








**PAPER**

Femtosecond pulse parameter estimation from photoelectron momenta using machine learning

OPEN ACCESS**RECEIVED**
2 May 2023**REVISED**
19 July 2023**ACCEPTED FOR PUBLICATION**
8 August 2023**PUBLISHED**
24 August 2023Original Content from
this work may be used
under the terms of the
[Creative Commons
Attribution 4.0 licence](#).Any further distribution
of this work must
maintain attribution to
the author(s) and the title
of the work, journal
citation and DOI.**Tomasz Szoldra**^{1,2} , **Marcelo F Ciappina**^{3,4,5} , **Nicholas Werby**^{6,7} , **Philip H Bucksbaum**^{6,7,8} ,
Maciej Lewenstein^{9,10} , **Jakub Zakrzewski**^{2,11}  and **Andrew S Maxwell**^{12,*} ¹ Doctoral School of Exact and Natural Sciences, Jagiellonian University, Łojasiewicza 11, PL-30-348 Kraków, Poland² Instytut Fizyki Teoretycznej, Wydział Fizyki, Astronomii i Informatyki Stosowanej, Uniwersytet Jagielloński, Łojasiewicza 11, PL-30-348 Kraków, Poland³ Department of Physics, Guangdong Technion—Israel Institute of Technology, 241 Daxue Road, Shantou, Guangdong 515063, People's Republic of China⁴ Technion—Israel Institute of Technology, Haifa 32000, Israel⁵ Guangdong Provincial Key Laboratory of Materials and Technologies for Energy Conversion, Guangdong Technion—Israel Institute of Technology, 241 Daxue Road, Shantou, Guangdong 515063, People's Republic of China⁶ Stanford PULSE Institute, SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, United States of America⁷ Department of Physics, Stanford University, Stanford, CA 94305, United States of America⁸ Department of Applied Physics, Stanford University, Stanford, CA 94305, United States of America⁹ ICFO-Institut de Ciències Fotoniques, The Barcelona Institute of Science and Technology, Av. Carl Friedrich Gauss 3, 08860 Castelldefels (Barcelona), Spain¹⁰ ICREA, Pg. Lluís Companys 23, 08010 Barcelona, Spain¹¹ Mark Kac Complex Systems Research Center, Jagiellonian University, Łojasiewicza 11, PL-30-348 Kraków, Poland¹² Department of Physics and Astronomy, Aarhus University, DK-8000 Aarhus C, Denmark

* Author to whom any correspondence should be addressed.

E-mail: andrew.maxwell@phys.au.dk**Keywords:** convolutional neural networks, femtosecond pulse characterization, machine learning, strong-field ionizationSupplementary material for this article is available [online](#)**Abstract**

Deep learning models have provided huge interpretation power for image-like data. Specifically, convolutional neural networks (CNNs) have demonstrated incredible acuity for tasks such as feature extraction or parameter estimation. Here we test CNNs on strong-field ionization photoelectron spectra, training on theoretical data sets to ‘invert’ experimental data. Pulse characterization is used as a ‘testing ground’, specifically we retrieve the laser intensity, where ‘traditional’ measurements typically lead to 20% uncertainty. We report on crucial data augmentation techniques required to successfully train on theoretical data and return consistent results from experiments, including accounting for detector saturation. The same procedure can be repeated to apply CNNs in a range of scenarios for strong-field ionization. Using a predictive uncertainty estimation, reliable laser intensity uncertainties of a few percent can be extracted, which are consistently lower than those given by traditional techniques. Using interpretability methods can reveal parts of the distribution that are most sensitive to laser intensity, which can be directly associated with holographic interferences. The CNNs employed provide an accurate and convenient ways to extract parameters, and represent a novel interpretational tool for strong-field ionization spectra.

1. Introduction

Machine learning (ML) has been transformative for science over the last two decades, providing a huge range of new analytical tools. This has affected nearly every avenue of research, with major use across the physical sciences, particularly in particle physics [1], astrophysics [2], and condensed matter physics [3, 4]. Convolutional neural networks (CNNs) have enabled major leaps in computer vision and language processing [5, 6]. This makes CNNs well-suited for pattern recognition and parameter estimation in

scientific data. For example, CNNs have been used for determining crystal symmetries in electron diffraction [7], and estimating parameters related to gravitational lensing [8]. However, in the field of strong field physics and attoscience the high interpretability power of CNNs has not been fully explored.

Strong-field physics and attoscience exploit recent advances for producing intense and short laser pulses to image and control matter over attosecond (10^{-18} s) timescales [9–12]. These capabilities have led to a wide range of atomic and molecular imaging procedures, e.g. high-order harmonic spectroscopy [13], laser-induced electron diffraction [14], photoelectron holography [15, 16], attosecond streaking [17, 18], and reconstruction of attosecond harmonic beating by interference of two-photon transitions [19, 20]. However, despite ever more accuracy from experiment and theory, due to the nonlinear nature of the interactions, interpretation of the data is often very challenging. This provides an opportunity for ML methods to be used to extract parameters and physical trends from experimental data sets.

A growing number of studies have begun to use ML techniques for strong-field physics. For example, studies using neural networks to classify semi-classical trajectories [21], deep learning to predict in spectra of high-harmonic generation [22], and optimization of ‘quantum pathways’ in enhanced ionization of diatomic molecules [23]. In terms of parameter estimation, in a recent theoretical study, CNNs were used to extract internuclear distances and laser intensities, using data generated solving the time-dependent Schrödinger equation (TDSE) [24, 25]. The power of CNNs to extract useful information from experimental images has big implications for strong-field physics and attoscience. Most studies, however, focus on a proof of principle, using only theoretical data. Notable exceptions are, reference [26], where CNNs were used to extract molecular structure parameters from experimental laser-induced electron diffraction images, and reference [27], where deep neural networks were applied to streaking traces for parameter extraction and prediction of uncertainties. Unfortunately, the analytical power of ML-assisted imaging is limited if the laser pulse parameters cannot be accurately measured.

The characterization of laser pulses in strong-field and attosecond physics has posed a persistent problem. The high intensity of the laser pulse means that a direct measurement (see e.g. [28, 29]) of the intensity leads to significant uncertainties, in the range of 10%–20% for the strong-field regime [29, 30]. An alternative approach is to use the high sensitivity of the non-linear phenomena in question, to estimate the laser pulse parameters, known as an *in situ* measurement. Using this approach, laser intensity uncertainties as low as 1% have been reached, by matching experimental results to TDSE theory, under highly controlled experimental conditions for atomic hydrogen [29, 30]. Despite this success, such low uncertainties are not common, and would be more difficult to achieve routinely in standard experimental conditions. Simply fitting photoelectron spectra is less effective and more advanced methods, using all the available information in photoelectron momentum distributions (PMDs), are called for. Recent results, using quantum metrology tools, suggest the uncertainty from *in situ* measurements could be significantly reduced by exploring quantum interferences present in the PMDs in strong-field ionization [31]. There is a variety of *in situ* methods for determining laser parameters that are implemented by hand, whose performances vary across parameter regimes. The most consistent and powerful method is to use the whole PMD, matching it to that obtained with accurate theoretical methods. As such, ML schemes, and in particular CNNs, are an ideal tool for *in situ* extraction of laser parameters from experimental data.

The task of using ML for laser pulse characterization has been addressed in the relativistic regime where, a theoretical study used CNNs to predict laser intensities by using proton dynamics [32]. In free electron lasers (FELs), neural networks have been used to accurately reconstruct pulses, by training a model on a small set of fully diagnosed pulses [33]. CNNs were also used to characterize the FEL pulses by training on simulated data [34]. CNNs have also been used in all optical measurement schemes, employing interferometric cross-correlation between pulses to characterize one of the pulses [35]. Recently, ML tools have been used for pulse characterization for strong-field ionization [36] using purely theoretical data. Here, the autocorrelation function of the ionization yield from two identical pulse was used to extract the pulse duration, spectral width and relative carrier-envelope phase (CEP), but the method was insensitive to laser intensity.

In this work, we investigate the power of a CNN as an analysis tool for strong-field physics. We use laser pulse characterization in strong-field ionization as a testing ground, retrieving parameters from PMDs. We train the CNNs on a TDSE model, and test this on large experimental datasets, focusing on retrieving laser intensities, over a larger parameter regime than has previously been considered, for an argon target. The CNNs trained may be used on any experimental data within the parameter range, without special requirements, and the CNN models are available online for testing. Important modifications to theoretical training data, are presented, that ensure the CNN models are insensitive to common experimental imperfections. We also include predictive uncertainty estimation, which goes beyond previous methods to extract uncertainty in strong-field studies. We produce so-called ‘explainability’ figures that are able to highlight regions of the PMD that contribute the most to a laser intensity prediction, and connect these to the physical interference process that are most sensitive to changes in laser intensity. As such, CNNs represent

the easiest way to extract laser parameters in strong-field ionization, making a key step towards producing a general tool for parameters estimation, while also providing more interpretational power. Atomic units are used unless otherwise stated.

2. Datasets

2.1. QProp

The main workhorse for generating our theoretical data set is the single-active-electron (SAE) TDSE solver QPROP. The latest version of QPROP [37], implements a fast and accurate method for the calculation of PMDs. QPROP is a velocity gauge three-dimensional TDSE solver in the dipole approximation that allows studies within the SAE approximation using model pseudopotentials¹³. For our SAE model of the Ar atom, we have employed the model potential of [38], which has the form

$$V(r) = -\frac{Z+f(r)}{r} \quad (1)$$

with

$$f(r) = a_1 e^{-a_2 r} + a_3 r e^{-a_4 r} + a_5 e^{-a_6 r}, \quad (2)$$

where $Z = 1$. For argon the coefficients a_i are $a_1 = 16.039$, $a_2 = 2.007$, $a_3 = -25.543$, $a_4 = 4.525$, $a_5 = 0.961$ and $a_6 = 0.443$ [38], which gives the correct ionization potential of $I_p = 0.579$ a.u. In this computation, we considered angular momenta up to $l = 55$. Total data set consists of 15712 PMDs with ponderomotive potentials ranging from $U_p = 0.0075$ to $U_p = 0.95$, laser cycles number $N = 2$ up to $N = 61$. The number of CEP values depends on the pulse length, i.e. for under $N = 13$ cycles we cover an interval of length π with 10 values of CEP, and for longer cycles, we have 5 values of CEP spanning a shifted interval of the same length.

2.2. Strong field approximation (SFA)

An extensive review of the SFA can be found in [39]. Here, we use the transition amplitude for direct ATI from an initial bound state $|\psi_0\rangle$ to a final Volkov state with drift momentum \mathbf{p} given by [31, 40–44]

$$M(\mathbf{p}) = -i \lim_{t \rightarrow \infty} e^{iS(\mathbf{p}, t)} \int_{-\infty}^t dt' d(\mathbf{p}, t') e^{iS(\mathbf{p}, t')}, \quad (3)$$

where $d(\mathbf{p}, t') = \langle \mathbf{p} + \mathbf{A}(t') | \mathbf{r} \cdot \mathbf{E}(t') | \Psi_0 \rangle$ is the dipole prefactor, which in this study we will neglect as we retain only terms correct to exponential accuracy. The action is given by

$$S(\mathbf{p}, t) = I_p t + \frac{1}{2} \int dt' (\mathbf{p} + \mathbf{A}(t'))^2. \quad (4)$$

Here, I_p is the ionization potential of our target. We employ the saddle point approximation, seeking the stationary action for the integration variable t' , $2I_p + (\mathbf{p} + \mathbf{A}(t_s))^2 = 0$. Now the probability distribution can be computed from equation (3) as

$$M(\mathbf{p}) = \sum_s c(\mathbf{p}, t_s, t) d(\mathbf{p}, t_s) e^{iS(\mathbf{p}, t_s)}, \quad (5)$$

where the prefactor $c(\mathbf{p}, t_s, t)$, derived from the application of the saddle point approximation, includes the t' independent phase from equation (3), is given by

$$c(\mathbf{p}, t_s, t) = -i e^{iS(\mathbf{p}, t)} \sqrt{\frac{2\pi i}{\partial^2 S(\mathbf{p}, t_s) / \partial t_s^2}}. \quad (6)$$

In both QPROP and SFA calculations we use the vector potential:

$$\mathbf{A}(t) = 2\sqrt{U_p} \sin^2\left(\frac{\omega t}{2N}\right) \cos(\omega t + \phi), \quad (7)$$

where N is the number of laser cycles, while U_p is the ponderomotive energy or quiver energy of a free electron in the laser field, which is proportional to the peak laser intensity $I_0 = 2U_p c \epsilon_0 \omega^2$, where ϵ_0 and c are

¹³ There is also an implementation of many-electron systems via the solution of the time-dependent Kohn–Sham equations.

the vacuum permittivity and the speed of light, respectively. The angular frequency is given by ω and the CEP is given by ϕ . We also perform focal (FA) and CEP averaging (CA), to account for variations of the intensity across the focal volume and CEP fluctuations between laser pulses, respectively. CEP-averaged Q_{PROP} datasets contain 3071 PMDs. Details are given in the supplemental material.

2.3. Experimental methods

Our experimental data sets consist of PMDs of argon gas photoionized by intense, linearly polarized, 800 nm laser pulses generated by a 1 kHz commercial Ti:sapphire laser system. These laser pulses are focused in ultra-high vacuum and intersect a pulsed beam of argon gas delivered by an Even–Lavie valve [45]. Photoelectrons are collected in a velocity map imaging spectrometer, and impact a microchannel plate and a fast phosphor detector system. A camera records the intensified phosphor, and on-the-fly peak finding is applied to the live camera feed to extract individual electron impacts.

Six experimental data sets with 125 PMDs in total are analyzed here. Four of these sets, labeled E1–E4, were collected using an experimental schema in which the ionizing pulse energy was the variable parameter. The pulse energy is controlled using a motorized rotation mount which manipulates a half-wave plate to rotate the pulse polarization. It then passes through a polarizing beamsplitter cube, which transmits only the component of the laser pulse polarized parallel to the optical table. Another data set, ‘E HS’, was collected for intensities over a larger range and include the highest values of intensity. The final set, labeled E (single), is a single-intensity PMD, which is included as it has the highest signal-to-noise ratio. Each dataset was collected on the timescale of approximately two days, and contains in total $O(10^9)$ electron counts, distributed between their intensity slices.

The data sets presented have all been Abel inverted using the standard technique of polar onion peeling [46, 47] to extract their cylindrical momentum cross sections. The ponderomotive potential, U_p , of the laser fields generating each slice of each data set is computed in a two-step process. First, it is roughly calculated by examining the direct electrons, which form a disk of radius $2U_p$. Then, that rough value is refined by comparing nodes found along the ‘spider-leg’ holographic feature to those predicted by the Coulomb quantum orbit SFA, see [16, 48–50], at nearby intensities. This procedure is described in more detail in [47]. This has an error of approximately $\pm 10\%$.

3. Deep neural network approach

Our task is the following: given an experimental PMD X , find a physical parameter y , for example, the intensity of the laser pulse, that has been used to produce such PMD. We assume an underlying theoretical model of the strong-field process that allows us to generate PMDs expected at given physical parameters. The richness of the features present in the PMDs makes the task very challenging. Moreover, various imperfections are present in the experimental data, complicating the comparison with theory further, cf histograms in figure 1(g) showing a dramatic quantitative difference between the PMD values in both datasets.

To address this demanding work, we then use a deep learning approach. By generating a dataset of PMDs labeled by many different physical parameters, we reformulate the problem as a standard supervised regression task. In section 3.1, we describe our choice of deep neural network architecture.

3.1. CNNs and transfer learning

CNNs [51] are designed to work with data incorporating spatial correlations such as pictures. The main building block of CNN is a convolution matrix with trainable parameters that slides over the input image and produces its filtered representation. The composition of many such filters gives a feature map of the image, ready to be used for further processing in the final fully connected part of the network.

Although deep CNNs achieve state-of-the-art in image recognition [52–54], sometimes this requires the use of prohibitively large training datasets and computing power. However, one can train deep models through the transfer learning paradigm [55, 56]: one takes a pre-trained deep model and fine-tunes it on a smaller dataset. The first few layers of the network extract general features such as edges, so often retraining only the last few layers of the model may be sufficient to achieve good performance on the new dataset. In this work we benchmark four pre-trained architectures called VGG16 [57], Xception [58], EfficientNetB7 [59], and EfficientNetV2L [60] that achieved state-of-the-art accuracy in classification tasks on the Imagenet dataset [53]. Models are ordered from least to most sophisticated. We do not describe the substantial innovations introduced in each of them but concentrate on the comparison of their general performance. Because the models were originally designed for classification and not regression, we remove the last classification layer, and replace it with a fully connected layer with a linear (identity) activation function, see

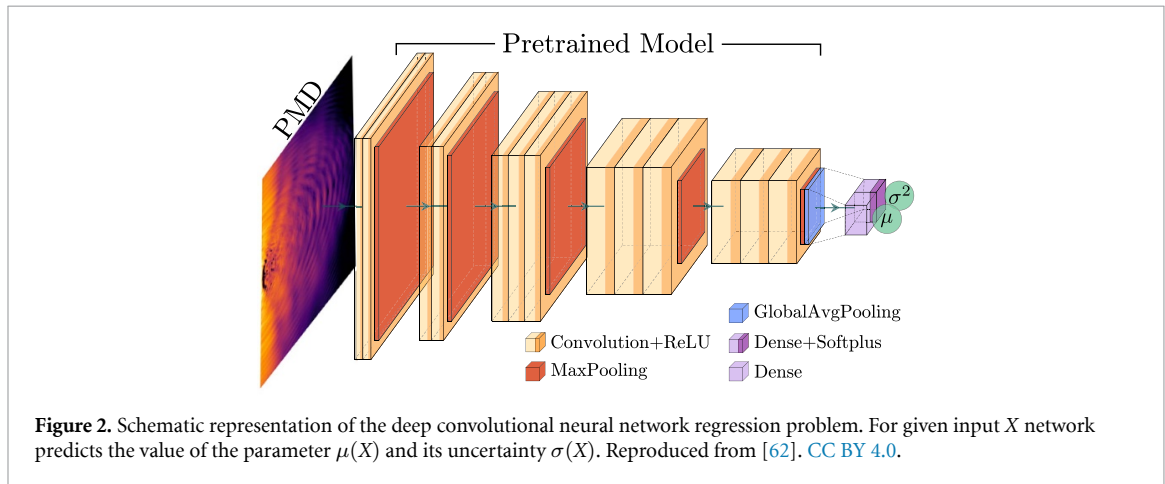
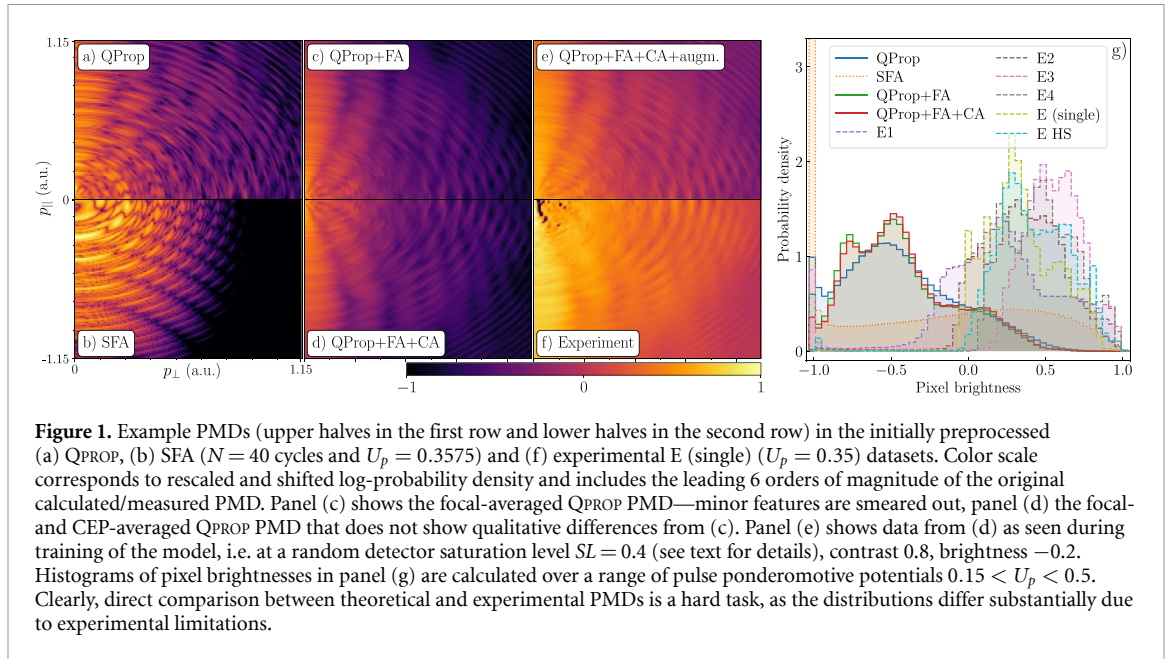


figure 2. Models are implemented using Keras library [61] and are available in our code repository <https://github.com/tszoldra/attoDNN>.

3.2. Predictive uncertainty estimation

Along with the predicted label, we aim to provide an estimate of the model uncertainty for a given input [63, 64]. We slightly modify the model and the loss function [65, 66]: instead of predicting a single value of the label $y_{\text{pred}}(X)$, we assume the label comes from a normal distribution $p(y_{\text{true}}|X) = \mathcal{N}(\mu(X), \sigma(X))$ and the model predicts its parameters $\mu(X), \sigma^2(X)$ for a given input X . $\sigma^2(X)$ is the output of an additional fully connected layer with a Softplus(x) = $\ln(\exp(x) + 1)$ activation function, see figure 2. The loss function to minimize is the negative log-likelihood,

$$\text{NLL} = -\langle \ln p(y_{\text{true}}|X) \rangle = \frac{1}{2} \left\langle \ln \sigma^2(X) + \frac{(y_{\text{true}} - \mu(X))^2}{\sigma^2(X)} \right\rangle, \quad (8)$$

where $\langle \cdot \rangle$ denotes the mean over the training dataset. To further improve the reliability of the predictions and predictive uncertainties, we combine M models trained on random subsets of the training dataset into a deep ensemble [66, 67]. Further assuming that the ensemble prediction is a Gaussian, the ensemble mean and variance read

$$\mu_*(X) = \frac{1}{M} \sum_{m=1}^M \mu_m(X), \quad (9)$$

$$\sigma_*^2(X) = \frac{1}{M} \sum_{m=1}^M (\sigma_m^2(X) + \mu_m^2(X)) - \mu_*^2(X), \quad (10)$$

where $\mu_m(X)$, $\sigma_m(X)$ are the mean and variance predicted by the m th model in the ensemble [66]. This goes beyond studies such as [27], where an ensemble of predictions is used to produce the uncertainty.

3.3. Data preprocessing and augmentation

In this section, we describe a few technical steps that were necessary to preprocess the PMDs to form a viable image input for the CNNs. As a first step, we identified a common range of momenta accessible in all datasets to form a rectangle with $p_{\perp} \in [0.001 \text{ a.u.}, 1.15 \text{ a.u.}]$, $p_{\parallel} \in [-1.15 \text{ a.u.}, 1.15 \text{ a.u.}]$, where the discretization is given by the resolution of the experimental dataset E1 $\Delta p_{\perp} = \Delta p_{\parallel} = 0.0049 \text{ a.u.}$. Then, we interpolated the theoretical data on the same 2-dimensional momentum grid.

The natural scale for features contained in the PMD is logarithmic and the signal has to be transformed accordingly, see e.g. [68] for CNNs applied to diffraction images with similar properties. Thus, we take the logarithm of the PMDs, rescale and apply an offset to the pixel values so that in the end they fill the interval $[-1, 1]$. Thus, the pixel value 1 represents the largest peak probability density in each PMD and -1 is the value smaller by a factor of 10^{-6} . Each pixel is clipped according to $X_{ij} \rightarrow \max(-1, X_{ij})$. Six orders of magnitude are selected based on heuristics that will capture all features in the experimental data, at the same time not misleading the network by showing extremely precise, low values in the theoretical data. All images are then resized to 224 by 224 pixels with 3 (repeated) color channels to match the standard of the Imagenet dataset expected by the pre-trained models.

We split the theoretical datasets, see section 2, into training (80%), validation (10%), and test (10%) subsets. For each model in the ensemble, the training/validation split is different and random, while the test dataset is constructed once by a random selection from the full dataset.

During the training phase, we perform image augmentation [69]. Each input image is randomly reflected in the up-down and left-right axes, and its contrast and brightness are randomly set from the interval $(0.1, 1.0)$ and $(-1, 1)$, respectively, using the built in Keras [61] functionalities, see figure 1(e). The final image is clipped to a fixed range $[-1, 1]$. While the testing data does not include reflected images, during training we apply reflections to help the network learn the same ‘shapes’ in four different settings with the aim of reducing overfitting. On the other hand, the testing data has inherently varying levels of contrast and background signal (‘brightness’) and we deliberately make the network insensitive to them.

In our efforts to make the models useful for experimentalists, we encountered an obstacle: while the models performed well on theoretical data (see next section 4), they failed for experiments. We fixed it by adding a single extra step in the augmentation pipeline, motivated by the histogram in figure 1(g), that shows significant differences between distributions of pixel values in theory and experiment, especially for the brightest pixels. This suggests there was some uncontrolled detector saturation effect in the experiment, at a level not necessarily fixed between experiments. Thus, prior to all augmentations described earlier, we simulate a random detector saturation level. For each sample PMD, we draw a random ‘saturation’ value x from the interval $[SL, 1]$ and transform each pixel according to $X_{ij} \rightarrow \min(X_{i,j}, x) + 1 - x$. The lower bound on the random saturation value SL is a parameter that has to be found by checking the performance of the model on part of the experimental data. By making the saturation level random, we increase the training difficulty, but at the same time make the model insensitive to detector saturation that occurs in a real experiment¹⁴.

3.4. Training

We train an ensemble of size $M = 5$ of pretrained models VGG16, Xception, EfficientNetB7, EfficientNetV2L on four datasets: QProp, QProp+CA, QProp+FA, QProp+CA+FA, for a set of random detector saturation level lower bounds $SL \in \{-1, -0.5, 0, 0.5, 1\}$, yielding 400 models in total. We choose the Adam optimizer [70] and a batch size of 32. During the first 50 training epochs, the base model has fixed weights and only the added, two randomly initialized dense layers each with 1-dimensional output $\mu(X)$ and $\sigma(X)$ are being updated at a learning rate of 10^{-3} . This roughly sets up the last layer while not destroying pretrained filters. In the next 150 iterations the model is fine-tuned: all weights are updated at a learning rate 10^{-4} , which decreases by a factor of 0.5 every 50 iterations. We stop the training if the loss on the evaluation

¹⁴ Since pixel brightness values are limited to the range $[-1, 1]$, $SL = -1$ corresponds to a fully random saturation level, whereas $SL = 1$ to no detector saturation effect present at all (case of a ‘perfect detector’).

Table 1. Loss metrics for the training dataset QProp+FA+CA. For each CNN architecture, we show only saturation level SL used in the image augmentation for which NLL on dataset E4 is the lowest. On E4, the best accuracy is obtained for EfficientNetV2L which simultaneously gives acceptable errors for other experimental data. Error metrics are similar on test/train subsets of QProp+FA+CA (first two rows) which is a signature of good model generalization. The lowest values of MAPE for each dataset are in bold.

Dataset	VGG16 @ $SL = 0.0$				Xception @ $SL = -0.5$				EfficientNetB7 @ $SL = 0.0$				EfficientNetV2L @ $SL = -0.5$			
	NLL	RMSE	$\langle \sigma_* \rangle$	MAPE	NLL	RMSE	$\langle \sigma_* \rangle$	MAPE	NLL	RMSE	$\langle \sigma_* \rangle$	MAPE	NLL	RMSE	$\langle \sigma_* \rangle$	MAPE
train	-4.2	0.0042	0.018	0.77	-2.0	0.11	0.12	14.	-4.4	0.0098	0.0097	1.3	-4.1	0.015	0.012	1.7
test	-4.1	0.0051	0.019	0.87	-1.9	0.12	0.12	14.	-4.4	0.011	0.010	1.4	-4.1	0.016	0.013	1.9
E1	-2.2	0.039	0.020	15.	-1.9	0.083	0.11	33.	1.7	0.038	0.010	15.	-0.99	0.040	0.016	17.
E2	-0.93	0.046	0.020	15.	-2.5	0.0099	0.083	3.1	5.9	0.032	0.0077	12.	-0.70	0.034	0.012	12.
E3	0.61	0.12	0.050	29.	-2.5	0.035	0.077	7.2	-0.89	0.037	0.021	8.5	-2.3	0.043	0.029	9.8
E4	-3.4	0.021	0.015	6.2	-2.2	0.061	0.089	22.	-3.4	0.014	0.0075	4.0	-4.1	0.0091	0.014	2.9
E (single)	-3.8	0.010	0.020	2.9	-2.4	0.0072	0.087	2.0	-3.1	0.015	0.0083	4.3	-4.3	0.0067	0.014	0.19
E HS	0.70	0.11	0.041	19.	-1.9	0.11	0.11	15.	-2.1	0.035	0.023	4.5	-3.1	0.025	0.025	4.5

dataset does not decrease for more than 100 epochs to save on computing time. All models can be trained in parallel¹⁵.

We were unable to train any models if they were initialized with random weights. It demonstrates the power of transfer learning from real-world images to physical experiments. The models are already capable of extracting basic shapes from images and need fine-tuning only.

4. Results

The quality of all 400 trained models is measured in terms of the NLL (see equation (8)), root mean squared error (RMSE) and the mean absolute percentage error (MAPE) achieved on the test datasets. For theoretical data sets, the ‘true’ intensity is known exactly, so RMSE and MAPE give the error on model prediction. For experimental data sets, the ‘true’ intensity carries 10% uncertainty, so the MAPE only needs to be within this bound. The mean predictive uncertainty $\sigma_*(X)$ is the models’ prediction of the uncertainty. This can be compared with the RMSE to assess if the models ‘know when they are wrong’. All these values are fully tabulated in the supplemental material. Here we give a general overview of these results and describe a method of model post-selection that allows us to find the best model candidates for experimental data presented in table 1.

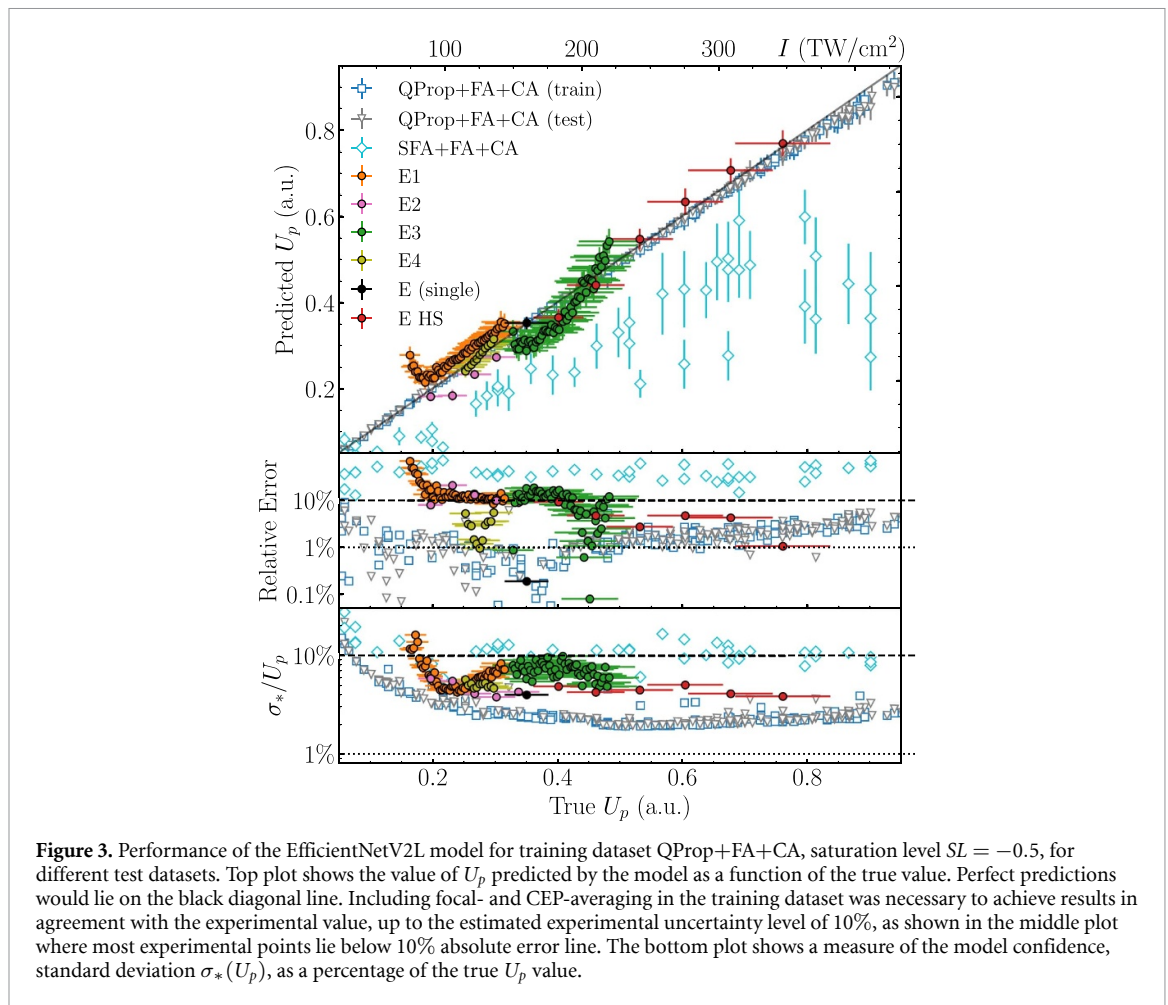
As a standard practice, no augmentation techniques are applied in the testing phase unless explicitly noted. While this can decrease performance of some models on theoretical testing data due to input distribution shifts, we concentrate more on the performance of experimental data which naturally includes imperfections. In the ‘perfect detector’ augmentation scenario, $SL = 1$, all models are trainable on all theoretical QPROP datasets with a testing MAPE $< 1\%$ (except for the VGG16 model and QProp+CA dataset where MAPE = 4.9%).

Testing on experimental data, we notice that including focal averaging and CEP-averaging in the training results in a smaller error. This is supported by a visual inspection of the PMDs in figures 1(a), (d) and (f), which unveils a greater resemblance between experimental and QProp+FA+CA rather than QProp data with more small-sized features. Thus, we restrict our further discussion to the QProp+FA+CA training dataset. Moreover, we observe that the quality of the prediction is always improved if detector saturation effects are included ($SL \leq 0.5$) than if they are not ($SL = 1$).

Aiming for the highest-quality models for experiments, we post-select the saturation level based on the best NLL on a single evaluation dataset E4. Obtained metrics are presented in table 1. We find that the fittest model is the most sophisticated EfficientNetV2L trained with a saturation level $SL = -0.5$, reaching a MAPE of 2.9% on E4, which is well below the experimental error. The predictive uncertainty is $\sigma_*(X) = 0.014$ a.u., corresponding to 6×10^{12} W cm⁻² in typical intensity units. This is close to the reported RMSE, signaling a good calibration of the model confidence. Almost all errors on other datasets for this model also fall within the error bars of the experimental label. In figure 3, we plot the predicted value of ponderomotive energy U_p , proportional to the laser intensity, as a function of the true U_p , for experimental and theoretical (QPROP and SFA) inputs. All predictions are presented with uncertainties computed using equation (10), given by the vertical error bars in figure 3 (upper panel) and as a percentage (lower panel).

Predictions on the test/train QProp+FA+CA dataset in figure 3 lie within the uncertainty estimate around the true value up to $U_p \approx 0.5$. For larger U_p the model slightly deviates, finishing with an error of

¹⁵ The training effectiveness may be improved by a recent training scheme [71] not applied here: in the first few iterations one should optimize the mean, while keeping the variance fixed.



around 7% at $U_p = 0.95$. We expect this drop in performance is associated with a limited range of momenta present in the training dataset. The $p = 2\sqrt{U_p}$ peak, used for labeling PMDs manually, is located at the border of accessible momenta at around $U_p \sim 0.66$. Thus, some other, possibly less expressed features in the PMD have to be used by the CNN.

Before we proceed to experimental data, we quickly cross-check the output of the model on the SFA dataset. Looking at the strong qualitative difference between sample images in figures 1(a) and (b), not to mention the dramatic dissimilarity of the histograms of both datasets in figure 1(g), it is surprising that the model finds any structure at all in the SFA+FA+CA data, i.e. there is a significant positive correlation coefficient between the true value and the prediction. However, the uncertainty of SFA+FA+CA is strongly underestimated, particularly for the largest values. This is a warning that on the strongly out-of-distribution data¹⁶, the model may fail to predict the true value and report a relatively high confidence.

Testing the same model on experimental datasets we notice that errors generally stay equal or lower than the experimental uncertainty of 10% (middle panel). Note that the model was selected based on its performance on evaluation dataset E4, yet its predictions are consistent for other datasets. Overall, there is good agreement with the vast majority of intensity predictions carrying an uncertainty below that of attainable through traditional methods, getting as low as 4% in a number of cases. In particular, the high-statistic dataset E HS gives a very good agreement over a wide range of intensities. It was crucial to consider focal averaging here, otherwise, the ‘true’ labels deviated from the predicted value, with an absolute percentage error up to $\sim 50\%$ for large intensities. Notably, a large relative error is observed for a few points of the dataset E1 at low intensities. We believe this is primarily caused by a low contrast in the input image due to a lower number of electron counts in this setting, see supplemental material for an example image, since this deviation can be manually removed by increasing the contrast of the input images.

¹⁶ Out-of-distribution data describes data that is far from the kind of data that was used for training, where the model is likely to give unpredictable results.

5. Explanations

5.1. Methods

The high accuracy of the models presented above makes them a readily useful tool for parameter estimation. On the other hand, due to a rather complex flow of information in image regression, the understanding of why a certain output is produced, is lacking, i.e. we deal with a ‘black box’. This hinders progress in the development of new, more accurate models, and, more importantly, does not give any insight into the underlying physical reasoning. These issues are addressed in the following section using so-called explainability techniques that quantify how certain features of the input contribute to the output, see recent review [72, 73] for a more general survey.

The most popular explainability techniques were designed for classifiers, and special care needs to be taken when using them for regression, see [74]. Here we adopt the simple yet powerful strategy of [75], where explanations of deep regression models are obtained directly using methods for classification.

Three basic approaches to explainable regression have been developed to date and a variety of algorithms can be found in each category [74]. The most straightforward are removal-based explanations [76], measuring the importance of a given subset of input features by hiding it from the model. Because there is an exponential number of subsets to check, usually these methods are limited to at most 15–20 features before the analysis of images becomes infeasible. Another set of methods are gradient-based explanations that rely on the computation of the gradient of the input in a single forward/backward propagation of the signal. They are built on the intuition that if some region of the image is important for the prediction, a small change in this region will noticeably change the output. Finally, propagation-based explanations aim to leverage the neural network structure to produce the feature attribution map. In particular, the layer-wise relevance propagation (LRP) algorithm [77] assigns a relevance score R_i to each neuron i based on the activations of the neurons in the next layer. Scores in a single layer are conserved, i.e. sum up to the final prediction. The relevance scores are calculated layer by layer in the backward pass from the output, until the input is reached.

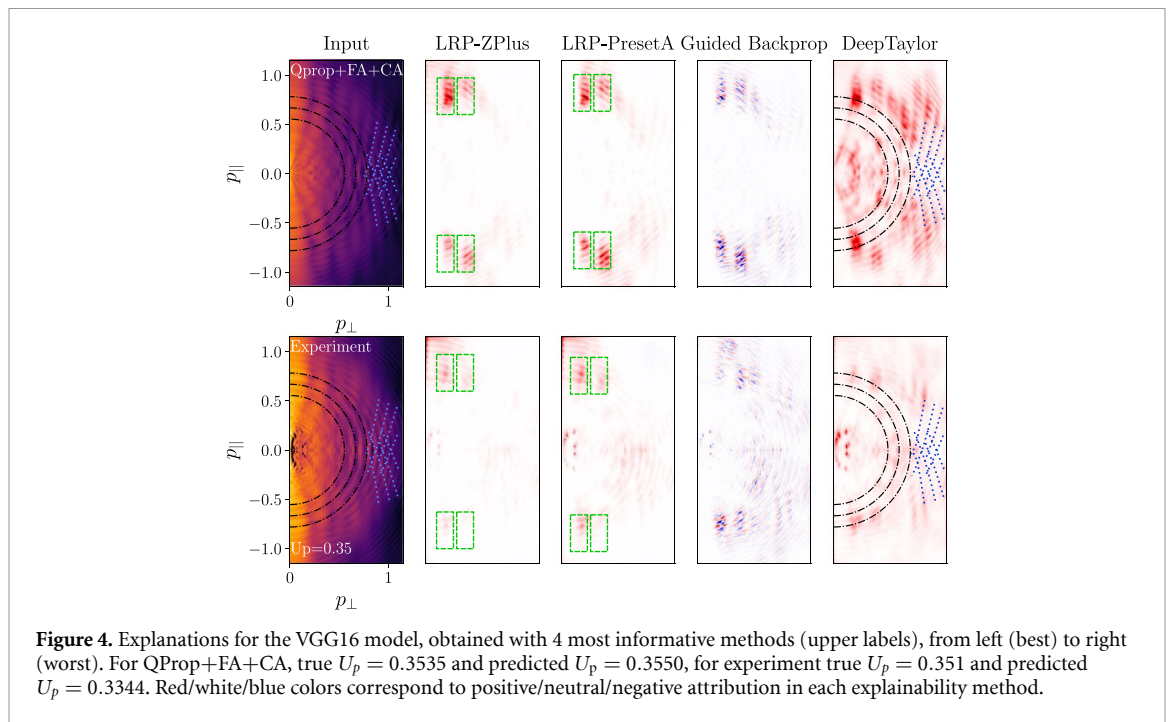
We apply 10 different explainability algorithms available in the iNNvestigate toolbox [78]. Our analysis is restricted to the VGG16 model at $SL = 0.0$ instead of the ‘best performing’ EfficientNetV2L at $SL = -0.5$, presented in figure 3, due to the large size of the latter, making most methods intractable due to memory requirements, and a ‘swish’ activation function which is not compatible with many algorithms. Out of all tested algorithms, for presentation, we post-select the four most relevant ones by scoring them following [75, 79]. Each explanation image is divided into 8 by 8 regions and perturbed one region at a time, from most to least important, according to the output of a given explanation algorithm. If the regions marked by the algorithm are indeed relevant for predictions, the accuracy of the model drops faster than when the perturbations are applied in random order. In our case, 9 out of 10 tested methods perform better than a random one, and the four presented in figure 4 are noticeably better than the rest, see supplemental material for further details.

5.2. What the model learns

In figure 4, we used four explainability techniques: two variants of the LRP [77], guided backpropagation [80], and DeepTaylor [81] to highlight regions on the PMD that have the most effect on the predicted value. This can also be interpreted as highlighting the features that are the most sensitive to changes in the ponderomotive energy, and thus it is a unique way to extract physical meaning. The highlighted regions can be identified as known interference features that are used in photoelectron holography. The guided backpropagation technique picks out two regions. The first region is at the end of the ‘legs’ of the so-called spider-like structure [15, 82], see green dashed rectangles in figure 4. This is formed via the interference between pairs of electronic wavepackets that are forward scattered/deflected off by the residual ion and have a differing degree of interaction with the core.

The explainability diagrams specifically pick out modulations along the spider legs above the direct boundary ($p = 2\sqrt{U_p}$). These modulations were discussed in [47], where their sensitivity to the ponderomotive energy was already exploited for determining the laser intensity. Crucially, these modulations have been shown previously [48] to be described approximately by circles with their centers determined by U_p . Thus, explaining the sensitivity to the ponderomotive energy.

Another region highlighted, with high perpendicular momentum, is the so-called carpet-like [83, 84] or spiral-like [85] structure, see blue dotted lines in figure 4. Around $p_{||} = 0$, the interference maxima can be described by $I_p + U_p + E = 2n\omega$ (with $n \in \mathbb{Z}$), which clearly encodes the ponderomotive energy. Away from $p_{||} = 0$, in the DeepTaylor method, we see the strongest contribution. Here, the above equation will not hold exactly, but the fringes will be dependent on the interplay of two rescattered wavepackets, that undergo different rescattering angles, which will be heavily dependent on the laser intensity/ponderomotive energy, as it determines the tunnel exit and initial scattering velocity.



Across many regions, and particularly in the DeepTaylor method, we can see the above-threshold ionization rings, which are ring-shaped interferences due to nearly identical wavepackets released at an integer number of laser cycles apart, see black dot-dashed circles in figure 4. The maxima may be described by a similar equation to the carpet-like structure $I_p + U_p + E = n\omega$, which is clearly sensitive to the ponderomotive energy. In previous work, this has been shown to provide important sensitivity for determining laser intensity [31].

6. Conclusions

We have proposed and tested deep learning models, powerful enough to detect objects in real world images, as a versatile analysis tool for strong-field ionization processes, adopting deep CNNs as our main workhorse. We have found they are capable of extracting physical parameters of interest—the laser peak intensity—and connect the parameter back to a specific feature, such as the interferences observed in the PMDs. We have overcome key difficulties using theoretically-trained CNNs with experimental data, which paves the way for CNNs to be used in a variety of settings for strong-field ionization, particularly when characterizing the laser field or target. The CNNs have been tested via pulse characterization, in particular determining the laser field intensity, on a large experimental data set, consistently yielding lower uncertainties than are achievable in traditional methods. For the prediction of uncertainty, we have used a reliable predictive uncertainty approach, that provides additional evaluation of the experimental conditions. We have also verified that other laser field parameters such as the pulse length could easily be extracted.

Deep CNNs can utilize information present in the picture to its full extent, while a human expert would typically be limited to using only a small subset of physical effects that are most sensitive to a change in the parameters. As such, we have developed a novel tool that can be applied to strong-field ionization photoelectron momentum spectra without any special requirements from the experimental data. We achieved this generalizability from our CNN by using state-of-the-art pretrained models. Training such models from scratch, to work well with experimental data, was not possible with our data set. Instead using the concept of transfer learning, we found the pretrained networks to work exceptionally well. This approach is easily repeatable as these models are freely available and it significantly reduces size of the required training data. The idea of transfer learning may be exploited in future work in order to expand the parameter range of the model, without having to use as much training data. Here, the exceptional capability and large capacity of the CNN models that we use could help make them more generalizable.

We show-cased the ‘explainability’ capability of CNNs, which highlighted the most relevant features in the PMDs, which could be associated directly with holographic interferences that display considerable sensitivity to changes in the ponderomotive energy. Thus, directly connecting the CNNs predictions to

fundamental physical processes. Explainability methods take away some of the black box nature of CNNs and can be used to highlight new physics, providing a powerful way to connect experiment and theory.

This study paves the way for further exploitation of CNNs to analyze strong-field ionization data, yielding new physical insights or confirming existing understanding. The recipe we developed, training the neural networks to be insensitive to various types of imperfections through the use of data augmentation techniques, made them ideal candidates for robust parameter extraction. We emphasize that the same procedure can be repeated and used to develop a range of analysis tools, which, for instance, could be highly useful for extracting atomic targets and/or pulse shapes, or further developing photoelectron holographic imaging, where inversion of experimental data is very difficult. Using these techniques, universal extraction of physical parameters is possible from existing and future experimental data, regardless of whether all details of the physical processes at play are fully understood.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: <https://chaos.if.uj.edu.pl/ZOA/index.php?which=opendata&dataset=/2303.13940>.

Acknowledgments

We gratefully acknowledge Poland's high-performance computing infrastructure PLGrid (HPC Centers: ACK Cyfronet AGH) for providing computer facilities and support within computational Grant No. PLG/2022/015830. T S is supported by National Science Centre (Poland) under Grant 2019/35/B/ST2/00034. This research was also funded by National Science Centre (Poland) under the OPUS call within the WEAVE Programme 2021/43/I/ST3/01142 (J Z) A partial support by the Strategic Programme Excellence Initiative at Jagiellonian University as well as that within the QuantEra II Programme that has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 101017733 DYNAMITE (M L and J Z). M F C acknowledges financial support from the Guangdong Province Science and Technology Major Project (Future functional materials under extreme conditions, No. 2021B0301030005) and the Guangdong Natural Science Foundation (General Program Project No. 2023A1515010871). A S M acknowledges funding support from: The European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement, SSFI No. 887153.

N W and P H B are supported by the U.S. Department of Energy, Office of Science, Basic Energy Sciences (BES), Chemical Sciences, Geosciences, and Biosciences Division, AMOS Program.

M L acknowledges support from: ERC AdG NOQIA; Ministerio de Ciencia y Innovation Agencia Estatal de Investigaciones (PGC2018-097027-B-I00/10.13039/501100011033, CEX2019-000910-S/10.13039/501100011033, Plan National FIDEUA PID2019-106901GB-I00, FPI, QUANTERA MAQS PCI2019-111828-2, QUANTERA DYNAMITE PCI2022-132919, Proyectos de I+D+I 'Retos Colaboración' QUSPIN RTC2019-007196-7); MICIIN with funding from European Union NextGeneration EU (PRTR-C17.I1) and by Generalitat de Catalunya; Fundació Cellex; Fundació Mir-Puig; Generalitat de Catalunya (European Social Fund FEDER and CERCA program, AGAUR Grant No. 2021 SGR 01452, QuantumCAT U16-011424, co-funded by ERDF Operational Program of Catalonia 2014-2020); Barcelona Supercomputing Center MareNostrum (FI-2022-1-0042); EU Horizon 2020 FET-OPEN OPTologic (Grant No. 899794); EU Horizon Europe Program (Grant Agreement 101080086—NeQST), National Science Centre, Poland (Symfonia Grant No. 2016/20/W/ST4/00314); ICFO Internal 'QuantumGaudi' project; European Union's Horizon 2020 research and innovation program under the Marie-Skłodowska-Curie Grant Agreement No. 101029393 (STREDCH) and No. 847648 ('La Caixa' Junior Leaders fellowships ID100010434: LCF/BQ/PI19/11690013, LCF/BQ/PI20/11760031, LCF/BQ/PR20/11770012, LCF/BQ/PR21/11840013). Views and opinions expressed in this work are, however, those of the author(s) only and do not necessarily reflect those of the European Union, European Climate, Infrastructure and Environment Executive Agency (CINEA), nor any other granting authority. Neither the European Union nor any granting authority can be held responsible for them.

ORCID iDs

Tomasz Szoldra  <https://orcid.org/0000-0002-2897-0506>

Marcelo F Ciappina  <https://orcid.org/0000-0002-1123-6460>

Nicholas Werby  <https://orcid.org/0000-0002-1789-3170>

Philip H Bucksbaum  <https://orcid.org/0000-0003-1258-5571>

Maciej Lewenstein  <https://orcid.org/0000-0002-0210-7800>

Jakub Zakrzewski  <https://orcid.org/0000-0003-0998-9460>

Andrew S Maxwell  <https://orcid.org/0000-0002-6503-4661>

References

- [1] Karagiorgi G, Kasieczka G, Kravitz S, Nachman B and Shih D 2022 *Nat. Rev. Phys.* **4** 399
- [2] VanderPlas J, Connolly A J, Ivezić Ž and Gray A 2012 *2012 Conf. on Intelligent Data Understanding* pp 47–54
- [3] Carleo G, Cirac I, Cranmer K, Daudet L, Schuld M, Tishby N, Vogt-Maranto L and Zdeborová L 2019 *Rev. Mod. Phys.* **91** 045002
- [4] Dawid A et al 2022 Modern applications of machine learning in quantum sciences (arXiv:2204.04198 [cond-mat, physics:quant-ph])
- [5] Rawat W and Wang Z 2017 *Neural Comput.* **29** 2352
- [6] Li Z, Liu F, Yang W, Peng S and Zhou J 2022 *IEEE Trans. Neural Netw. Learn. Syst.* **33** 6999
- [7] Kaufmann K, Zhu C, Rosengarten A S, Maryanovsky D, Harrington T J, Marin E and Vecchio K S 2020 *Science* **367** 564
- [8] Hezaveh Y D, Lévassieur L P and Marshall P J 2017 *Nature* **548** 555
- [9] Krausz F and Ivanov M 2009 *Rev. Mod. Phys.* **81** 163
- [10] Salières P, L’Huillier A, Antoine P and Lewenstein M 1999 *Advances in Atomic, Molecular and Optical Physics* vol 41, ed B Bederson and H Walther (Academic) pp 83–142
- [11] Lewenstein M and L’Huillier A 2009 *Strong Field Laser Physics (Springer Series in Optical Sciences)* ed T Brabec (Springer) pp 147–83
- [12] Ciappina M F et al 2017 *Rep. Prog. Phys.* **80** 054401
- [13] Itatani J, Levesque J, Zeidler D, Niikura H, Pépin H, Kieffer J C, Corkum P B and Villeneuve D M 2004 *Nature* **432** 867
- [14] Zuo T, Bandrauk A D and Corkum P B 1996 *Chem. Phys. Lett.* **259** 313
- [15] Huismans Y et al 2011 *Science* **331** 61
- [16] Figueira de Morisson Faria C and Maxwell A S 2020 *Rep. Prog. Phys.* **83** 034401
- [17] Hentschel M, Kienberger R, Spielmann C, Reider G A, Milosevic N, Brabec T, Corkum P, Heinzmann U, Drescher M and Krausz F 2001 *Nature* **414** 509
- [18] Itatani J, Quéré F, Yudin G L, Ivanov M Y, Krausz F and Corkum P B 2002 *Phys. Rev. Lett.* **88** 173903
- [19] Paul P M, Toma E S, Breger P, Mullot G, Augé F, Balcou P, Muller H G and Agostini P 2001 *Science* **292** 1689
- [20] Muller H 2002 *Appl. Phys. B* **74** s17
- [21] Liu X, Zhang G, Li J, Shi G, Zhou M, Huang B, Tang Y, Song X and Yang W 2020 *Phys. Rev. Lett.* **124** 113202
- [22] Lytova M, Spanner M and Tamblyn I 2022 *Can. J. Phys.* **101** 132–40
- [23] Chomet H, Plesnik S, Nicolae D C, Dunham J, Gover L, Weaving T and Figueira de Morisson Faria C 2022 *J. Phys. B: At. Mol. Opt. Phys.* **55** 245501
- [24] Shvetsov-Shilovski N I and Lein M 2022 *Phys. Rev. A* **105** L021102
- [25] Shvetsov-Shilovski N I and Lein M 2023 *Phys. Rev. A* **107** 033106
- [26] Liu X, Amini K, Sanchez A, Belsa B, Steinle T and Biegert J 2021 *Commun. Chem.* **4** 154
- [27] Brunner C, Duensing A, Schröder C, Mittermair M, Golkov V, Pollanka M, Cremers D and Kienberger R 2022 *Opt. Express* **30** 15669
- [28] Trebino R, DeLong K W, Fittinghoff D N, Sweetser J N, Krumbügel M A, Richman B A and Kane D J 1997 *Rev. Sci. Instrum.* **68** 3277
- [29] Kiełpinski D, Sang R T and Litvinyuk I V 2014 *J. Phys. B* **47** 204003
- [30] Pullen M G et al 2013 *Phys. Rev. A* **87** 053411
- [31] Maxwell A S, Serafini A, Bose S and Figueira de Morisson Faria C 2021 *Phys. Rev. A* **103** 043519
- [32] Bukharskii N D, Vais O E, Korneev P A and Bychenkov V Y 2023 *Matter Radiat. Extremes* **8** 014404
- [33] Sanchez-Gonzalez A et al 2017 *Nat. Commun.* **8** 15461
- [34] Ren X, Edelen A, Lutman A, Marcus G, Maxwell T and Ratner D 2020 *Phys. Rev. Accel. Beams* **23** 040701
- [35] Kolesnichenko P V and Zigmantas D 2023 *Opt. Express* **31** 11806
- [36] Geffert O, Kolbasova D, Trabattoni A, Calegari F and Santra R 2022 *Opt. Lett.* **47** 3992
- [37] Tulskey V and Bauer D 2019 *Comput. Phys. Commun.* **251** 107098
- [38] Tong X M and Lin C D 2005 *J. Phys. B: At. Mol. Opt. Phys.* **38** 2593
- [39] Amini K et al 2019 *Rep. Prog. Phys.* **82** 116001
- [40] Becker W, Grasbon F, Kopold R, Milošević D B, Paulus G G and Walther H 2002 *Advances in Atomic, Molecular and Optical Physics* vol 48, ed B Bederson and H Walther (Academic) pp 35–98
- [41] Figueira de Morisson Faria C, Schomerus H and Becker W 2002 *Phys. Rev. A* **66** 043413
- [42] Keldysh L V 1965 *Sov. Phys.—JETP* **20** 1307
- [43] Faisal F H M 1973 *J. Phys. B: At. Mol. Phys.* **6** L89
- [44] Reiss H R 1980 *Phys. Rev. A* **22** 1786
- [45] Even U 2015 *EPJ Tech. Instrum.* **2** 17
- [46] Roberts G M, Nixon J L, Lecointre J, Wrede E and Verlet J R R 2009 *Rev. Sci. Instrum.* **80** 053104
- [47] Werby N, Maxwell A S, Forbes R, Bucksbaum P H and Figueira de Morisson Faria C 2021 *Phys. Rev. A* **104** 013109
- [48] Maxwell A S, Al-Jawahiry A, Das T and Figueira de Morisson Faria C 2017 *Phys. Rev. A* **96** 023420
- [49] Maxwell A S, Al-Jawahiry A, Lai X Y and Figueira de Morisson Faria C 2018 *J. Phys. B: At. Mol. Opt. Phys.* **51** 044004
- [50] Maxwell A S and Figueira de Morisson Faria C 2018 *J. Phys. B: At. Mol. Opt. Phys.* **51** 124001
- [51] LeCun Y, Boser B, Denker J S, Henderson D, Howard R E, Hubbard W and Jackel L D 1989 *Neural Comput.* **1** 541
- [52] Krizhevsky A, Sutskever I and Hinton G E 2012 *Advances in Neural Information Processing Systems* vol 25, ed F Pereira, C Burges, L Bottou and K Weinberger (Curran Associates, Inc.)
- [53] Russakovsky O et al 2014 Imagenet large scale visual recognition challenge (arXiv:1409.0575 [cs.CV])
- [54] Chen X et al 2023 Symbolic discovery of optimization algorithms (arXiv:2302.06675 [cs.LG])
- [55] Tan C, Sun F, Kong T, Zhang W, Yang C and Liu C 2018 A survey on deep transfer learning (arXiv:1808.01974 [cs.LG])
- [56] Plested J and Gedeon T 2022 Deep transfer learning for image classification: a survey (arXiv:2205.09904 [cs.CV])
- [57] Simonyan K and Zisserman A 2015 Very deep convolutional networks for large-scale image recognition (arXiv:1409.1556 [cs])
- [58] Chollet F 2017 Xception: deep learning with depthwise separable convolutions (arXiv:1610.02357 [cs])
- [59] Tan M and Le Q V 2020 EfficientNet: rethinking model scaling for convolutional neural networks (arXiv:1905.11946 [cs, stat])
- [60] Tan M and Le Q V 2021 EfficientNetV2: smaller models and faster training (arXiv:2104.00298 [cs])

- [61] Chollet F et al 2015 Keras (available at: <https://keras.io>)
- [62] Iqbal H 2018 Harisiqbal88/plotneuralnet (v1.0.0) (Zenodo) <https://doi.org/10.5281/zenodo.2526396>
- [63] Gawlikowski J et al 2022 A survey of uncertainty in deep neural networks (arXiv:2107.03342 [cs, stat])
- [64] Abdar M et al 2021 *Inf. Fusion* **76** 243
- [65] Nix D and Weigend A 1994 *Proc. 1994 IEEE Int. Conf. on Neural Networks (ICNN'94)* vol 1 pp 55–60
- [66] Lakshminarayanan B, Pritzel A and Blundell C 2017 Simple and scalable predictive uncertainty estimation using deep ensembles (arXiv:1612.01474 [cs, stat])
- [67] Ovadia Y, Fertig E, Ren J, Nado Z, Sculley D, Nowozin S, Dillon J V, Lakshminarayanan B and Snoek J 2019 Can you trust your model's uncertainty? Evaluating predictive uncertainty under dataset shift (arXiv:1906.02530 [cs, stat])
- [68] Zimmermann J et al 2019 *Phys. Rev. E* **99** 063309
- [69] Shorten C and Khoshgoftaar T M 2019 *J. Big Data* **6** 60
- [70] Kingma D P and Ba J 2014 Adam: a method for stochastic optimization (arXiv:1412.6980 [cs.LG])
- [71] Sluijterman L, Cator E and Heskes T 2023 Optimal training of mean variance estimation neural networks (arXiv:2302.08875 [stat.ML])
- [72] Linardatos P, Papastefanopoulos V and Kotsiantis S 2020 *Entropy* **23** 18
- [73] Mohseni S, Zarei N and Ragan E D 2021 *ACM Trans. Interact. Intell. Syst.* **11** 1
- [74] Letzgs S, Wagner P, Lederer J, Samek W, Müller K-R and Montavon G 2021 Toward explainable AI for regression models (arXiv:2112.11407 [cs, stat])
- [75] Zhang J, Petitjean C, Yger F and Ainouz S 2020 *Workshop on Interpretability of Machine Intelligence in Medical Image Computing at MICCAI 2020 (Lima, Peru)*
- [76] Covert I, Lundberg S and Lee S-I 2021 *J. Mach. Learn. Res.* **22** 1 (arXiv:2011.14878)
- [77] Bach S, Binder A, Montavon G, Klauschen F, Müller K-R, Samek W and Suarez O D 2015 *PLoS One* **10** e0130140
- [78] Alber M, Lapuschkin S, Seegerer P, Hägele M, Schütt K T, Montavon G, Samek W, Müller K-R, Dähne S and Kindermans P-J 2019 *J. Mach. Learn. Res.* **20** 1
- [79] Samek W, Binder A, Montavon G, Lapuschkin S and Müller K-R 2017 *IEEE Trans. Neural Netw. Learn. Syst.* **28** 2660
- [80] Springenberg J T, Dosovitskiy A, Brox T and Riedmiller M 2015 Striving for simplicity: the all convolutional net (arXiv:1412.6806 [cs])
- [81] Montavon G, Lapuschkin S, Binder A, Samek W and Müller K-R 2017 *Pattern Recognit.* **65** 211
- [82] Hickstein D D et al 2012 *Phys. Rev. Lett.* **109** 073004
- [83] Korneev P A et al 2012 *Phys. Rev. Lett.* **108** 223601
- [84] Kang H et al 2020 *Phys. Rev. A* **102** 013109
- [85] Maxwell A S, Figueira de Morisson Faria C, Lai X Y, Sun R P and Liu X J 2020 *Phys. Rev. A* **102** 033111