

Single-Exposure Absorption Imaging of Ultracold Atoms Using Deep Learning

Gal Ness[✉], Anastasiya Vainbaum, Constantine Shkedrov[✉], Yanay Florshaim, and Yoav Sagi^{✉*}

Physics Department and Solid State Institute, Technion—Israel Institute of Technology, Haifa 32000, Israel



(Received 4 April 2020; accepted 20 May 2020; published 6 July 2020)

Absorption imaging is the most common probing technique in experiments with ultracold atoms. The standard procedure involves the division of two frames acquired at successive exposures, one with the atomic absorption signal and one without. A well-known problem is the presence of residual structured noise in the final image, due to small differences between the imaging light in the two exposures. Here, we solve this problem by performing absorption imaging with only a single exposure, where instead of a second exposure the reference frame is generated by an image-completion autoencoder neural network. The network is trained on images without an absorption signal, such that it can infer the noise overlaying the atomic signal based only on the information in the region encircling the signal. We demonstrate our approach on data captured with a quantum degenerate Fermi gas. The average residual noise in the resulting images is below that of the standard double-shot technique. Our method simplifies the experimental sequence, reduces the hardware requirements, and can improve the accuracy of extracted physical observables. The trained network and its generating scripts are available as an open-source repository (absDL.github.io).

DOI: [10.1103/PhysRevApplied.14.014011](https://doi.org/10.1103/PhysRevApplied.14.014011)

I. INTRODUCTION

Ultracold atomic gases are unique systems that allow the study of few- and many-body physics in a highly precise and tunable manner. The atomic ensembles are exquisitely isolated from the surroundings, as they are held in an ultrahigh-vacuum environment; therefore, probing them is almost always restricted to the analysis of their optical response. The most widely used probing technique is absorption imaging, where a collimated resonant laser beam is passed through the cloud and the shadow cast by the atoms is recorded by a digital camera [1]. The spatial atomic distribution is then extracted from the position-dependent absorption coefficient. The coherence length of the probe beam is typically much longer than the distances between optical interfaces in the experiment; hence unwanted reflections interfere and generate a characteristic patterns of stripes and Newton's rings in the recorded image. These patterns pose a problem in distinguishing between the signal and the nonuniform background.

The standard solution is to employ a double-exposure scheme: the first exposure is performed while the atoms are present, while the second *reference* exposure is performed shortly after and without the atoms. The exposure without atoms can be done either by waiting for the atoms to move out of the frame or by optically pumping them into a dark state. The line-of-sight integrated optical density (OD) image is formed by subtracting the logarithms

of the pixel counts in the two frames, with and without the atoms. However, due to acoustic noises and other dynamical processes, the noise patterns in the two images are typically not identical. This results in a residual-structured-noise pattern in the final image [Fig. 1(a)]. The lower signal-to-noise ratio imposed by the fringes is particularly problematic in low-OD images. Linear approaches for background completion have recently been suggested [2,3] but, as we show, they are sensitive to small changes in the noise pattern that evolve over time.

In this work, we tackle the noisy-background problem using machine learning, a term describing a set of algorithms that effectively perform specific tasks relying on patterns and inference. Among these, deep learning refers to a class of models that involves information propagation via multiple structures, enabling the translation of a given input to a certain prediction. The use of deep learning has become widespread in recent years for problems where an analytical mapping does not exist or when numerical solutions are intractable [4–8]. Image completion is an excellent example of such an application, particularly in a scenario where there are typical recurrent but varying patterns in the image. Machine-learning techniques have also been used for the optimization of the cooling sequences of ultracold atoms [9–12] and to execute related numerical calculations [13]. They have also been suggested [14] and demonstrated [15] to be useful for fluorescence detection of pinned atoms and ions.

Here, we report on an alternative approach for absorption imaging that uses a deep neural network (DNN)

*yoavsagi@technion.ac.il

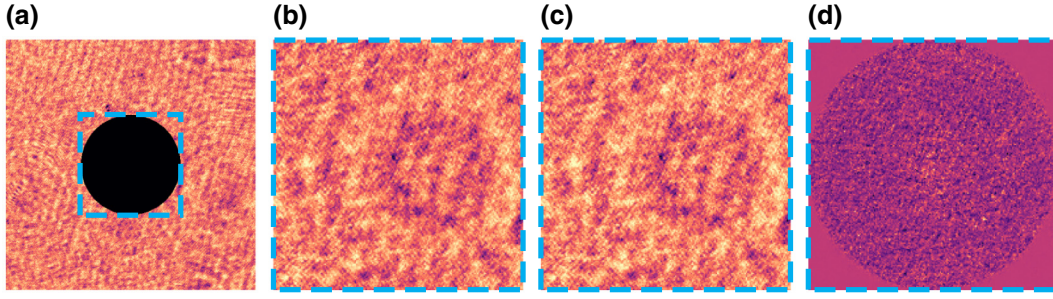


FIG. 1. The completion of a background frame by the neural net—an example of network evaluation of a typical image without atoms from the validation set. (a) The input log image, with its central part masked. The network task is to complete the image in the central cyan square. (b) The network prediction for the central square. (c) The original central part of the image (the “ground truth”). (d) The difference between the network prediction and the ground truth, multiplied by 5 to enhance the contrast. The residual OD root-mean-square error of this example is 0.061 for both the single-exposure and double-exposure techniques.

to generate an ideal reference frame from a single image that includes the atomic absorption signal. The reference image is constructed by masking out the part of the image containing the atomic shadow and using the network for image completion of the background. We demonstrate the method with data acquired with ultracold ^{40}K gas and show that the images captured by the single-exposure technique feature lower noise levels and therefore allow for a more accurate extraction of physical observables. In addition to the improvement in the data quality, our single-shot approach simplifies the experimental sequence and eases the hardware requirements from the camera. The DNN model successfully adapts to both short and long time variations and therefore it constitutes a robust solution.

II. METHODS

A. Experimental apparatus

The experiments are conducted with a quantum degenerate Fermi gas of ^{40}K atoms with an equal mixture of the two lowest energy states in the $F = 9/2$ manifold at a magnetic field of 185 G. Our experimental system and cooling procedure are the same as described in Refs. [16] and [17]. The frames without atoms are captured deliberately over a period of seven months to test the DNN in realistic conditions. We acquire data with atomic clouds at different conditions by modifying the evaporation cooling sequence. For training and validation of the DNN, we also acquire images without atoms. To this end, we set the initial position of the optical-transfer trap to about 2 cm away from their location at the magnetic trap; hence no atoms are shuttled to the position at which the images are recorded. In all cases, the first exposure is taken between 12 and 18 ms after the optical-dipole trap is turned off abruptly.

The images are taken with a laser tuned to the cycling transition $F = 9/2$, $m_F = -9/2 \rightarrow F' = 11/2$, $m_F = -11/2$ in the D_2 manifold, at a wavelength of approximately 766.7 nm. The laser line width is about

100 kHz, much narrower than the D_2 natural line width of approximately $2\pi \times 6$ MHz. The illumination is pulsed for 80 μs and recorded by a 14-bit CCD camera [18]. The reference frame (for the conventional absorption imaging) is recorded with a second pulse given after 50 ms, when the atoms have already moved out of the camera field of view. We also capture “dark frames” without probe illumination at all, which serve as the zero references. The dark images do not have to be taken often, since they only account for any remaining light that is not due to the probe beam and for electronic noise in the camera. Prior to analyzing the two images in the conventional absorption-imaging technique, we correct for small differences that may exist between the intensity of the illumination in both exposures. These differences are typically of a few percent. The second exposure is taken only in order to compare our technique with the conventional method and is required neither for the application of the DNN nor for its training.

Two physical observables that are commonly used in ultracold atomic experiments are the temperature and the number of atoms. In the presented results, the number of atoms in the cloud and its temperature are controlled by changing the final trap depth in the optical evaporation. We extract the observables from the momentum distribution, which is measured after 15 ms of a ballistic expansion. To extract the observables, we fit the OD images with [1]

$$n(x, y) = n_{\text{peak}} \frac{\text{Li}_2 \left(-ze^{-(x-x_0)^2/2\sigma_x^2 - (y-y_0)^2/2\sigma_y^2} \right)}{\text{Li}_2(-z)} + B, \quad (1)$$

where $\text{Li}_n(z)$ denotes Jonquière’s polylogarithm function, $z = e^{\mu/k_B T}$ is the fugacity, and B accounts for any remaining constant background in the OD image. From the fugacity, we extract the relative temperature $T/T_F = [-6\text{Li}_3(-z)]^{-1/3}$, where $T_F = (6N)^{1/3} \hbar \bar{\omega}$ is the Fermi temperature and $\bar{\omega}$ is the geometrically averaged trapping frequency, which we measure and rescale according to the

trapping laser power. The number of atoms, N , is obtained by integrating over the fitted momentum distribution.

B. DNN architecture and training

DNNs establish a pipeline where the input (in our case, the information in the masked OD image) undergoes multiple convolutional transformations and dimensional variations. These transformations distill the features of the underlying spatial pattern and their result is the prediction of the DNN. The network is trained to optimally recover the structure of the illumination in the region where the atomic signal appears. The training phase is performed using images captured *without atoms* and therefore constitutes the “ground truth” for the noise-pattern reconstruction. At each optimization step, the prediction of the network is compared to the ground-truth values in the masked area and the weights of the model are varied to minimize the loss, i.e., the mean-square error (L_2 norm) between the ground truth and the prediction. At the end of the training, we obtain an optimized model ready for prediction (inference) on new images *with atoms*. The network produces an ideal reference regardless of whether atoms appeared in the original image or not, because the relevant region is masked out. Since the involved convolutions are relatively simple, the evaluation of the model for inference on new inputs is rapid and therefore the integration of a trained network into the infrastructure of another calculation is straightforward.

We subtract the dark frames from the raw images and then take the logarithm of their pixel values. The convolutional network is an autoencoder of a U-net architecture [19]. The input to the network is the OD image cropped to 476×476 pixels around the position of the atoms, from which we mask out the central circle with a diameter of 190 pixels [20], which may include an absorption signal if atoms are present. This mask diameter is larger by at least a factor of 2 relative to the size of the typical atomic cloud, to ensure that there is no absorption signal in the region used by the DNN to predict the background. For training, we use a generator to riffle through the stored TIFF images, apply the mask on the input, and feed the DNN input with eight-frame batches. To evaluate the DNN on an atomic frame, we store the model inference as a binary file and subtract the input frame to obtain the atomic OD. By minimizing the loss over the square circumscribing the masked region [the dashed cyan square in Fig. 1(a)], we ensure continuity at the corners, where the background is unmasked. Effectively, $1 - \pi/4$ of the loss is dedicated to image duplication rather than completion, in order to eliminate any offset between the input and output frames, which might be translated into an error in the number of atoms.

The feed-forward network consists of about 20×10^6 parameters arranged in 27 layers. These parameters are optimized by running over about 30×10^3 frames captured

without atoms, with an additional approximately 7×10^3 images for loss validation, comparing the network output to the original central part of each image, and minimizing the mean-square-error loss function. We implement the DNN training over the KERAS PYTHON package [21] on the TensorFlow backend [22] and use the ADAM optimizer [23] and Glorot initialization [24] for the parameter optimization, applying 99% batch normalization [25]. For this application, labeling of the input frames is unnecessary, as the network output is compared directly against its input before masking. The only prior knowledge is the absence of atoms in the peripheral region and, only for the training set, also in the central area. Notably, generative adversarial networks [26], which have been found very successful in natural-scene image-competition tasks, might be destructive for this study case, as there is a given unique ground truth.

III. RESULTS

A. DNN performance on the validation set

First, we examine the residual noise in inferences on the validation set, which was not used for training and does not include atomic signal. The convergence of the model is depicted in Fig. 2, where we present the decay of the residual loss during the training process for both the training (purple) and validation (black) data sets. The

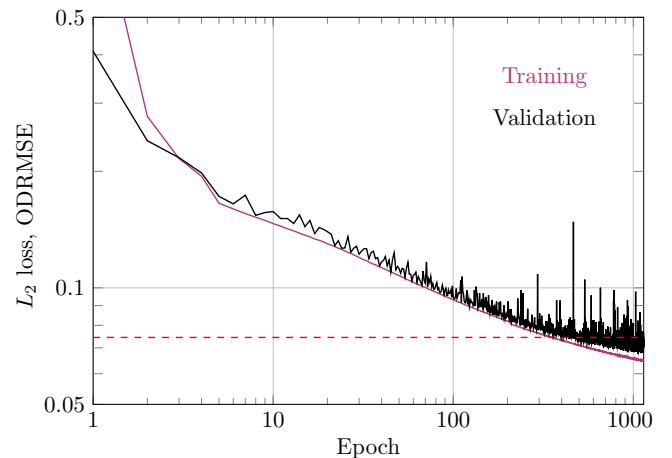


FIG. 2. The minimization of the residual error along the DNN training. The optical-density root-mean-square error between the model prediction and the ground truth as a function of the number of training iterations (epochs). Lower values mean better performance. The purple curve represents the residual loss of the training set, which is minimized in the optimization process. The black curve is the residual error on the validation set, which was not used for training. The dashed red line designates the mean residual noise in the standard double-shot scheme, multiplied by $\pi/4$ to correctly compare with the residual noise of the DNN prediction in the central circle [see Fig. 1(d)].

decay in both data sets on a log-log scale features a sub-power-law scaling. It exceeds the reference level, set by the average double-shot residual noise (the dashed red line), after approximately 100 training epochs, which mainly points to a reliable extraction of the bias, but noise features still exist. In principle, the training should continue as long as the validation loss decreases. In practice, the loss decay slows dramatically after a few hundred epochs and therefore we cease the training after 1133 epochs. An example for image completion without atoms is displayed in Fig. 1, with the DNN input [Fig. 1(a)] and the corresponding prediction of the network [Fig. 1(b)], which closely resembles the original data [Fig. 1(c)]. Notably, there are no significant spatial correlations in the difference between the desired and the predicted frame [Fig. 1(d)].

The lowest residual error is an 0.0681 optical-depth root-mean-square error (ODRMSE), for the whole validation data set captured intermittently over a period of seven months. As most of the residual error results from the inner circle of the square output image [see Fig. 1(d)], a fair comparison for the loss is against $\pi/4$ of the averaged-double-shot error, indicated by a dashed red line in Fig. 2. This reference value is 0.0745 (ODRMSE), 9.4% higher than the minimal validation loss obtained during the first 1139 epochs.

In Fig. 3, we compare the histograms of the residual loss on the validation set using the DNN-based single-shot technique (upper panel), the conventional double-exposure technique (middle panel), and background completion using the principal-component-analysis (PCA) technique (lower panel) [2]. The histogram for the DNN technique exhibits a single narrow peak, while for the two other approaches it is markedly wider and multistructured. To illustrate the source of this behavior, we color the histograms based on the elapsed time when taking the corresponding data set, relative to the first set. We find that the double-peak structure of the conventional double-shot technique is correlated with the time variations, probably due to slow drifts in the probe light intensity. An exacerbation of this variation is observable in the PCA results. We substantiate that it directly emerges from the variations in the set from which the PCA basis is taken by repeating the PCA analysis but with the basis taken over 600 frames, all from the first day of image acquisition. In this case, we find that the PCA approach yields excellent results for same-day frames, at 0.08(2) ODRMSE. Nonetheless, its performance deteriorates dramatically with long-term drifts—we find distinct date-dependent peaks in the histogram (not shown in the figure) and for the 206–210-day data sets the error distribution lies at 0.42(1) ODRMSE. We can conclude that in order for the PCA approach to maintain adequate performance, recurrent data-set accumulation and analysis is needed, almost on a daily basis. In contrast, the DNN technique is robust and insensitive to these variations. It derives its robustness from the variance

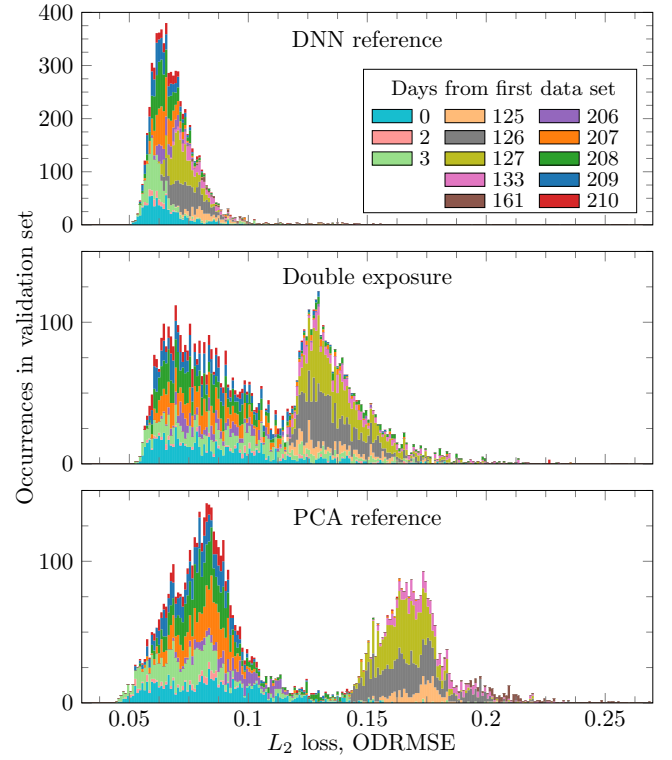


FIG. 3. The residual-error distribution of the difference images in the validation set. The upper histogram indicates the optical-density root-mean-square error of the DNN single-exposure technique following 1133 training epochs. The middle histogram represents the residual error in the standard double-exposure technique, after correction for probe intensity fluctuations. The lower histogram depicts the residual error of PCA-based reference generation images [2]. The set of PCA vectors is extracted from the 300 significant components out of 600 random images taken from the DNN training set. Different colors distinguish the validation set constituent frames by date, counting from the first partial set.

in the substantially broader data set, which is tractable due to the sequential training of the network.

The results on the validation set show that the DNN single-exposure approach achieves lower residual noise levels and deals better with variations in the imaging conditions when compared to the conventional double-shot scheme or linear algorithms. The residual noise of the DNN technique can, in principle, be further reduced by additional training. To assess the usefulness of the time invested in such prolonged training, one should take into account whether it has a measurable effect on physical observables, as we describe in the next section.

B. Single-shot imaging evaluation

In this section, we present single-shot absorption images of a quantum degenerate fermionic potassium gas at different conditions. A typical analysis of a low-OD image

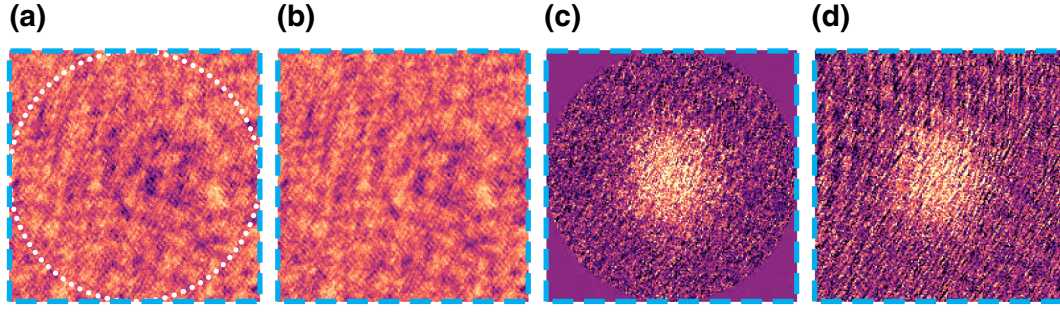


FIG. 4. The reconstruction of a single-shot image with atoms, exemplified with a cloud of about 30×10^3 atoms. (a) The central square of a single-shot log image, including the masked area (the dotted white circle). (b) The network prediction. (c) The difference between prediction and input, multiplied by 5, resulting in a fringes-free single-shot absorption image. (d) The result of the conventional two-exposures technique in the same experiment, where the second exposure is taken 50 ms after the first one (also multiplied by 5).

following a ballistic expansion from an approximately 80-nK-deep trap is shown in Fig. 4. In Fig. 4(a), we present the inner square part of the input log image. In this example, there are approximately 30×10^3 atoms; hence, to the naked eye, the atomic signal is hardly discernible from the background. When it is subtracted from the network prediction in Fig. 4(b), a clean OD image is obtained [Fig. 4(c)]. As a comparison, Fig. 4(d) shows the conventional absorption image obtained from two exposures in the same experiment. Evidently, the single-shot approach eliminates the remaining fringe pattern and yields an OD image that is better overall. More examples for different trap depths are presented in the upper panel of Fig. 5 and show the same behavior regardless of the atomic conditions.

C. Effect on physical observables

In Fig. 6, we plot the number of atoms and the temperature for different trap depths as extracted by the single-shot

(purple diamonds) and the two-exposures (black circles) techniques. Notice that the DNN-based technique does not introduce any systematic error in extraction of these important observables. The error bars represent the 1σ shot-to-shot variation in the experimental conditions combined with the fitting extraction error. Since both of these terms are of a similar magnitude, it is hard to observe the improvement in the single-exposure technique. To emphasize this improvement, we present in the insets only the fitting extraction relative error averaged over the ten experimental realizations in each trap depth. We find that the extraction uncertainty of both observables is smaller by about 17% using the single-exposure technique.

IV. SUMMARY AND OUTLOOK

We demonstrate a single-shot absorption imaging based on a deep convolutional-network background completion. We show that this approach can accurately reconstruct

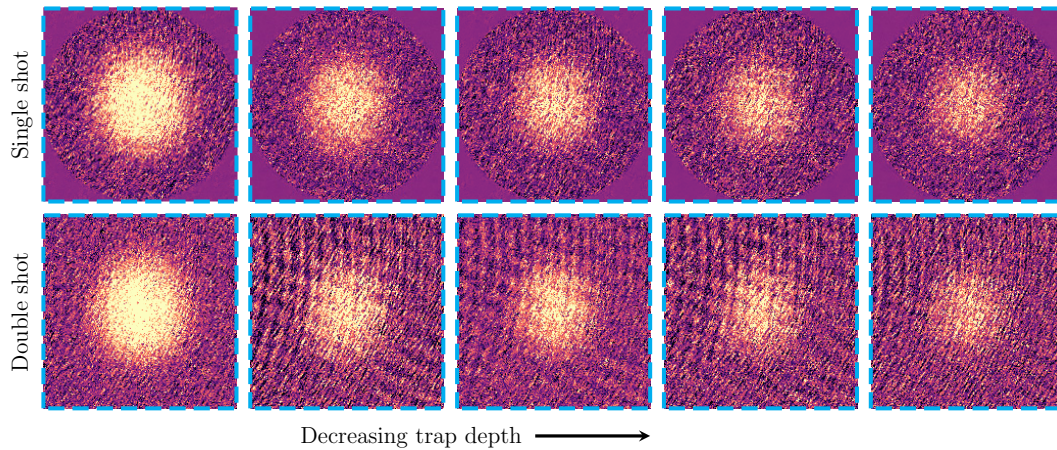


FIG. 5. Additional examples of inferences of the neural network on images with atoms (upper panel) for different conditions of the atomic cloud. The lower panel presents the correlative results using the standard double-shot technique. The numbers of atoms in these examples are, from left to right, $99(11) \times 10^3$, $55(7) \times 10^3$, $45(9) \times 10^3$, $37(9) \times 10^3$, and $28(7) \times 10^3$ and they were released, respectively, from 190-, 117-, 89-, 76-, and 57-nK-deep traps. All examples are displayed using the same color scale as in Fig. 4(d).

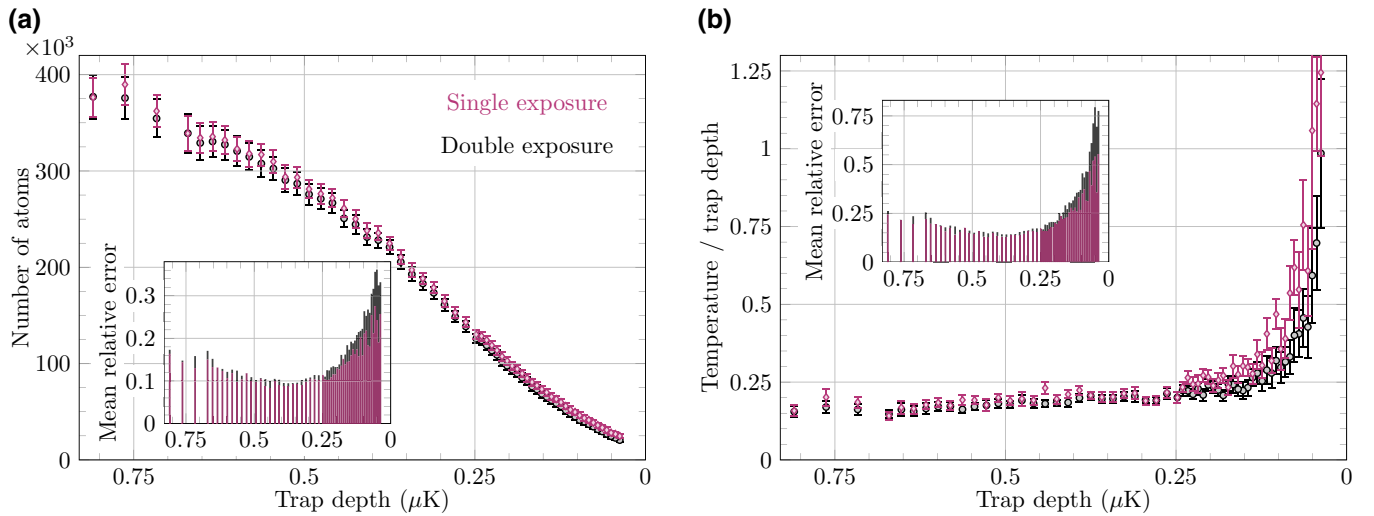


FIG. 6. Characterization of the resulting images: (a) the number of atoms and (b) the temperature extracted by fitting a Fermi-Dirac distribution to the data. The conditions of the atomic clouds are controlled by the final trap depth in the optical evaporation. The black circles mark the results of the conventional double-exposure technique, while the purple diamonds mark the results with the single-shot DNN approach. The error bars combine extraction uncertainty with shot-to-shot variation over ten experimental realizations. The insets show the average fit extraction error. The single-exposure technique achieves a better accuracy in both observables.

atomic density profiles and yield smaller errors on the extracted physical quantities compared to the standard double-exposure technique. The single-shot imaging eliminates the need for fast cameras and facilitates multiframed acquisitions. The corresponding simplification directly enables simpler and cleaner designs for cold atomic systems. We also demonstrate the ability of the DNN to adapt to variations in the working conditions that develop through time.

Our network can be improved in several aspects. First, further optimization of the DNN hyperparameters can be undertaken, e.g., improved loss functions, different network and layer depths, and utilization of additional optimizers. Second, the masked area can be enlarged to achieve even better robustness. Also, by training the network over random patches in the uncropped OD image, the position dependency of the results can be further reduced. Another interesting direction is the implementation of an *online learning* scheme, where images are routinely added to the data set and the model is continuously updated between inferences.

DATA AVAILABILITY

The trained network and its generating scripts are publicly available as an open-source PYTHON software package [27] to facilitate their deployment by other experimental groups. Using the provided repository, single-shot imaging can be realized on any imaging apparatus, following local parameter training. The ideal training-data-set size depends on the noise parameters and may vary between different experimental systems. We find that approximately

30×10^3 frames are sufficient to produce good results in our case.

ACKNOWLEDGMENTS

This research was supported by the Israel Science Foundation (ISF) Grant No. 1779/19 and by the United States–Israel Binational Science Foundation (BSF), Jerusalem, Israel, Grant No. 2018264. The GeForce TITAN V used for the local network training was donated by the Nvidia Corporation. Remote training power was granted by the Google Cloud Platform research credits program. G.N. would like to thank Amit Oved for inspiring discussions.

-
- [1] M. W. Z. W. Ketterle, Making, probing and understanding ultracold Fermi gases, *La Rivista del Nuovo Cimento* **1**, 247 (2008).
 - [2] L. Niu, X. Guo, Y. Zhan, X. Chen, W. M. Liu, and X. Zhou, Optimized fringe removal algorithm for absorption images, *Appl. Phys. Lett.* **113**, 144103 (2018).
 - [3] B. Song, C. He, Z. Ren, E. Zhao, J. Lee, and G.-B. Jo, Effective statistical fringe removal algorithm for high-sensitivity imaging of ultracold atoms, arXiv:2002.10053 (2020).
 - [4] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, *Nature* **521**, 436 (2015).
 - [5] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Quantum machine learning, *Nature* **549**, 195 (2017).
 - [6] P. Mehta, M. Bukov, C.-H. Wang, A. G. Day, C. Richardson, C. K. Fisher, and D. J. Schwab, A high-bias,

- low-variance introduction to machine learning for physicists, *Phys. Rep.* **810**, 1 (2019).
- [7] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Machine learning and the physical sciences, *Rev. Mod. Phys.* **91**, 045002 (2019).
- [8] M. Raghu and E. Schmidt, A survey of deep learning for scientific discovery, arXiv:2003.11755 (2020).
- [9] 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, Fast machine-learning online optimization of ultra-cold-atom experiments, *Sci. Rep.* **6**, 25890 (2016).
- [10] A. D. Tranter, H. J. Slatyer, M. R. Hush, A. C. Leung, J. L. Everett, K. V. Paul, P. Vernaz-Gris, P. K. Lam, B. C. Buchler, and G. T. Campbell, Multiparameter optimisation of a magneto-optical trap using deep learning, *Nat. Commun.* **9**, 4360 (2018).
- [11] I. Nakamura, A. Kanemura, T. Nakaso, R. Yamamoto, and T. Fukuhara, Non-standard trajectories found by machine learning for evaporative cooling of ^{87}Rb atoms, *Opt. Express* **27**, 20435 (2019).
- [12] A. J. Barker, H. Style, K. Luksch, S. Sunami, D. Garrick, F. Hill, C. J. Foot, and E. Bentine, Applying machine learning optimization methods to the production of a quantum gas, *Mach. Learn.: Sci. Technol.* **1**, 015007 (2020).
- [13] S. Pilati and P. Pieri, Supervised machine learning of ultracold atoms with speckle disorder, *Sci. Rep.* **9**, 5613 (2019).
- [14] L. R. B. Picard, M. J. Mark, F. Ferlaino, and R. van Bijnen, Deep learning-assisted classification of site-resolved quantum gas microscope images, *Meas. Sci. Technol.* **31**, 025201 (2019).
- [15] Z.-H. Ding, J.-M. Cui, Y.-F. Huang, C.-F. Li, T. Tu, and G.-C. Guo, Fast High-Fidelity Readout of a Single Trapped-Ion Qubit via Machine-Learning Methods, *Phys. Rev. Appl.* **12**, 014038 (2019).
- [16] C. Shkedrov, Y. Florshaim, G. Ness, A. Gandman, and Y. Sagi, High-Sensitivity rf Spectroscopy of a Strongly Interacting Fermi Gas, *Phys. Rev. Lett.* **121**, 093402 (2018).
- [17] G. Ness, C. Shkedrov, Y. Florshaim, and Y. Sagi, Realistic shortcuts to adiabaticity in optical transfer, *New J. Phys.* **20**, 095002 (2018).
- [18] pco.pixelfly USB (ICX285AL CCD).
- [19] O. Ronneberger, P. Fischer, and T. Brox, in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer, Munich, 2015), p. 234.
- [20] The imaging system translates the mask diameter of 190 pixels into $626\text{ }\mu\text{m}$ at the atoms plane.
- [21] F. Chollet *et al.*, KERAS, <https://keras.io> (2015).
- [22] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin *et al.*, TensorFlow: Large-scale machine learning on heterogeneous distributed systems, arXiv:1603.04467 (2016).
- [23] D. P. Kingma and J. Ba, in *proceedings of the 3rd International Conference on Learning Representations* (an Diego, 2015).
- [24] X. Glorot and Y. Bengio, in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* (2010), p. 249.
- [25] S. Ioffe and C. Szegedy, in *Proceedings of the 32nd International Conference on Machine Learning, Proceedings of Machine Learning Research* (PMLR, Lille, 2015), Vol. 37, p. 448.
- [26] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, in *Advances in Neural Information Processing Systems 27* (Curran Associates, Inc., Montreal, 2014), p. 2672.
- [27] The absDL repository is available at <http://absDL.github.io>.