Visualizing strange metallic correlations in the two-dimensional Fermi-Hubbard model with artificial intelligence

Ehsan Khatami,[1,*] Elmer Guardado-Sanchez,[2] Benjamin M. Spar,[2] Juan Felipe Carrasquilla,[3] Waseem S. Bakr,[2] and Richard T. Scalettar[4]

[1]*Department of Physics and Astronomy, San José State University, San José, California 95192, USA*
[2]*Department of Physics, Princeton University, Princeton, New Jersey 08544, USA*
[3]*Vector Institute, MaRS Centre, Toronto, Ontario, Canada M5G 1M1*
[4]*Department of Physics, University of California, Davis, California 95616, USA*

(Received 29 May 2020; accepted 20 August 2020; published 17 September 2020)

Strongly correlated phases of matter are often described in terms of straightforward electronic patterns. This has so far been the basis for studying the Fermi-Hubbard model realized with ultracold atoms. Here, we show that artificial intelligence (AI) can provide an unbiased alternative to this paradigm for phases with subtle, or even unknown, patterns. Long- and short-range spin correlations spontaneously emerge in filters of a convolutional neural network trained on snapshots of single atomic species. In the less well-understood strange metallic phase of the model, we find that a more complex network trained on snapshots of local moments produces an effective order parameter for the non-Fermi-liquid behavior. Our technique can be employed to characterize correlations unique to other phases with no obvious order parameters or signatures in projective measurements, and has implications for science discovery through AI beyond strongly correlated systems.

DOI: 10.1103/PhysRevA.102.033326

I. INTRODUCTION

Strongly correlated phases of matter are often described in terms of relatively simple real space order parameters, which are theoretically understood using Landau symmetry-breaking theory [1]. For instance, ferromagnetism on a square lattice involves a uniform arrangement where the electrons' spins align and create a magnetic state with a wave vector $\mathbf{q} = \mathbf{0}$. Antiferromagnetism, slightly more complex, is revealed by a $\mathbf{q} = \pi$ alternation of the electrons' spin state on two sublattices. These choices, and incommensurate (spiral) order which bridges them at general $\mathbf{q}$, can be characterized in a unified way through the magnetic structure factor, $S(\mathbf{q})$, and further generalized to include time-domain patterns via the dynamic susceptibility, $\chi(\mathbf{q}, \omega)$. Similar statements apply to charge density wave and other phases involving diagonal long-range order.

While many of our theoretical and experimental probes of interacting quantum systems have been constructed with coupling to these patterns in mind, there is an increasing realization that the most interesting strongly correlated phases might not be immediately accessible via such observables. Cuprate and iron pnictide superconductors, which combine closely entwined conventional phases with well-established order parameters, and much less well-understood non-Fermi-liquid (NFL) or pseudogap phases with so far "hidden orders" are examples [2–4], as is the zoo of orbital ferromagnetism, superconductivity, and Mott-insulating behavior in twisted bilayer graphene [5,6]. The community of strongly correlated quantum systems is thus faced with the challenge of developing new means of identifying complex phases.

Here, we introduce an unbiased approach in which artificial intelligence (AI) is used to extract hidden features from raw images of quantum many-body systems. We test our approach using projective measurements on a two-dimensional (2D) Fermi-Hubbard model, obtained through quantum gas microscopy of ultracold fermionic atoms in an optical lattice. We find that filters of a convolutional neural network (CNN), trained to recognize snapshots of fermions, capture features at different densities that have clear interpretation in terms of short- and long-range magnetic correlations. We further show that a more complex CNN can produce an effective order parameter for the NFL phase, based on the interplay of multiple types of density fluctuations, reflecting the more enigmatic nature of the correlations in this phase.

In the experiment, the 2D Fermi-Hubbard model is realized using a spin-balanced mixture of the first and third lowest energy states of ^{6}Li loaded into a square optical lattice. We work at a magnetic field of 615 G in the vicinity of the Feshbach resonance near 690 G, which gives a scattering length of $1056(10)a_0$, where a_0 is the Bohr radius. The lattice depth is $7.25(2)E_R$, where E_R is the lattice recoil energy and $E_R/h = 14.66$ kHz. For these parameters we obtain $t/h = 850(20)$ Hz and $U/t = 8.0(1)$. Here, t and U are the nearest-neighbor hopping matrix element and the strength of on-site repulsive interaction, respectively, in the Hubbard model (see Appendix A).

Using quantum gas microscopy techniques [7], we image the atoms in the lattice with single site resolution with a fidelity of 98%. When a fluorescence image is taken, atoms on doubly occupied sites undergo light-assisted collisions and

*ehsan.khatami@sjsu.edu

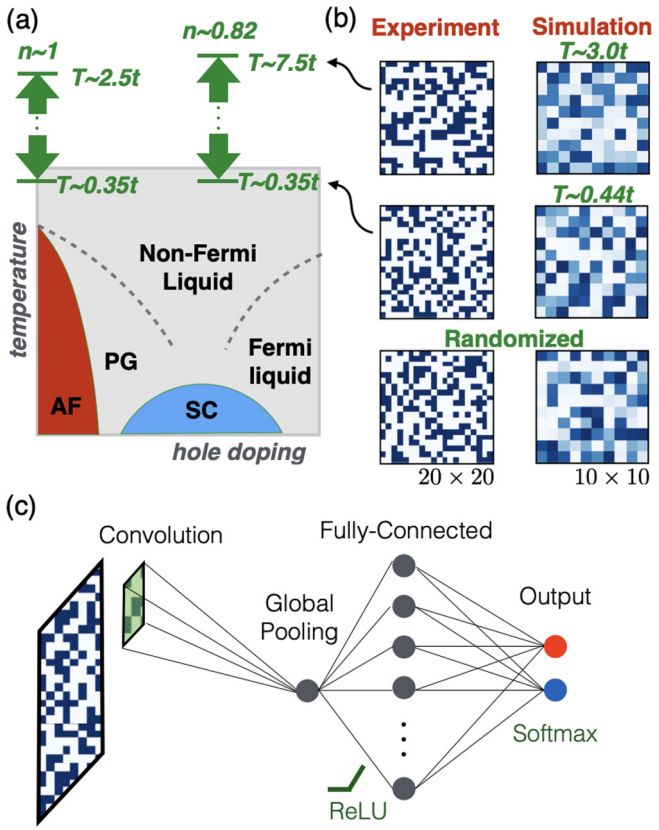

FIG. 1. Phase diagram, sample snapshots, and CNN architecture. (a) Schematic phase diagram of cuprate high-temperature superconductors in the space of temperature and hole doping. AF, PG, and SC stand for antiferromagnetic, pseudogap, and superconducting phases, respectively. (b) Sample experimental (left) and determinant quantum Monte Carlo (right) snapshots taken for $U = 8t$ at the density $n \sim 0.82$ at different temperatures. The bottom row shows the lower-temperature snapshots whose pixels have been randomly shuffled, i.e., "fake" snapshots. (c) The main convolutional neural network architecture used in this study. It contains a convolutional layer with one filter and one feature map followed by a global pooling layer, a hidden layer with eight fully connected neurons, and an output softmax layer with two neurons, each associated with a temperature limit. We use the rectified linear unit (ReLU) as the activation function in all but the output layer. In our experiments we observe that the presence of the fully connected layer accelerates the training of the neural network.

appear empty. An image taken this way allows us to extract the local moment on each site. Alternatively, we can apply a short pulse of resonant light prior to taking an image to eject atoms of one of the two hyperfine states. This allows us to measure the single component density of the remaining hyperfine state.

Our lattice beams produce a harmonic trapping potential, which if uncompensated leads to significant variations of the local density. To study regions of uniform density, we flatten the potential using light shaped using a spatial light modulator [8]. In the subsequent analysis, we work with a flattened region of 20×20 lattice sites.

Figure 1 shows two randomly chosen samples of binarized occupancy snapshots at an average density of $n = 0.82(2)$ at two extreme temperatures of $T \sim U \sim 8\,t$ and $T \sim 0.35\,t$. These parameters place us within the NFL region of a typical

TABLE I. Number of available experimental snapshots at $T \sim 0.35$ for different densities.

$n \rightarrow$	0.97	0.835	0.82	0.735	0.70	0.64	0.58
Spin-up/spin-down	402		216				
Singles	201	281	5023	290	342	281	330

cuprate phase diagram [see Fig. 1(a)]. Thermometry is performed using averages of various correlation functions taken over such snapshots [8].

The increasingly large number of snapshots taken in quantum gas microscope experiments in various regions of the parameter space lends itself to data-driven approaches for science discovery, such as the enlisting of AI (see Table I for the number of snapshots used in this study). In fact, early implementations of machine learning techniques for the study of quantum many-body systems demonstrated great potential [9–15]. Recent applications to experimental data have directly led to the discovery of new physics [16–20], modeling of their distribution [21], or the optimization of experimental processes [22,23], including those related to quantum gas microscopy.

CNNs offer an ideal platform for the detection of patterns in the experimental snapshots. Not only can they efficiently compress the information in images and use them for classification, but also their trained filters provide a window into the relevant features observed [24]. Figure 1(c) shows the main CNN architecture we have used. After labeling them according to their temperature, hundreds of snapshots taken at the extreme temperatures along with their labels are provided to the CNN for training. During the training, the network adjusts its free parameters to minimize the difference between given labels and its prediction (see Appendix B). The convolutional layer in our CNN interacts directly with the input snapshots and, therefore, examining the filter after the completion of training can teach us about the most important feature the network has picked up.

II. RESULTS

Figure 2(a) shows a sample 5×5 filter for a CNN that is trained to distinguish experimental snapshots of a single species of fermions at the highest temperature ($T \sim 2.5t$) from those at the lowest temperature ($T \sim 0.35t$) when $n \sim 1$. If we expect mostly random behavior at high temperature, of the same order as the largest energy scale in the system, the features that spontaneously develop in the filters during training will most likely represent patterns found in the low-temperature snapshots. We find that the CNN consistently makes the distinction with more than 91% accuracy, and it does so using filters showing a distinctive pattern indicative of long-range antiferromagnetic (AF) correlations.

Training the CNN using similar snapshots obtained for $n = 0.82$ at $T \sim 7.5t$ and $0.35t$ results in filters that reflect a shorter-range anticorrelation between neighboring fermions of the same species [see Fig. 2(b)]. The nearest-neighbor checkerboard pattern emerging in the filters is consistent with the fact that the correlation length in the NFL region is about one lattice spacing [25]. We find that this feature appears

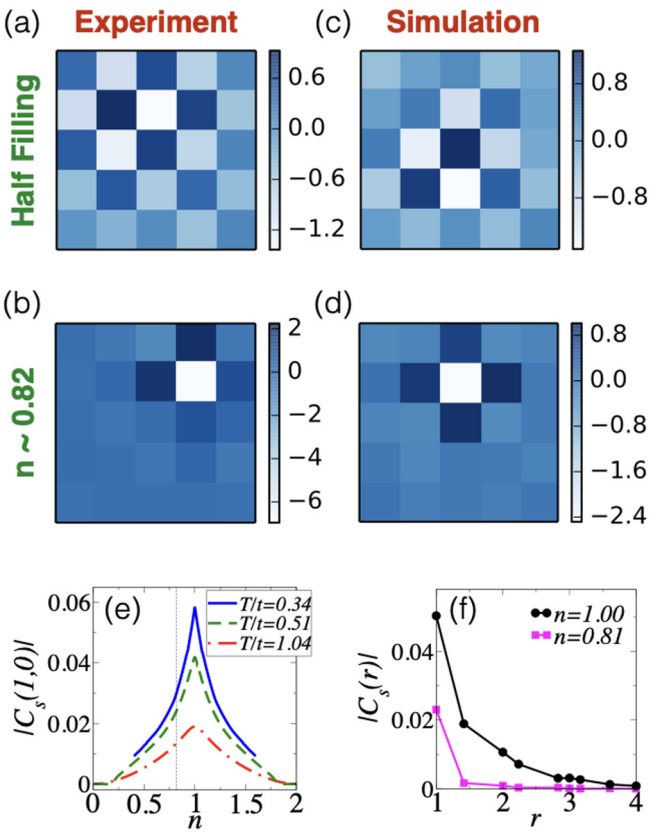

FIG. 2. Analysis of single-species snapshots using CNNs with one filter. Sample 5×5 trained filters for (a) $n \sim 1$ and (b) $n = 0.82$. The CNN architecture is shown in Fig. 1(c). The testing accuracies are between 91% and 96%. The visual pattern in (a) is consistent with recognizing long-range antiferromagnetic (AF) order near half filling. The filter in (b) indicates a pattern capturing short-range AF correlations. (c), (d) Similar sample filters evolved from training runs using determinant quantum Monte Carlo (DQMC) simulations at $T = 0.1t$ and $0.44t$, with testing accuracies of 91% and 68%, respectively. (e), (f) Theory data for the nearest-neighbor spin-spin correlation for $U = 8t$ vs density at different temperatures (numerical linked cluster expansion), and vs distance for $n = 0.81$ and $n = 1.00$ at $T = 0.44t$ (DQMC). These results illustrate that AI can capture the correct trends in magnetic behavior of the Hubbard model, and that the trained filters carry a clear physical interpretation.

at different locations in the filter for different training runs, which points to a redundancy: on average the filter must reflect the translational symmetry of the underlying system. These findings suggest that the network effectively uses the strength of AF correlations as a measure for classifying snapshots of a single species of fermions. Figures 2(e) and 2(f) show that the density, distance, and temperature dependence of the magnetic correlations of the model, $C_s(\mathbf{r})$ (see Appendix A), which are calculated here on a 10×10 cluster using the determinant quantum Monte Carlo (DQMC) method [26], or in the thermodynamic limit using the numerical linked cluster expansion (NLCE) [27,28], support this observation. We note that the unbiased quantum Monte Carlo simulations also provide a platform to corroborate these findings. However, except in one spatial dimension, these simulations cannot provide projective measurements in the density basis. Instead, theory "snapshots" can be constructed via expectation values of local charge or spin density using instances of auxiliary field variables during a simulation; for example, the ith pixel of a spin-up DQMC snapshot is $\langle \hat{n}_{i\uparrow} \rangle_h = 1 - \mathcal{G}_{ii\uparrow}(h)$, where $\mathcal{G}_{ii\uparrow}(h)$ is the ith diagonal element of the spin-up equal time Green's function matrix for the auxiliary field instance h. We perform the simulations for a 10×10 site Hubbard system with $U = 8t$ at several average average densities and temperatures (see Appendix D).

At high temperatures, of the order of $3t$, we find that density snapshots are fuzzy with no clear empty sites; mostly fluctuations about an average background density can be seen. This fuzziness is less of a concern for single-species snapshots [see Fig. 1(b)], although they too lose their pixelated character at higher temperatures. For this reason, to eliminate fuzziness as an obvious feature for the CNN to learn, instead of high-temperature snapshots, we use low-temperature images whose pixels have been randomly shuffled, effectively destroying any physical correlations. In the following, we refer to the latter as *fake* (as opposed to *real*) snapshots.

Figures 2(c) and 2(d) show sample filters from training experiments using theory snapshots of single species at half filling and $n = 0.82$. Despite reduced accuracies of about 68% for the latter, which we believe is due to the exacerbation of the issue with the nonprojective nature of simulated images at this density, we find that the trained features are in excellent agreement with those obtained with quantum gas microscope snapshots. Together, they demonstrate that relevant spin correlations can be captured in an unbiased fashion through CNNs.

Studies of the origin of the NFL behavior, a central question in any theory of high-temperature superconductivity [29], have for decades been focused on its possible connections to the order parameter fluctuations of a magnetic quantum critical point [29–35]. Here, we are in a position to ask whether any such fluctuations manifest themselves in charge correlations too, and to what extent they can be inferred from the other type of snapshots available in the experiment, those of local moments.

A similar analysis using images at the two extreme temperatures, however, is largely affected by the abundance of doubly occupied sites at $T \sim 7.5t$, and their lack of representation in the snapshots of local moments. Upon lowering the temperature to $T \sim 0.35t$, the fraction of doubly occupied sites at 18% doping reduces roughly by a factor of 4, from 12% to about 3% [28], providing the CNN again with an obvious feature with which to perform classification. Removing this bias by randomly populating pixels to create "fake" replacements for high-temperature snapshots in the training brings the accuracy down dramatically when $n = 0.82$.

Figure 3 shows that the accuracy of CNNs trained on snapshots of local moments largely depends on the strength of short-range correlations between the moments. The largest accuracies (almost 90%) are typically achieved near quarter filling, where the correlations are the most negative. Patterns observed in filters trained in this region are also consistent with the anticorrelation of neighboring moments [Fig. 3(a)]. The accuracy drops to around 60% at $n = 0.82$, near the zero crossing of the correlator, which is shown in Fig. 3(b). Typical trained filters do not display any immediately recognizable patterns either [see Fig. 3(c)]. Near half filling, the correla-

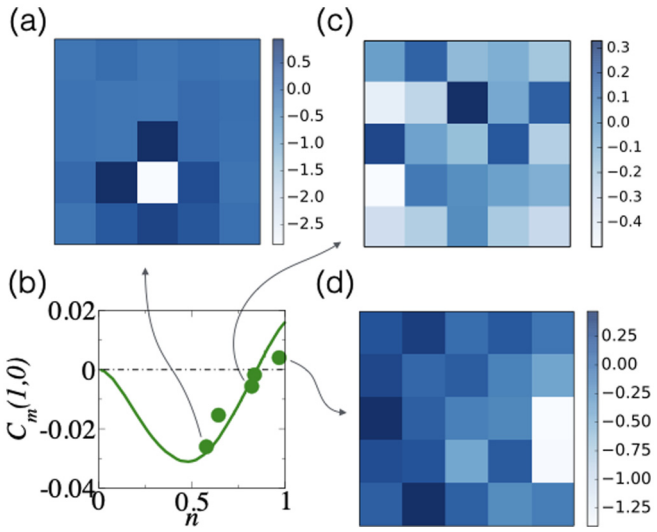

FIG. 3. Analysis of local moment snapshots using CNNs with one filter. Representative 5×5 filters for runs at (a) $n = 0.58$, (c) $n = 0.82$, and (d) $n = 0.97$ using the same CNN architecture as one used for snapshots of single species. (b) The nearest-neighbor local moment correlation function from DQMC (solid line) on an 8×8 cluster at $T = 0.3t$ and from the experiment (circles) at similar low temperatures. The testing accuracy reaches (a) 89%, (c) 62%, and (d) 60%.

tions between local moments are positive due to the bunching of holes and doubly occupied sites [36]. Here, we find that, despite the relatively low accuracies ($\lesssim 65\%$), trained filters often do reflect the bunching of empty sites [Fig. 3(d)].

The comparison of Fig. 3(c) ($n = 0.82$) with Figs. 3(a) and 3(d) ($n = 0.58, 0.97$) makes it clear that snapshots of local moments in the NFL region of the Hubbard model do not contain a single dominant ordering pattern and that a more advanced treatment may be necessary to capture the physics. In Fig. 4, we show results of a training with a CNN, modified to include six 7×7 filters in its convolutional layer (see Appendix B). The bigger data set we have available for snapshots of local moments at this density allows us to experiment with different filter sizes and numbers of filters. We find that including more than one filter in the CNN improves the best accuracies only marginally in this case, up to around 65%, and having too many and/or much larger filters can still result in overfitting. We also find that using a deeper CNN with two convolutional layers does not significantly improve the accuracy.

Figure 4(a) shows six filters of a sample CNN trained on the local moment snapshots. Their fuzzy patterns offer some insight into possible spatial arrangements of local moments at low temperatures. As we see below, patterns in filters $m = 1$, 4, 5, and 6 are more frequently associated by the network with real low-temperature snapshots at this filling, whereas patterns in filter $m = 2$ are more frequently associated with fake snapshots.

By transferring the knowledge of the CNN to other densities, we find that the network is the most sensitive to correlations around the NFL region. Figure 4(b) shows the difference in probabilities that a snapshot and its fake counterpart are categorized as belonging to the NFL region, effectively eliminating density itself as a factor in the signal. We find this quantity to be maximal in the vicinity of $n = 0.8$, suggesting that the CNN as a whole is in fact focusing on local moment correlations more unique to the NFL region and slightly lower densities.

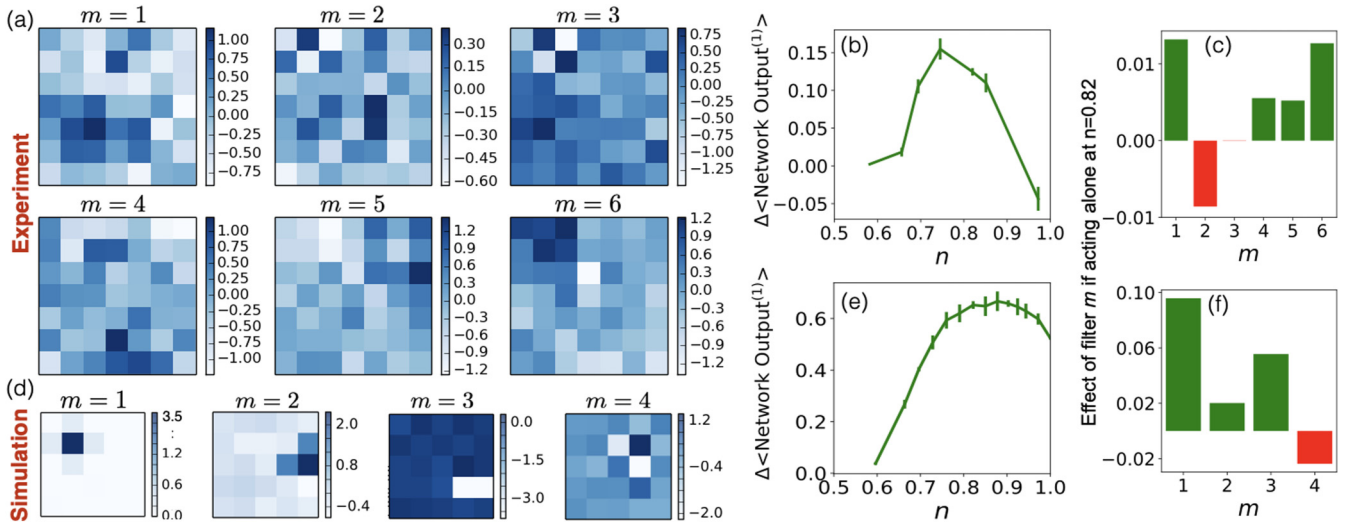

FIG. 4. Analysis of local moment snapshots using CNNs with multiple filters. (a) The six filters of a trained CNN. Training is performed with the 5023 experimental local moment snapshots taken at $n = 0.82$ and $T \sim 0.35$. The testing accuracy remains at 62% with less than 1% variation over the last 20 epochs (see Appendix B). (b) The difference in the average network output for $T \sim 0.35t$ real and fake snapshots as a function of the density when all six filters are present: $\Delta\langle\text{Network Output}^{(1)}\rangle \equiv \langle\text{Network Output}^{(1)}(\mathbf{X}^{\text{real}})\rangle - \langle\text{Network Output}^{(1)}(\mathbf{X}^{\text{fake}})\rangle$. Superscript (1) indicates the value at the output neuron responsible for real low-temperature snapshots (see Appendix B). This quantity indicates roughly the percentage of the output attributable to factors other than the density. (c) Similar to (b) at $n = 0.82$ when the CNN has access to one filter at a time. (d) Four representative filters of a CNN with sixteen 5×5 filters trained using DQMC snapshots of local moments (see Appendix D). (e), (f) Same as (b) and (c), but obtained using the CNN in (d).

While the contribution of individual filters to the CNN's decision making cannot be completely isolated, we can study what the network output would be if each filter were to act alone (see Appendix B). Figure 4(c) shows this quantity averaged over samples at $n = 0.82$, after subtracting the value for the corresponding fake snapshot, for each of the six filters shown in Fig. 4(a). The results suggest that filters 1 and 6, if acting alone, would have the largest effect on the decision making at this average density, followed by filters 2, 4, and 5, while filter 3 plays almost no role at all. The negative value for filter 2 indicates that the network signal is larger on average for fake snapshots in that case.

Using DQMC, we verify that similar trends can be observed in simulated snapshots of local moments. However, unlike with the experimental snapshots, here, we find that the accuracy generally increases with increasing the number of filters in the CNN, while increasing the filter size does not necessarily improve the performance. We attribute these to the fundamental difference between the two types of snapshots (projective vs nonprojective). Figure 4(d) highlights a representative sample of 5×5 filters of a CNN with 16 such filters, trained on simulated snapshots reaching to an accuracy of 87% (see Appendix C). They appear to measure a variety of short-range correlations to assist the network in making decisions. Figure 4(e) shows the overall signal of the CNN for correlations unique to the NFL phase, plotted across densities. It has a broad peak around the NFL region. As shown in Fig. 4(f), patterns in the first three filters seem to be mostly associated with real snapshots in the NFL region, while the pattern in the $m = 4$ filter is mostly associated with the fake snapshots. We find that including the information about doubly occupied sites, i.e., using full density snapshots, generally improves the diversity of features seen in the trained filters while yielding the same basic trends.

III. DISCUSSION

Using specially designed artificial neural networks, we have developed algorithms for extracting organizing patterns of correlated particles from raw quantum many-body data in an unbiased fashion and without any prior theoretical knowledge. When applied to the snapshots of one of the two species of fermionic atoms in a 2D optical lattice, our approach yields patterns indicative of long- and short-range AF correlation near and away from the commensurate filling, consistent with theory. We show that these features can be reproduced using nonprojective measurements from DQMC simulations.

Our analysis provides a window into the signatures of the NFL phase, one of the most mysterious and theoretically challenging phases in correlated electron systems, in snapshots of local moments from the quantum gas microscope. We show that in this case, a more complex neural network can be constructed and trained to be sensitive to correlations specific to the strange metallic phase. A similar analysis of snapshots with information about both species of particles in future experiments [37] may further reveal the interplay between spin and charge fluctuations in this region.

Nonlinearities in the neural network model make the interpretation of features seen in the filters vis-à-vis correlations in the physical snapshots challenging since the knowledge of the network can be divided in nontrivial ways among its different components. An early example of this was the surprisingly successful classification of snapshots of the Ising lattice gauge theory at $T = 0$ and $T = \infty$, despite the lack of an order parameter, using CNNs with multiple filters [10]. The procedure we have introduced here overcomes aspects of this challenge. We point out that the nonlinearities are key for the success of our CNNs; for example, the principal component analysis [38], a linear unsupervised learning method, largely fails to draw any meaningful distinction between sets of snapshots (see Appendix E).

The techniques developed in this work for the AI-assisted feature extraction in projective measurements can be adapted to peek into other mysterious phenomena for the Fermi-Hubbard model, such as the pseudogap phase [39], or the magnetic polaron which has been observed closer to half filling [40]. They can also be employed to study other microscopic models of correlated systems. Our work paves the way for AI-related studies that go beyond mere categorization and the quest for gaining more predictive power and focus instead on the inner workings of the machines to advance our understanding of complicated natural phenomena.

ACKNOWLEDGMENTS

We thank Christie Chiu, Annabelle Bohrdt, and Neil Switz for useful discussions. E.K. acknowledges support from the National Science Foundation (NSF) under Grant No. DMR-1609560. Computations were performed in part on the Spartan high-performance computer at San José State University, which is supported by the NSF under Grant No. OAC-1626645. The work of R.T.S. was supported by Grant No. DE-SC0014671 funded by the U.S. Department of Energy, Office of Science. The work of E.G.-S., B.M.S., and W.S.B. was supported by the NSF under Grant No. DMR-1607277, the David and Lucile Packard Foundation under Grant No. 2016-65128, and the AFOSR Young Investigator Research Program under Grant No. FA9550-16-1-0269. J.C. acknowledges support from the Natural Sciences and Engineering Research Council of Canada (NSERC), the Shared Hierarchical Academic Research Computing Network (SHARCNET), Compute Canada, Google Quantum Research Award, and the Canada CIFAR AI chair program.

APPENDIX A: THE FERMI-HUBBARD MODEL

The Hamiltonian for the 2D Fermi-Hubbard model in particle-hole-symmetric form is expressed as

$$\begin{aligned}\hat{H} = &-t \sum_{\langle \mathbf{i},\mathbf{j}\rangle\,\sigma} \left(\hat{c}^{\dagger}_{\mathbf{i}\sigma}\hat{c}_{\mathbf{j}\sigma} + \text{H.c.}\right) \\ &+ U \sum_{\mathbf{i}} \left(\hat{n}_{\mathbf{i}\uparrow} - \frac{1}{2}\right)\left(\hat{n}_{\mathbf{i}\downarrow} - \frac{1}{2}\right) \\ &- \mu \sum_{\mathbf{i}} (\hat{n}_{\mathbf{i}\uparrow} + \hat{n}_{\mathbf{i}\downarrow}),\end{aligned} \tag{A1}$$

where $\hat{c}^{\dagger}_{\mathbf{i}\sigma}$ ($\hat{c}_{\mathbf{i}\sigma}$) creates (annihilates) a fermion with spin σ on site $\mathbf{i}$, and $\hat{n}_{\mathbf{i}\sigma} = \hat{c}^{\dagger}_{\mathbf{i}\sigma}\hat{c}_{\mathbf{i}\sigma}$ is the number operator. $\langle \mathbf{i}, \mathbf{j}\rangle$ denotes nearest neighbors on a square lattice, $U = 8t$ is the strength of the on-site repulsive interaction in the numerical simulations,

and μ is the chemical potential. $\mu = 0$ corresponds to half filling, although density fluctuations around half filling exist in our grand canonical ensemble. $t = 1$ (also $\hbar = 1$ and $k_B = 1$) sets the energy scale. The spin correlation function is calculated as $C_s(\mathbf{r}) = \langle \hat{S}_{z,\mathbf{i}} \hat{S}_{z,\mathbf{i}+\mathbf{r}} \rangle$, where $\hat{S}_{z,\mathbf{i}} = \frac{1}{2}(\hat{n}_{\mathbf{i}\uparrow} - \hat{n}_{\mathbf{i}\downarrow})$ and $\langle \cdots \rangle$ denotes the expectation value. The local moment correlation function is calculated as $C_m(\mathbf{r}) = \langle \hat{m}^2_{z,\mathbf{i}}\, \hat{m}^2_{z,\mathbf{i}+\mathbf{r}} \rangle$, where $\hat{m}^2_{z,\mathbf{i}} = (\hat{n}_{\mathbf{i}\uparrow} - \hat{n}_{\mathbf{i}\downarrow})^2$.

APPENDIX B: TRAINING THE CONVOLUTIONAL NEURAL NETWORK

We implement our CNNs using TENSORFLOW [41]. The minimalistic design in Fig. 1(c) we have adopted reflects the need to reduce the number of free parameters to avoid overfitting given the sizes of our data sets. To train, we assign a label, Y, to each snapshot based on the temperature at which it is taken, or whether it is real or fake. Each label is stored in the one-hot format, i.e., a binary array of two numbers, one of which is 1 and the other 0. The index for 1 indicates the category (high or low temperature, or real or fake) to which each snapshot belongs. Given an input image $\mathbf{X}$, the value arriving at the first of the two output neurons of the CNN shown in Fig. 1(c), e.g., at the neuron we have associated in our labels to the low-temperature (or real) snapshots, is

$$O_1^{\text{out}}(\mathbf{X}) = \sum_h O_h^{\text{hid}}(\mathbf{X}) \times W_h^{\text{out (1)}} + b^{\text{out (1)}}, \tag{B1}$$

where the sum is over hidden neurons,

$$O_h^{\text{hid}}(\mathbf{X}) = \text{ReLU}\left(\left[\frac{1}{N_s}\sum_{\text{stride:}s} \text{ReLU}\left(\mathbf{W}^{\text{filter}} \cdot \mathbf{X}(s) + b^{\text{filter}}\right)\right] \times W_h^{\text{hid}} + b_h^{\text{hid}}\right), \tag{B2}$$

N_s is the number of strides the filter takes around the image convolving with different sections, $\mathbf{W}^{\text{filter}}$ is the matrix of pixel values for the filter, $\mathbf{X}(s)$ is the matrix of pixel values for the section of the image the filter is convolving with in stride s, ReLU is the rectified linear unit activation function, and b^{filter}, W_h^{hid}, b_h^{hid}, $W_h^{\text{out}(1)}$, and $b^{\text{out}(1)}$ are numbers representing other weights and biases in the network. O_1^{out}, along with the value arriving at the second output layer $O_2^{\text{out}}(\mathbf{X})$, are then passed through the softmax activation function to obtain two probabilities as network outputs:

$$\left[\text{Network Output}^{(1)}(\mathbf{X}),\ \text{Network Output}^{(2)}(\mathbf{X})\right] = \text{softmax}\left[O_1^{\text{out}}(\mathbf{X}),\ O_2^{\text{out}}(\mathbf{X})\right]. \tag{B3}$$

The input snapshot is classified as belonging to category i if O_i^{net} is the higher probability, where $O_i^{\text{net}} = \text{Network Output}^{(i)}$ for brevity. The accuracy is defined as the percentage of correct classifications given known labels Y. The convolution of the trained filter with sections of the input image as it moves around in strides of one in every direction creates a "feature map" in which large overlaps between patterns in the filter and the image are highlighted.

For training, we use the Adam optimizer, which is an extension of stochastic gradient descent, to minimize the cross-entropy cost function, defined as

$$c = -\frac{1}{N_d}\sum_{\mathbf{X}}\sum_{i=1}^{2}\left(Y_i(\mathbf{X}) \ln\left[O_i^{\text{net}}(\mathbf{X})\right] + [1 - Y_i(\mathbf{X})] \ln\left[1 - O_i^{\text{net}}(\mathbf{X})\right]\right), \tag{B4}$$

where N_d is the number of data. During the training, we keep between 10% and 20% of the snapshots from the CNN and use them to perform unbiased testing of the accuracy.

1. CNN with more than one filter

In cases where we have more than one filter in the convolutional layer, we have modified the architecture to have no fully connected hidden layer in order to reduce the total number of network parameters; the output of each filter after pooling is instead fully connected to the output layer. The value arriving at the output neuron that is responsible for firing when a real snapshot $\mathbf{X}$ is provided to the input, $O_1^{\text{out}}(\mathbf{X})$, can then be expressed as a linear combination of contributions from individual filters:

$$O_1^{\text{out}}(\mathbf{X}) = \sum_{m=1}^{N_f} F_m^{(1)}(\mathbf{X}), \tag{B5}$$

$$F_m^{(1)}(\mathbf{X}) = \left[\frac{1}{N_s}\sum_{\text{stride:}s} \text{ReLU}\left(\mathbf{W}_m^{\text{filter}} \cdot \mathbf{X}(s) + b_m^{\text{filter}}\right) \times W_m^{\text{out}(1)}\right] + \frac{b^{\text{out}(1)}}{N_f}, \tag{B6}$$

where N_f is the number of filters, and $W_m^{\text{out}(1)}$ and b_m^{filter} are again numbers representing other weights and biases in the network. As in the case of the CNN with one filter, the network output is obtained using Eq. (B3).

2. Effect of individual filters

To estimate the effect of filter m on the outcome, we replace $O^{(1,2)}(\mathbf{X})$ with $F_m^{(1,2)}(\mathbf{X})$ before the softmax function,

$$\left[\text{Network Output}_m^{(1)}(\mathbf{X}), \text{Network Output}_m^{(2)}(\mathbf{X})\right] = \text{softmax}\left[F_m^{(1)}(\mathbf{X}), F_m^{(2)}(\mathbf{X})\right], \tag{B7}$$

so that we can interpret $[\text{Network Output}_m^{(1)}(\mathbf{X}^{\text{real}}) - \text{Network Output}_m^{(1)}(\mathbf{X}^{\text{fake}})]$ as the percentage the network output for $\mathbf{X}$, based on the action of filter m alone, has to do with factors other than the density itself.

3. Augmentation of data

We augment [42] our data for the experimental single-species snapshots before training [Figs. 2(a) and 2(b)] by applying point-group symmetries of the square lattice to each snapshot. This will increase the data by a factor of 8, and in these cases helps make the training smoother and/or faster. However, we find that generally, such augmentation of data, or breaking the 20×20 images into smaller subregions, does not significantly affect the final accuracy. This is consistent with the fact that the network architectures we have used in

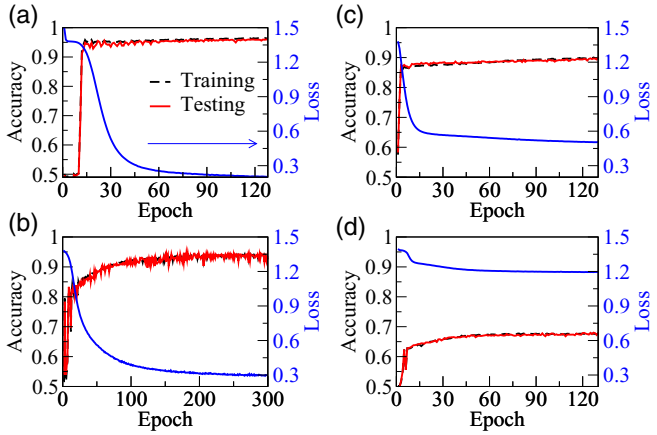

FIG. 5. Progression of CNN accuracy during training with single-species snapshots. (a) Training and unbiased testing accuracies, along with the value of the loss function when using experimental snapshots at $n \sim 1$ to train the CNN with one filter, leading to the filter shown in Fig. 2(a). (b–d) Same as (a), but for trainings leading to filters in Figs. 2(b)–2(d). DQMC snapshots are used in (c) and (d). (a) and (c) correspond to trainings near half filling and (b) and (d) correspond to trainings at $n = 0.82$.

this work are not deep and do not show serious overfitting even before the data augmentation.

4. Training progression

To monitor the training progression and look for signs of overfitting, especially in the case of CNNs with more than one filter, we track the training and the unbiased testing accuracies as well as the loss function, defined in Eq. (B4), over *epochs*. An epoch is when the network has gone over the entire data set once.

Figure 5 shows the results for various trainings using single-species snapshots that lead to filters presented in Fig. 2. In each case, we stop the training roughly when the loss function reaches a plateau. As can be seen, the training and the unbiased testing accuracies, which agree throughout the training process, also settle to their maxima.

Figure 6(a) shows similar results for the sample training with snapshots of local moments that lead to the filter presented in Fig. 3(c). For this case, we find that it takes longer

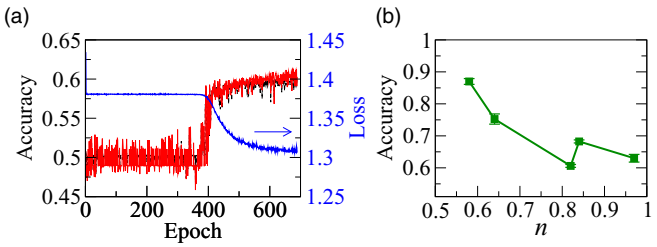

FIG. 6. Progression of CNN accuracy during training with local moment snapshots. (a) Same as Fig. 5(a), but for the training with snapshots of local moments at $n = 0.82$ leading to the filter shown in Fig. 3(c). (b) The average of maximum testing accuracies, obtained through four or five separate trainings, vs density.

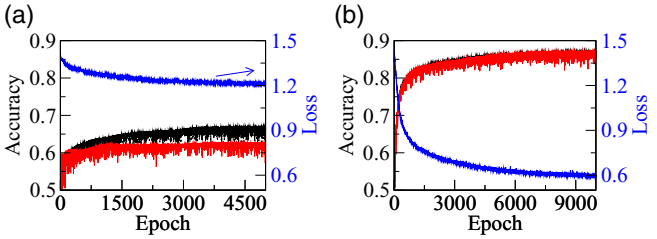

FIG. 7. Progression of accuracy of CNNs with multiple filters during trainings with local moment snapshots. Same as Fig. 5, except that (a) experimental and (b) DQMC snapshots of local moments are used to train CNNs with 6 and 16 filters, showcased in Fig. 4.

for any training to be achieved and that the fluctuations in the accuracies are larger. Figure 6(b) shows the average of maximum accuracies achieved through four or five different training runs with different random number seeds over the range of densities for which snapshots of local moments were available. We see that as the strength of local correlations decreases by increasing the density towards half filling [shown in Fig. 3(b)], the maximum accuracies we can achieve also generally decrease.

Figure 7 shows the same training progressions for the case of CNNs with multiple filters, trained on snapshots of local moments at $n = 0.82$, leading to results shown in Fig. 4. Despite the deviation of the average of two accuracies from each other beyond ∼1000 epochs when experimental snapshots are used [Fig. 7(a)], signaling the beginning of overfitting due to the relatively large number of free parameters in the CNN, large fluctuations in the accuracies cause overlapping of the two curves even after 5000 epochs. When DQMC snapshots are used, the signs of overfitting are observed only after increasing the number of 5×5 filters to about 60.

APPENDIX C: CNN WITH 16 FILTERS FOR THEORY SNAPSHOTS OF LOCAL MOMENTS

Figure 8 shows the result of training a CNN with 16 filters using 15 000 10×10 theory snapshots of local moments. This is the same CNN whose filters are featured in Fig. 4(d). To obtain the snapshots in the DQMC approach, we note that for a particular auxiliary field, the expectation value of the local double occupancy reduces to its uncorrelated value: the product of the expectation values for spin-up and spin-down occupancies. Therefore, the local moment at site i can be written as $\langle \hat{n}_{i\uparrow} \rangle + \langle \hat{n}_{i\downarrow} \rangle - 2\langle \hat{n}_{i\uparrow} \rangle \langle \hat{n}_{i\downarrow} \rangle$.

Figure 8(a) shows the general improvement of the testing accuracy by increasing the number of filters. Figure 8(b) shows the 16 trained filters, which similarly to what we find for spin, point to only short-range fluctuations. One can find many redundancies. However, a few representative patterns [those features in Fig. 4(d)] emerge. Figure 8(c) shows the average network output when the network trained at $n = 0.82$ is tested on configurations across a range of densities. We also test the network on fake snapshots. A nonmonotonic behavior emerges in both cases. Subtracting the average network output for the real and fake snapshots at each density results in a curve that has a broad peak around $n = 0.85$ and showcases the extent of learned correlations between local moments in

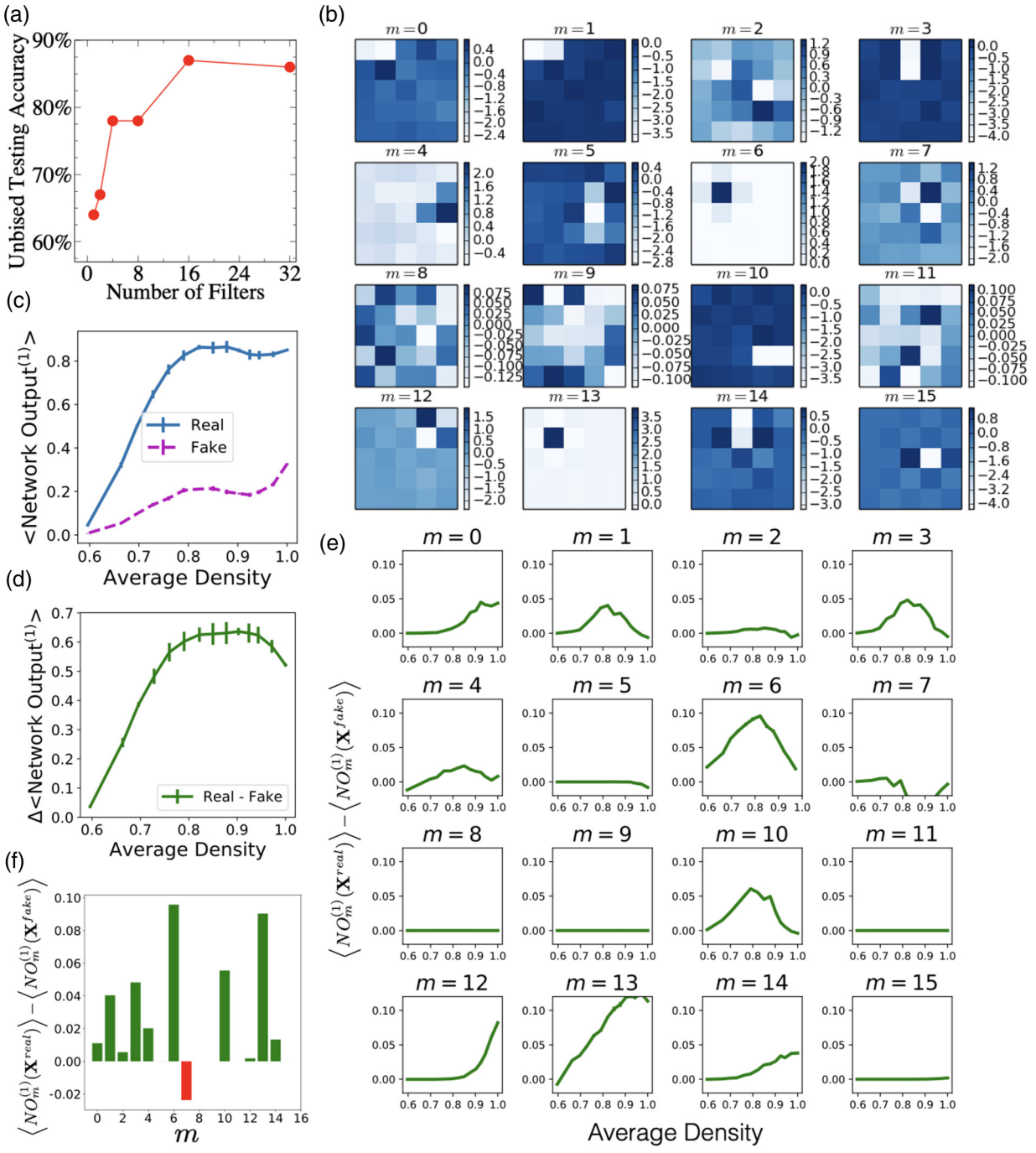

FIG. 8. Analysis of DQMC local moment snapshots using a CNN with 16 filters. (a) The general improvement in the prediction accuracy of the CNN by increasing the number of filters. (b) Trained filters when $n = 0.82$ and $T = 0.44t$. Note that pixel values or their range in one filter should not be compared with those in other filters since a filter-dependent bias is added to the result of the convolution before it is passed through the ReLU activation function (see Appendix B). The testing accuracy is around 87%. (c) Average network output when a real or fake DQMC snapshot is provided as input, as a function of average density for this CNN. Here, 1 means all the snapshots are classified as likely to be real $n = 0.82$ snapshots, 0 means all the snapshots are classified as likely to be fake $n = 0.82$ snapshots, and 0.5 means neither of the two options is preferred. (d) The difference between the two curves in (c), representing the average percentage the decisions made by the CNN have to do with factors other than the density itself. (e) Similar to (d) when the CNN has access to one filter at a time (see Appendix B). (f) Same as in (e), but at $n = 0.82$.

this CNN that have to do with factors other than the average density itself [Fig. 8(d)]. See Appendix B for more details.

To attribute certain features seen in trained filters in Fig. 8(b) to correlations unique to the NFL region, we have to rule out their dominance at other densities. Figure 8(e) shows the performance of individual filters over the same range of densities we used to study the network output. Figure 8(f) further highlights the values in Fig. 8(e) at $n = 0.82$. Based on

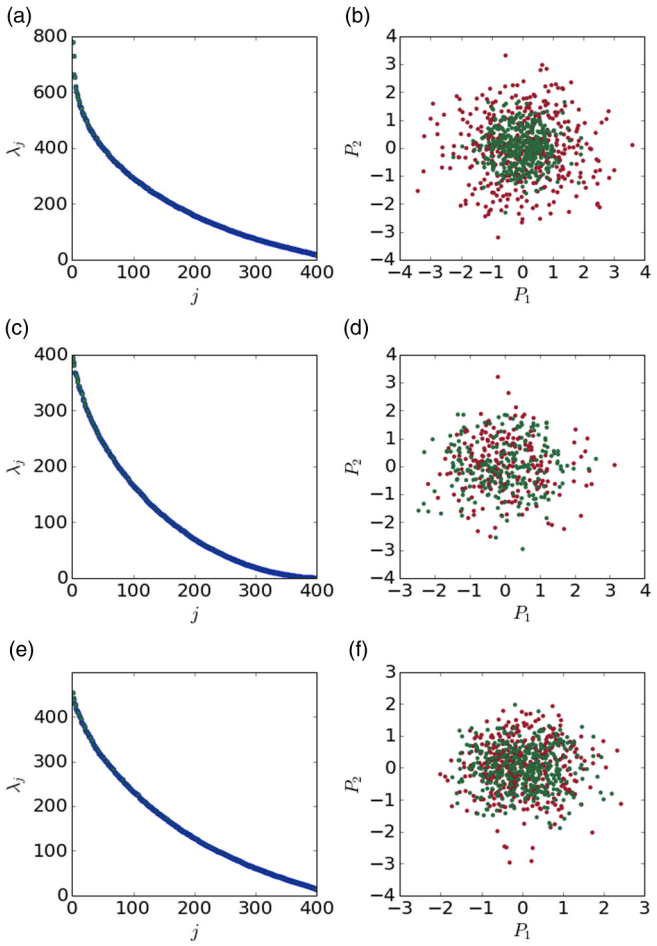

FIG. 9. PCA of the experimental snapshots. (a), (b) Eigenvalues of the covariance matrix of data for single-species snapshots at $n \sim 1$ and their projection to the space of the two largest principal components. The same 402 snapshots at each of the two extreme temperatures, used for Fig. 2, are also used here. Although the data points corresponding to the lower temperature [green (dark gray) circles in (b)] are slightly less spread out than their high-temperature counterparts [red (light gray) circles in (b)], there is no obvious clustering of data based on temperature. (c), (d) The same as (a) and (b), except for $n = 0.82$ and that 216 available snapshots at each of the two extreme temperatures are used. (e), (f) The same as (c) and (d), except that 400 real and 400 fake snapshots of local moments are used. No discernible patterns emerge in these projections either.

these results, filters that significantly contribute to the CNN's decision-making process and are unique to the NFL phase and, therefore, are the best candidates for offering insight into local moment fluctuations are $m = 1$, 3, 6, 7, and 10. The most frequently seen correlation seems to be the one between two neighboring empty sites. Filters $m = 6$ and 13 signal that the network also partly uses the information about the density gradient near local moments to make a decision.

APPENDIX D: DETERMINANT QUANTUM MONTE CARLO SNAPSHOTS

The implementation of DQMC used in the work proceeds via the exact rewriting of the interacting electron-electron problem as independent electrons moving in a space-imaginary time auxiliary field $h(\mathbf{r}, \tau)$. This reformulation involves first expressing the partition function $\mathscr{Z}$ for the original Hubbard Hamiltonian as a path integral, and then the use of a Hubbard-Stratonovich transformation to decouple the electrons. The original fermionic degrees of freedom are then traced out analytically, leaving an equivalent expression for $\mathscr{Z}$ as an integral over $h(\mathbf{r}, \tau)$. Detailed descriptions can be found in Refs. [26,43,44].

Here we focus on aspects of DQMC which have specific implications to the machine learning process. The most crucial is that, unlike world-line quantum Monte Carlo methods or cold-atom experiments, which directly sample the 0 or 1 occupation of sites $\mathbf{r}$ by the fermions, at any point (snapshot) in a DQMC simulation, the fermionic occupation is represented by a real number giving the probability of occupation of that site in the specific $h(\mathbf{r}, \tau)$ currently being sampled. As the temperature is lowered below t, sharper images containing pixels more closely resembling binary pixels in the experimental snapshots emerge. In fact, one can show that in the atomic limit, local expectation values approach step functions as $T \to 0$. This "smearing" of the occupation makes some aspects of machine learning via these snapshots more challenging. However, whether individual snapshots present (0,1) fermion occupations or not, the strong correlation physics (magnetism, superconductivity, strange metallicity) of the Hubbard model needs to be built up from many thousands of snapshots. It is the task of uncovering these many-body effects that is shared by the theoretical and experimental images investigated here with AI.

The presence of the fermion "sign problem" [45–47] in DQMC simulations can complicate the interpretation of the theory snapshots. To avoid this complication to the extent possible, at $n = 0.82$, we work at temperatures where the problem is not severe, where at least 90% of the auxiliary field configurations have a positive sign. We also treat the network output during the testing process the same way we treat the expectation value of a conventional observable, O, by computing $\langle O\mathscr{S}\rangle/\langle\mathscr{S}\rangle$ in place of $\langle O\rangle$, where $\mathscr{S}$ is the sign associated with the auxiliary field configuration resulting in a snapshot [11]. Like the network output, O is often a nonlinear function of the auxiliary field.

APPENDIX E: PRINCIPAL COMPONENT ANALYSIS

We have employed the linear principal component analysis (PCA) [38], an unsupervised learning algorithm, to perform dimensional reduction on the experimental data and examine whether any features emerge in the low-dimensional space, revealing any potential linear relation between pixels as an indicator for identifying low-temperature snapshots.

In the PCA, one forms a *matrix of data*, $\mathbf{x}$, by flattening the matrix of pixel values for each snapshot and placing them as an array of 400 binary numbers in each row (of $\mathbf{x}$). In the next step, the covariance matrix of data is composed as $\mathbf{x}^T \cdot \mathbf{x}$. Diagonalizing the 400×400 matrix, we obtain its eigenvalues, whose magnitudes are a measure for the variance of the data along the *principal axes*, determined by the corresponding eigenvectors. As demonstrated for the two-dimensional Ising model in a pioneering application in physics [48], a dominant

eigenvalue is indicative of a clear distinguishing pattern in the snapshots that can be represented as a linear combination of its pixels (their projection to the corresponding principal axis).

Figure 9 shows our results for both single-species ($n \sim 1$ and $n = 0.82$) and local moment ($n = 0.82$) snapshots. The PCA does not seem to be able to draw any particular distinction between low and high temperature or real and fake snapshots in any of the cases as there are no clear signs of clustering of data points based on temperature or whether or not they correspond to real snapshots. Figures 9(a) and 9(b) show, respectively, the eigenvalues of the covariance matrix of data for snapshots of single species near half filling and their projections to the space formed by the first two principal axes, representing the directions of the largest variance in data. A large gap between the first few eigenvalues and the rest of them would indicate that there are clear linear indicators for distinguishing sets of data [48], something we do not observe here. This is also inferred from the projection of data in Fig. 9(b), where, other than a slightly larger spread of data points at the lower temperature, no clear separation between the hot and cold data points is formed in the space of the first two principal components. Figures 9(c)–9(f) show a similar trend for both the single species and local moment snapshots at $n = 0.82$.

[1] X. Wen, *Quantum Field Theory of Many-Body Systems: From the Origin of Sound to an Origin of Light and Electrons* (Oxford University Press, Oxford, 2004).

[2] J. Orenstein and A. J. Millis, Science **288**, 468 (2000).

[3] C. Varma, Z. Nussinov, and W. van Saarloos, Phys. Rep. **361**, 267 (2002).

[4] S. Sachdev and D. Chowdhury, Prog. Theor. Exp. Phys. **2016**, 12C102 (2016).

[5] Y. Cao, V. Fatemi, S. Fang, K. Watanabe, T. Taniguchi, E. Kaxiras, and P. Jarillo-Herrero, Nature (London) **556**, 43 (2018).

[6] Y. Cao, V. Fatemi, A. Demir, S. Fang, S. L. Tomarken, J. Y. Luo, J. D. Sanchez-Yamagishi, K. Watanabe, T. Taniguchi, E. Kaxiras, R. C. Ashoori, and P. Jarillo-Herrero, Nature (London) **556**, 80 (2018).

[7] P. T. Brown, D. Mitra, E. Guardado-Sanchez, P. Schauß, S. S. Kondov, E. Khatami, T. Paiva, N. Trivedi, D. A. Huse, and W. S. Bakr, Science **357**, 1385 (2017).

[8] P. T. Brown, D. Mitra, E. Guardado-Sanchez, R. Nourafkan, A. Reymbaut, C.-D. Hébert, S. Bergeron, A.-M. S. Tremblay, J. Kokalj, D. A. Huse, P. Schauß, and W. S. Bakr, Science **363**, 379 (2019).

[9] G. Carleo and M. Troyer, Science **355**, 602 (2017).

[10] J. Carrasquilla and R. G. Melko, Nat. Phys. **13**, 431 (2017).

[11] K. Ch'ng, J. Carrasquilla, R. G. Melko, and E. Khatami, Phys. Rev. X **7**, 031038 (2017).

[12] D.-L. Deng, X. Li, and S. Das Sarma, Phys. Rev. B **96**, 195145 (2017).

[13] E. P. L. van Nieuwenburg, Y.-H. Liu, and S. D. Huber, Nat. Phys. **13**, 435 (2017).

[14] Y. Zhang and E.-A. Kim, Phys. Rev. Lett. **118**, 216401 (2017).

[15] P. Zhang, H. Shen, and H. Zhai, Phys. Rev. Lett. **120**, 066401 (2018).

[16] Y. Zhang, A. Mesaros, K. Fujita, S. D. Edkins, M. H. Hamidian, K. Ch'ng, H. Eisaki, S. Uchida, J. C. S. Davis, E. Khatami, and E.-A. Kim, Nature (London) **570**, 484 (2019).

[17] A. Bohrdt, C. S. Chiu, G. Ji, M. Xu, D. Greif, M. Greiner, E. Demler, F. Grusdt, and M. Knap, Nat. Phys. **15**, 921 (2019).

[18] B. S. Rem, N. Käming, M. Tarnowski, L. Asteria, N. Fläschner, C. Becker, K. Sengstock, and C. Weitenberg, Nat. Phys. **15**, 917 (2019).

[19] G. Torlai, B. Timar, E. P. L. van Nieuwenburg, H. Levine, A. Omran, A. Keesling, H. Bernien, M. Greiner, V. Vuletić, M. D. Lukin, R. G. Melko, and M. Endres, Phys. Rev. Lett. **123**, 230504 (2019).

[20] A. M. Samarakoon, K. Barros, Y. W. Li, M. Eisenbach, Q. Zhang, F. Ye, V. Sharma, Z. L. Dun, H. Zhou, S. A. Grigera, C. D. Batista, and D. A. Tennant, Nat. Commun. **11**, 892 (2020).

[21] C. Casert, K. Mills, T. Vieijra, J. Ryckebusch, and I. Tamblyn, arXiv:2002.07055.

[22] P. B. Wigley, P. J. Everitt, A. van den Hengel, J. W. Bastian, M. A. Sooriyabandara, G. D. McDonald, K. S. Hardman, C. D. Quinlivan, P. Manju, C. C. N. Kuhn, I. R. Petersen, A. N. Luiten, J. J. Hope, N. P. Robins, and M. R. Hush, Sci. Rep. **6**, 25890 (2016).

[23] L. R. B. Picard, M. J. Mark, F. Ferlaino, and R. van Bijnen, Meas. Sci. Technol. **31**, 025201 (2019).

[24] M. D. Zeiler and R. Fergus, in *Computer Vision – ECCV 2014*, edited by D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Lecture Notes in Computer Science Vol. 8689 (Springer, Cham, 2014), pp. 818–833.

[25] J. Tranquada, in *Handbook of High-Temperature Superconductivity*, edited by J. R. Schrieffer and J. Brooks (Springer, New York, 2007).

[26] R. Blankenbecler, D. J. Scalapino, and R. L. Sugar, Phys. Rev. D **24**, 2278 (1981).

[27] M. Rigol, T. Bryant, and R. R. P. Singh, Phys. Rev. Lett. **97**, 187202 (2006).

[28] E. Khatami and M. Rigol, Phys. Rev. A **84**, 053611 (2011).

[29] S. Sachdev, Phys. Status Solidi (b) **247**, 537 (2010).

[30] J. A. Hertz, Phys. Rev. B **14**, 1165 (1976).

[31] S. Sachdev and J. Ye, Phys. Rev. Lett. **69**, 2411 (1992).

[32] A. J. Millis, Phys. Rev. B **48**, 7183 (1993).

[33] G. R. Stewart, Rev. Mod. Phys. **73**, 797 (2001).

[34] H. v. Löhneysen, A. Rosch, M. Vojta, and P. Wölfle, Rev. Mod. Phys. **79**, 1015 (2007).

[35] A. L. Fitzpatrick, S. Kachru, J. Kaplan, and S. Raghu, Phys. Rev. B **88**, 125116 (2013).

[36] L. W. Cheuk, M. A. Nichols, K. R. Lawrence, M. Okan, H. Zhang, E. Khatami, N. Trivedi, T. Paiva, M. Rigol, and M. W. Zwierlein, Science **353**, 1260 (2016).

[37] J. Koepsell, S. Hirthe, D. Bourgund, P. Sompet, J. Vijayan, G. Salomon, C. Gross, and I. Bloch, Phys. Rev. Lett. **125**, 010403 (2020).

[38] I. Jolliffe, *Principal Component Analysis* (Wiley, New York, 2002).

[39] C. S. Chiu, G. Ji, A. Bohrdt, M. Xu, M. Knap, E. Demler, F. Grusdt, M. Greiner, and D. Greif, Science **365**, 251 (2019).

[40] J. Koepsell, J. Vijayan, P. Sompet, F. Grusdt, T. A. Hilker, E. Demler, G. Salomon, I. Bloch, and C. Gross, Nature (London) **572**, 358 (2019).

[41] M. Abadi *et al.*, TENSORFLOW: Large-scale machine learning on heterogeneous systems 2015, Software available from tensorflow.org

[42] C. Shorten and T. M. Khoshgoftaar, J. Big Data **6**, 60 (2019).

[43] S. R. White, D. J. Scalapino, R. L. Sugar, E. Y. Loh, J. E. Gubernatis, and R. T. Scalettar, Phys. Rev. B **40**, 506 (1989).

[44] F. Assaad and H. Evertz, in *Computational Many-Particle Physics*, edited by H. Fehske, R. Schneider, and A. Weiße (Springer, Berlin, 2008), pp. 277–356.

[45] E. Y. Loh, J. E. Gubernatis, R. T. Scalettar, S. R. White, D. J. Scalapino, and R. L. Sugar, Phys. Rev. B **41**, 9301 (1990).

[46] M. Troyer and U.-J. Wiese, Phys. Rev. Lett. **94**, 170201 (2005).

[47] V. I. Iglovikov, E. Khatami, and R. T. Scalettar, Phys. Rev. B **92**, 045110 (2015).

[48] L. Wang, Phys. Rev. B **94**, 195105 (2016).