, as the information within the density profile is increasingly removed. However, the classification accuracy stays over 50% around $k_c = 50 \sim 70 \mu\text{m}^{-1}$, which strongly suggests that the high-momentum tails of the density distribution still contribute towards the classifications based on $SU(N)$. Beyond $k_c = 80 \mu\text{m}^{-1}$, where the atomic shot-noise becomes comparable to the background shot-noise level, the test accuracy rapidly drops for the images replaced by averaged images.

Density fluctuations and compressibility. The question still remains as to what dominant feature classifies spin multiplicity in the low-momentum regime. Based on the significant decrease of the accuracy with profiles being radially averaged in Fig. 2b, we hypothesize that the NNs utilize the density fluctuation along the azimuthal direction for classification. The amount of azimuthal density fluctuations can be revealed in the correlation spectrum (Fig. 2b) showing a strong signal in the original and binned images, while flattened at all angles for the radially averaged images.

To understand how the density fluctuations reveal spin multiplicity, we consider the fluctuation-dissipation theorem by

which the thermodynamic compressibility $\kappa = \frac{1}{n^2} \frac{\partial n}{\partial \mu}$ can be measured through density fluctuations (i.e., atom number fluctuations)^{19,29,40,41} where n is the local density and μ the local chemical potential. For repulsively interacting $SU(N)$ fermions, it is known that the compressibility κ decreases with increasing spin multiplicity N as $(\kappa/\kappa_0)^{-1} = 1 + \frac{2}{\pi}(k_F a_s)(N-1)(1 + \epsilon k_F a_s)$, where κ_0 is the compressibility of an ideal Fermi gas and $\epsilon = \frac{2}{15\pi}(22 - 4\ln 2)$ ⁴². Here, the atom number fluctuations are further suppressed by the Pauli blocking in the degenerate regime showing sub-Poissonian fluctuations as $\sigma_{N_a}^2/\overline{N}_a \propto k_B T$ where N_a indicates the atom number measured in the small volume. Therefore, one finds the relative atom number fluctuation $\sigma_{N_a}^2/\overline{N}_a$ is given as

$$\begin{aligned} \sigma_{N_a}^2/\overline{N}_a &= nk_B T \kappa \\ &= \frac{3}{2} \frac{T/T_F}{1 + \frac{2}{\pi} k_F a_s (N-1)(1 + \epsilon k_F a_s)} \end{aligned} \quad (1)$$

In our experiment, an atomic sample ballistically expands from the harmonic trap preserving occupation statistics of the phase space during the expansion^{43,44}. Instead of repeatedly producing identical samples and monitoring a small region at the certain position²⁹, the relative atom number fluctuations can be extracted along the azimuthal bins containing the same number of atoms on average (therefore, resulting in equivalent optical density) within a single image, even though a grouping of ideally equivalent bins is challenging and the fluctuation measurement is susceptible to the systematic variations. The successful classification of the spin multiplicity with NNs now guide us to subsequently investigate the atom number fluctuations with conventional analysis.

To verify this less-pronounced feature, we choose a series of bins containing ~ 450 atoms on average in a line-of-sight integrated density profile along the azimuthal direction (Fig. 4c) and analyze the statistics. To have a sufficiently large number of bins for

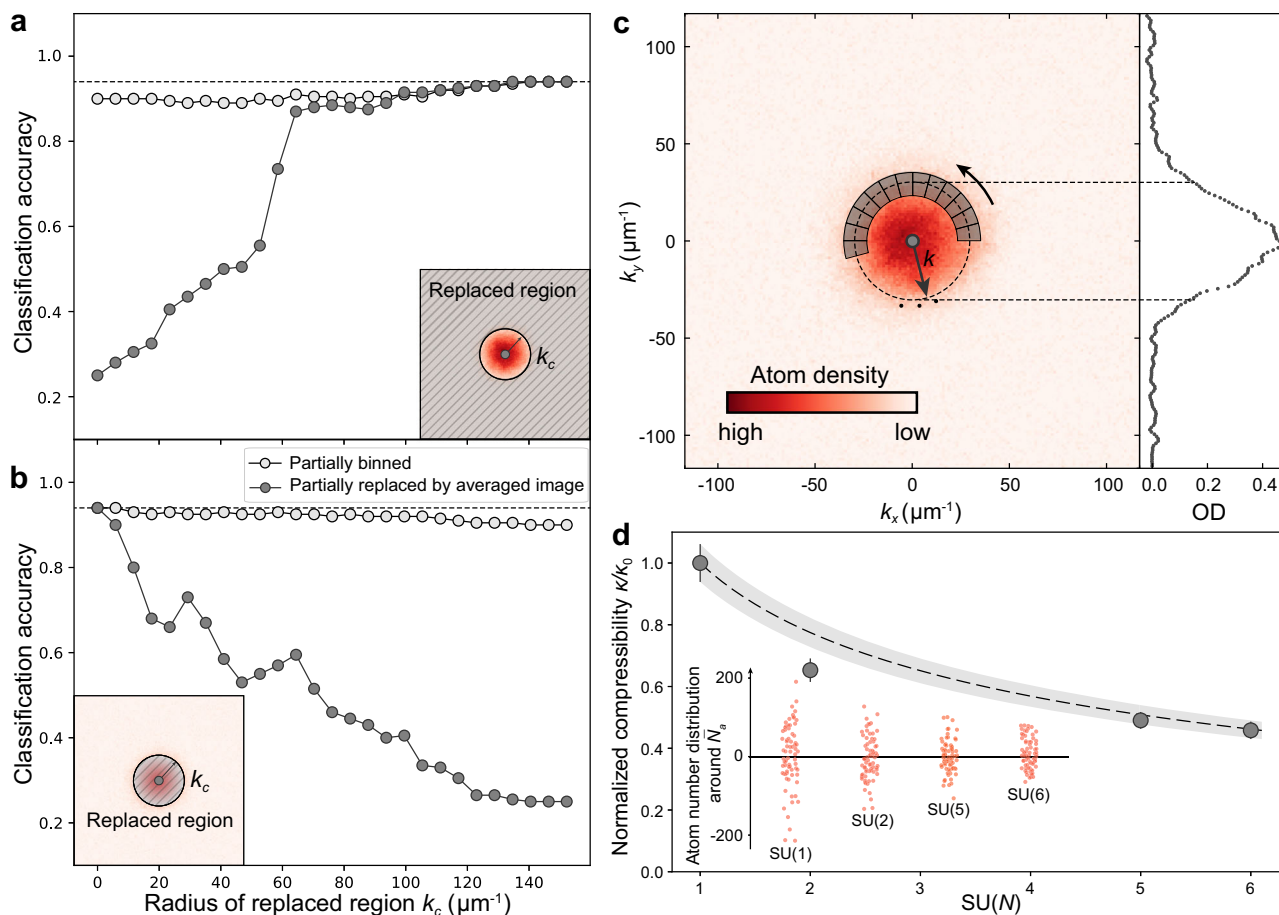


Fig. 4 Verification of ML-aided detection: density fluctuations and thermodynamic compressibility. **a, b** Classification accuracy of the correct class as a function of the cutoff momentum k_c of the mask. The dotted line indicates the accuracy of 94%. **c** Measurement of density fluctuations with a snapshot. In a line-of-sight integrated density profile, a series of bins containing on average \bar{N}_a atoms are chosen along the azimuthal direction. Each bin is about 10 (in azimuthal direction) by 17 (in radial direction) μm^{-1} , which is much larger than the optical resolution of the imaging system. The density profile at $k_x = 0$ is shown. **d** The normalized compressibility of $SU(N)$ fermions κ/κ_0 is measured by relative density fluctuations as $\kappa/\kappa_0 = \zeta_{SU(N)}/\zeta_{SU(1)}$. The error bar shows the SE. The dashed line indicates the theory curve $\kappa/\kappa_0 = [1 + \frac{2}{\pi} k_F a_s (N - 1)(1 + \epsilon k_F a_s)]^{-1}$ with the uncertainty represented by the shaded region considering the SE of $\zeta_{SU(1)}$. The inset shows the distribution of the atom number per bin from three images for each spin multiplicity. The distribution is plotted around the average normalized by the degenerate temperature, $(N - \bar{N}_a)/(T/T_F)$, where \bar{N}_a is the average atom number.

statistical analysis, we perform the measurements at varying momenta on the ring with $(k_x^2 + k_y^2)^{1/2} \simeq 28 \mu\text{m}^{-1}$ around the center of the distribution. Both the temperature and spin multiplicity are known to affect the atomic-density fluctuations through the change in compressibility. Therefore, we normalize relative atom fluctuations by the temperature as $\zeta_{SU(N)} = \frac{\sigma_{N_a}^2}{\bar{N}_a} / \frac{T}{T_F}$ to reveal $SU(N)$ interaction effect from a single snapshot. In Fig. 4d, we then plot the statistical value of $\zeta_{SU(N)}/\zeta_{SU(1)}$. This measurement indeed reveals the normalized thermodynamic compressibility $\kappa/\kappa_0 = \zeta_{SU(N)}/\zeta_{SU(1)}$ showing the enhanced interaction $k_F a_s (N - 1)$. The error bar indicates the SE from 150 different density profiles. Whereas the scaling of the measured density fluctuation with N is in good agreement with theoretical prediction, experimental results for $SU(N > 1)$ lie systematically below theoretical ones. The discrepancy may be due to interactions that remain finite during the expansion, which could slightly perturb the occupation statistics of the phase space. Considering the fact that the change of the compressibility is not significant for $N = 5$ and 6 in Fig. 4d, the high classification accuracy of NNs using the low-momentum part highlights the superior capabilities of a single snapshot approach using ML. In contrast to the conventional analysis that focuses on a single observable, NNs take

a holistic approach in utilizing multiple features simultaneously. Our measurement is consistent with recent experiments in which thermodynamics is studied by monitoring the density fluctuations and expansion dynamics in degenerate ^{87}Sr atoms¹⁹.

Discussion

To scrutinize the effects of $SU(N)$ symmetric interactions, we have provided the NN with altered images and probed specific attributes of the profiles independently. We found that the high-momentum tail and density fluctuation information significantly contribute to the $SU(N)$ classification process. First of all, the high-momentum tails of atomic-density distributions are expected to exhibit Tan’s contact, which encapsulates the many-body interactions through the set of universal relations. Although the previous work required averaging of hundreds of images for the detection of the $SU(N)$ -dependent contact¹⁷, the NN’s ability makes it possible to obtain the single-image distinguishability of the $SU(N)$ class after training. However, the exact mechanism behind how the trained network collects the required information for extraction of the contact, whether it is through superior noise suppression or signal enhancement, is not known and is left for future work. Furthermore, it is conceivable that the regression

algorithms can be used to extract the change of contact for different spin multiplicity in future works.

The second dominant feature for the $SU(N)$ classification is the density fluctuation within the profile. Both the temperature and spin multiplicity are known to affect the atomic-density fluctuations through the change in compressibility. Sub-Poissonian density distributions have been observed in degenerate Fermi gases of atoms^{29,40} and molecules⁴¹, where multiple images were used to obtain the statistics. The suppression of the density fluctuation was also observed in $SU(N)$ fermions allowing for the thermodynamic study¹⁹. For a single image, there exists multiple sets of density fluctuation measurements at varying momentum, where each measurements form a ring around the center of the distribution. Considering the decreased $SU(N)$ classification accuracy from the radially averaged datasets, the fluctuation information might have been utilized in addition to the contact, to reflect the effects of compressibility. Lastly, we found that the low-energy part of the density profile does not exhibit a signature as strong as the previous two features. Although there has been a report of $SU(N)$ -dependent modifications to the density distribution limited to the 1D case¹⁶, the corresponding beyond mean-field effects in 3D remains challenging to be measured experimentally.

In conclusion, we have demonstrated the capabilities of the proposed machinery by classifying $SU(N)$ Fermi gases with their time-of-flight density distributions. The NN provides classifications with an accuracy well beyond the conventional methods such as principal components analysis. By applying different types of manipulations, we also find that the NNs combine the features from a high-momentum signal and density fluctuations together, to distinguish $SU(N)$. Future directions include predictions of T/T_F of $SU(N)$ Fermi gases based on regression algorithms and explorations of human feedbacks to the ML process for feature extractions. Feature extraction through ML may guide us to investigate the right information and facilitate research in many-body quantum systems.

Methods

Sample preparation. We prepare a balanced ultracold Fermi gas of ^{173}Yb atoms with $SU(N)$ symmetric interactions as large as $N = 6$. (shown in Fig. 1a). We begin by loading a laser-cooled, six-component Fermi gas, where the nuclear spin states are equally populated, into a three-dimensional optical dipole trap (ODT). The atoms are further evaporatively cooled in the ODT to a temperature range of $0.2\text{--}1.0 T/T_F$, where T_F is the Fermi temperature. During the evaporation, different spin configurations are prepared via an optical pumping process using a narrow line-width transition of $^1S_0(F = 5/2) \rightarrow ^3P_1(F' = 7/2)$ at a wavelength of $\lambda = 556$ nm. The σ^z -polarized pumping light removes unwanted m_F states of the ground manifold of 1S_0 ⁴⁵ and produces a Fermi gas with tunable $SU(N)$ interactions, as the nuclear spin relaxation rates are negligible in our experiment. After the evaporative cooling, the ODT is further ramped up in 60 ms, to obtain large-enough trap frequencies $(\omega_x, \omega_y, \omega_z) = 2\pi \times (1400, 750, 250)$ Hz before 4 ms of time-of-flight expansion. We measure the density distributions by taking absorption images using a spin-insensitive $^1S_0(F = 5/2) \rightarrow ^1P_1(F' = 7/2)$ transition at 399 nm. The images are taken in random order with respect to their spin configurations, to avoid the possibility of a classification based on fluctuations in the background. The spin configuration of the sample can be monitored by the optical

Stern Gerlach measurement. In general, the atom number of different spin states has a fluctuation of $\pm 2\%$ of the total atom number.

Data preparation. All snapshots are first preprocessed by the fringe removal algorithm reported in ref. 46. Then, cropped images are loaded into the NN for further classification. For $SU(N)$ data, it is natural to prepare the same number of atoms per spin at constant T/T_F , in which the normalized density profile is the same for different $SU(N)$ cases. In this case, however, we find that the diffraction of the imaging light induces fringe patterns that depend on the total atom number in the experiment. One can normalize the image by the total atom number, but we inevitably change the level of background noise. Therefore, we keep the total atom number unchanged, otherwise the NN uses the background fringe patterns or noises to classify the $SU(N)$ data. In our experiment, we post-select 200 images per each $SU(N)$ class by using a Gaussian fitting, which allows us to obtain samples with similar profiles but different T/T_F . If we have kept different $SU(N)$ gases at constant T/T_F , the profiles are identical in the unit of k_F , instead of pixel. Subsequently, 75% of the data is used for training NNs and the remaining is for test.

Machine learning. ML, a sub-field of artificial intelligence, allows us to understand the structure of data and deduce models that explain the data. Traditionally, ML can be classified into two main categories, supervised and unsupervised learning, based on whether there are labels or not for training. Supervised learning usually trains a model from a known dataset of input data $\{x_i\}$ and output labels $\{y_i\}$ so that the model can find a correspondence rule $x_i \mapsto y_i$, which allows us to predict the labels of data beyond the training dataset. In contrast, unsupervised learning is used to classify the data into several different clusters based on the potential patterns or intrinsic structures in the dataset without any prior knowledge of the system or data properties.

Convolutional neural network. ML techniques used in this work are based on CNNs, which takes a supervised learning approach for classification task. NNs, inspired by the biological NNs that constitute animal brains, are composed of a series of artificial neurons, among which the connection is a real-valued function $f: \mathbb{R}^k \rightarrow \mathbb{R}$, parameterized by a vector of weights $(w_1, w_2, \dots, w_i, \dots) = \mathbf{w} \in \mathbb{R}^k$ and the activation function $\phi: \mathbb{R} \rightarrow \mathbb{R}$, given by

$$f(x) = \phi(\mathbf{w} \cdot \mathbf{x}) \quad \text{with} \quad \mathbf{x} = (x_1, \dots, x_i, \dots) \in \mathbb{R}^k. \quad (2)$$

By combining the artificial neurons in a network or in a layer of network, we obtain NNs. In recent years, CNNs have shown stronger validity and better performance than regular NN in image recognition. Similar to the regular NNs, CNNs also consist of a sequence of layers and each layer receives some inputs, performs a dot product, and optionally follows it with a nonlinear activation function. However, unlike a regular NN, a CNN usually has several convolutional layers where neurons are arranged in two dimensions, providing an efficient way of detecting the spatial structure. The convolutional layer first accepts an input 2D image from the previous layer or the whole NN. Then, the kernel of the convolutional layer slides (i.e., convolve) across the width and height of the input image, with dot products between the kernel and the input being computed. Consequently, we obtain a 2D feature map in which each pixel is the response at the corresponding position. If the convolutional layer has N different kernels, the same procedure will be repeated for each kernel and finally N 2D feature maps will be produced. These 2D feature maps will then be loaded into the next layer as input.

Training and evaluating the NN. The CNNs used in this study are realized by using the Tensorflow in Python⁴⁷. We have attempted different architectures and found that the result is not sensitive to the choice of architecture such as the number of layers or the kernel size. Therefore, we remove superfluous layers to simplify our model. The concrete parameters taken in this work are listed in Table 1.

To train the network on the data with different spin configurations, the model is compiled with a cross-entropy loss function. During the training process, the weights of model are updated based on Adam algorithm⁴⁸, to minimize the loss function with a learning rate of 1×10^{-4} , which is a hyper-parameter that controls

Table 1 Network architecture and parameters used in this work.

Layer	Layer name	Function	Description
1	Input	Image input	201 × 201 Images
2	Conv. layer	Convolution	24 24 × 24 Convolutions with stride (1,1)
3	ReLU	Activation function	ReLU function
4	Pool layer	Average pooling	2 × 2 With stride (1,1)
5	Dropout	Dropout	50% Dropout
6	Fully conn. layer	Fully connected	Fully connected layer with 4 neurons
7	Softmax	Activation function	Softmax function
8	Output	Classification output	Probability with classes $N=1,2,5,6$

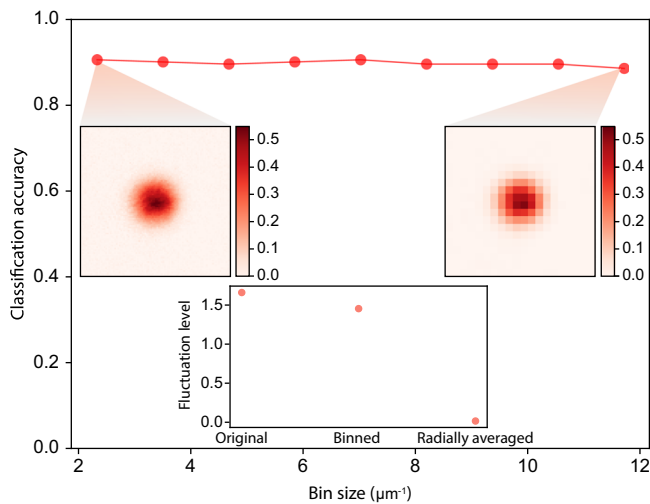


Fig. 5 Classification accuracy with different bin sizes. The classification accuracy is quite robust when bin size n is changed from 2.3 to $11.70 \mu\text{m}^{-1}$. The inset shows the comparison of fluctuation level among the original image, binned image, and radially averaged image.

how much the network changes the model each time. The maximum training epochs are limited to 1000, and the accuracy and loss are monitored during the training process for selecting the model with best performance. After full training, we evaluate the trained model on the test dataset.

We characterize the performance of trained NNs by obtaining the overall classification accuracy, which is defined as the ratio of number of samples with predictions matching true labels to the total sample number. For one single image loaded into the NNs, e.g., the softmax activation function normalizes the output values $\{\sigma_c\}$ by $P(c) = e^{\sigma_c} / \sum_{c=1}^N e^{\sigma_c}$, which allows a probabilistic interpretation for the different classes denoted by the subscript c . When calculating the classification accuracy, the class with highest probability $P(c)$ is selected as the prediction from NNs. As a complementary analysis of NNs, we also evaluate an output probability matrix that hints how the NNs perform among different classes, such as that shown in Figs. 1d and 3. In the probability matrix, every element A_{ij} represents the probability $P(c)$ averaged over the results for all images with the true label j and prediction i .

Manipulation of SU(N) data. In this work, we manipulate the experimental images to remove different types of information. In Fig. 2, we examine the binned image, radially averaged image, Fermi–Dirac fitting profile, and Gaussian fitting profile. The blurring of adjacent pixels effectively changes subtle features in SU(N) gases due to finite optical resolution, e.g., it will decrease the measured atom number variance²⁹. We minimize this effect by binning the data using a sufficiently large bin size²⁹. The whole image is partitioned into bins with the area of $n \mu\text{m}^{-1} \times n \mu\text{m}^{-1}$ without overlapping. In each bin, the averaged optical density with the bin is calculated and the value is subsequently used to fill all the pixels of the bin to maintain the original size of image. We attempted several different bin size n from 2.34 to $11.70 \mu\text{m}^{-1}$ and the result is robust against the bin size, as shown in Fig. 5.

For the radially averaged images, we first divided all the pixels into several bins based on the distance from the center of the atom cloud, then averaged the pixels in the same bin. The degenerate Fermi–Dirac and Gaussian profiles are fitted by the 2D Thomas–Fermi and Gaussian distribution, respectively. It is worth noting that both density fluctuations and high-momentum information are effectively removed from both fitting cases. Therefore, the comparison between Fermi–Dirac and Gaussian profiles may allow one to investigate possible next-order effects by which NNs detect the changes in T/T_F .

In Fig. 3, we first divide the whole image into two parts, low-momentum and high-momentum regions, based on whether the distance from the center of the atom cloud is larger than $70 \mu\text{m}^{-1}$ or not. Next, we replace one of the two parts with a fake image, which is generated by averaging the corresponding region of all the images ($N = 1, 2, 5, 6$) in the dataset. As all the test images are same in the replaced region, the information in that region can be considered as removed. In Fig. 4a, b, the procedure is the same as in Fig. 3 with variable cutoff momenta. For the partially binned images, the corresponding region is replaced by a binned image.

Data analysis. In Fig. 2a, we calculate the azimuthal correlation spectrum at $k = 58.5 \mu\text{m}^{-1}$, which is defined as $C_k(\theta_j) = \frac{\sum_i [P_k(\theta_i)P_k(\theta_{i+j})]}{\sum_i P_k^2(\theta_i)}$, where $P_k(\theta_i)$ represents the optical density for a specific pixel at $k \sim 58.5 \mu\text{m}^{-1}$ and angle $\sim \theta_i$. The formula can be further derived from the Fourier transform \mathcal{F} as $G = \frac{\mathcal{F}^{-1}(|\mathcal{F}(P_k)|^2)}{|P_k|^2}$. The azimuthal

correlation spectrum shows how the image looks like its own copy after rotating a specific angle. For a radially averaged image, the correlation becomes 1 at any angle, indicating no density fluctuations. When the image is binned, densities are only locally averaged resulting azimuthal correlation at nonzero angle < 1 .

Figure 4d shows density fluctuations measured along azimuthal bins containing the same number of atoms within a single image. Total 24 bins are chosen at the distance of $22 \sim 34 \mu\text{m}^{-1}$ from the center of the cloud. Redundant pixels are removed at the border to keep all bins having the same number of pixels. The size of each bin is much larger than the optical resolution of the imaging system.

Data availability

The data that support the finding of this work are available from the corresponding authors upon request.

Code availability

The code developed during the current study is available from the corresponding authors upon request.

Received: 23 June 2020; Accepted: 2 March 2021;

Published online: 31 March 2021

References

- Cazalilla, M. A. & A. M., Rey Ultracold fermi gases with emergent SU(N) symmetry. *Rep. Prog. Phys.* **77**, 124401 (2014).
- Castro Neto, A. H., Guinea, F., Peres, N. M. R., Novoselov, K. S. & Geim, A. K. The electronic properties of graphene. *Rev. Mod. Phys.* **81**, 109 (2009).
- Gell-Mann, M. Symmetries of baryons and mesons. *Phys. Rev.* **125**, 1067 (1962).
- Ne’eman, Y. Derivation of strong interactions from a gauge invariance. *Nuclear Phys.* **26**, 222–229 (1961).
- He, C., Hagiye, E., Ren, Z., Song, B. & Jo, G.-B. Recent progresses of ultracold two-electron atoms. *J. Phys. B Atom. Mol. Opt. Phys.* **52**, 102001 (2019).
- Cazalilla, M. A., Ho, A. F. & Ueda, M. Ultracold gases of ytterbium: ferromagnetism and Mott states in an SU(6) Fermi system. *New J. Phys.* **11**, 103033 (2009).
- Zhang, R., Cheng, Y., Zhang, P. & Zhai, H. Controlling the interaction of ultracold alkaline-earth atoms. *Nat. Rev. Phys.* **2**, 213–220 (2020).
- Gorshkov, A. V. et al. Two-orbital SU(N) magnetism with ultracold alkaline-earth atoms. *Nat. Phys.* **6**, 289–295 (2010).
- Taie, S., Yamazaki, R., Sugawa, S. & Takahashi, Y. An SU(6) Mott insulator of an atomic Fermi gas realized by large-spin Pomeranchuk cooling. *Nat. Phys.* **8**, 825–830 (2012).
- Zhang, X. et al. Spectroscopic observation of SU(N)-symmetric interactions in Sr orbital magnetism. *Science* **345**, 1467–1473 (2014).
- Cappellini, G. et al. Direct observation of coherent interorbital spin-exchange dynamics. *Phys. Rev. Lett.* **113**, 120402 (2014).
- Scazza, F., Hofrichter, C., Höfer, M., De Groot, P. C., Bloch, I. & Fölling, S. Observation of two-orbital spin-exchange interactions with ultracold SU(N)-symmetric fermions. *Nat. Phys.* **10**, 779–784 (2014).
- Hofrichter, C. et al. Direct probing of the Mott crossover in the SU(N) Fermi-Hubbard model. *Phys. Rev. X* **6**, 021030 (2016).
- Ozawa, H., Taie, S., Takasu, Y. & Takahashi, Y. Antiferromagnetic spin correlation of SU(N) Fermi gas in an optical superlattice. *Phys. Rev. Lett.* **121**, 225303 (2018).
- Goban, A. et al. Emergence of multi-body interactions in a fermionic lattice clock. *Nature* **563**, 369–373 (2018).
- Pagano, G. et al. A one-dimensional liquid of fermions with tunable spin. *Nat. Phys.* **10**, 198–201 (2014).
- Song, B. et al. Evidence for Bosonization in a three-dimensional gas of SU(N) fermions. *Phys. Rev. X* **10**, 041053 (2020).
- He, C. et al. Collective excitations in two-dimensional SU(N) Fermi gases with tunable spin. *Phys. Rev. Res.* **2**, 012028(R) (2020).
- Sonderhouse, L. et al. Thermodynamics of a deeply degenerate SU(N)-symmetric Fermi gas. *Nat. Phys.* **16**, 1216–1221 (2020).
- Carleo, G. & Troyer, M. Solving the quantum many-body problem with artificial neural networks. *Science* **355**, 602–606 (2017).
- Carrasquilla, J. & Melko, R. G. Machine learning phases of matter. *Nat. Phys.* **13**, 431–434 (2017).
- Deng, D.-L., Li, X. & Das Sarma, S. Quantum entanglement in neural network states. *Phys. Rev. X* **7**, 021021 (2017).
- Zhang, Y. et al. Machine learning in electronic-quantum-matter imaging experiments. *Nature* **570**, 484–490 (2019).
- Rem, B. S. et al. Identifying quantum phase transitions using artificial neural networks on experimental data. *Nat. Phys.* **15**, 917–920 (2019).

25. Feng, L., Hu, J., Clark, L. W. & Chin, C. Correlations in high-harmonic generation of matter-wave jets revealed by pattern recognition. *Science* **363**, 521–524 (2019).
26. Bohrdt, A. et al. Classifying snapshots of the doped Hubbard model with machine learning. *Nat. Phys.* **15**, 921–924 (2019).
27. Pearson, K. On lines and planes of closest fit to systems of points in space. *Phil. Mag.* **2**, 559–572 (1901).
28. Jackson, J. E. *A User's Guide to Principal Components* (Wiley, 2005).
29. Sanner, C. et al. Suppression of density fluctuations in a quantum degenerate fermi gas. *Phys. Rev. Lett.* **105**, 040402 (2010).
30. Tan, S. Energetics of a strongly correlated Fermi gas. *Ann. Phys.* **323**, 2952–2970 (2008).
31. Tan, S. Large momentum part of a strongly correlated Fermi gas. *Ann. Phys.* **323**, 2971–2986 (2008).
32. Tan, S. Generalized virial theorem and pressure relation for a strongly correlated Fermi gas. *Ann. Phys.* **323**, 2987–2990 (2008).
33. Stewart, J. T., Gaebler, J. P., Drake, T. E. & Jin, D. S. Verification of universal relations in a strongly interacting Fermi gas. *Phys. Rev. Lett.* **104**, 235301 (2010).
34. Kuhnle, E. D. et al. Universal behavior of pair correlations in a strongly interacting fermi gas. *Phys. Rev. Lett.* **105**, 070402 (2010).
35. Partridge, G. B., Strecker, K. E., Kamar, R. I., Jack, M. W. & Hulet, R. G. Molecular probe of pairing in the BEC-BCS crossover. *Phys. Rev. Lett.* **95**, 020404 (2005).
36. Fletcher, R. J. et al. Two- and three-body contacts in the unitary Bose gas. *Science* **355**, 377–380 (2017).
37. Laurent, S. et al. Connecting few-body inelastic decay to quantum correlations in a many-body system: a weakly coupled impurity in a resonant Fermi gas. *Phys. Rev. Lett.* **118**, 103403 (2017).
38. Yu, Z., Thywissen, J. H. & Zhang, S. Universal relations for a Fermi gas close to a p-wave interaction resonance. *Phys. Rev. Lett.* **115**, 135304 (2015).
39. Luciuk, C. et al. Evidence for universal relations describing a gas with p-wave interactions. *Nat. Phys.* **12**, 599–605 (2016).
40. Mueller, T. et al. Local observation of antibunching in a trapped Fermi gas. *Phys. Rev. Lett.* **105**, 040401 (2010).
41. Tobias, W. G. et al. Thermalization and sub-Poissonian density fluctuations in a degenerate molecular Fermi gas. *Phys. Rev. Lett.* **124**, 033401 (2020).
42. Yip, S. K., Huang, B.-L. & Kao, J.-S. Theory of SU(N) Fermi liquids. *Phys. Rev. A* **89**, 043610 (2014).
43. Bruun, G. M. & Clark, C. W. Ideal gases in time-dependent traps. *Phys. Rev. A* **61**, 061601(R) (2000).
44. Gupta, S., Hadzibabic, Z., Anglin, J. & Ketterle, W. Collisions in zero temperature Fermi gases. *Phys. Rev. Lett.* **92**, 100401 (2004).
45. Song, B. et al. Spin-orbit-coupled two-electron Fermi gases of ytterbium atoms. *Phys. Rev. A* **94**, 061604(R) (2016).
46. Song, B. et al. Effective statistical fringe removal algorithm for high-sensitivity imaging of ultracold atoms. *Phys. Rev. Appl.* **14**, 034006 (2020).
47. Abadi, M. et al. Tensorflow: A system for large-scale machine learning. In *Proc. 12th USENIX symposium on operating systems design and implementation (OSDI)*, 265–283 (2016).
48. Kingma, D. P. and Ba, J. Adam: a method for stochastic optimization. In *3rd International Conference on Learning Representations (ICLR)* (2015).

Acknowledgements

We thank Tin-Lun Ho for helpful discussion. G.-B.J. acknowledges supports from the RGC and the Croucher Foundation through 16311516, 16305317, 16304918, 16306119, C6005-17G, and N-HKUST601/17. This work is partially supported by the Innovative Exploratory Grant at HKUST. J.L. acknowledges supports from the RGC (N-HKUST626/18, 26302118, and 16305019).

Author contributions

E.Z. was responsible for the analysis of the work. E.Z., C.H., Z.R. and E.H. implemented the physical experiment. J. Lee., J. Liu. and G.-B.J. conceived the project and supervised the research. All authors contributed to discussions regarding the results and analysis contained in the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-22270-5>.

Correspondence and requests for materials should be addressed to G.-B.J.

Peer review information *Nature Communications* thanks Florian Schäfer and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021