Transforming a rare event search into a not-so-rare event search in real-time with deep learning-based object detection

J. Schueler,[1, *] H. M. Araújo,[2] S. N. Balashov,[3] J. E. Borg,[4] C. Brew,[3] F. M. Brunbauer,[5] C. Cazzaniga,[6]
A. Cottle,[7] C. D. Frost,[6] F. Garcia,[8] D. Hunt,[7] A. C. Kaboth,[9, 3] M. Kastriotou,[6] I. Katsioulas,[10] A. Khazov,[3]
P. Knights,[10] H. Kraus,[7] V. A. Kudryavtsev,[11] S. Lilley,[6] A. Lindote,[12] M. Lisowska,[5, 13] D. Loomba,[1]
M. I. Lopes,[12] E. Lopez Asamar,[14] P. Luna Dapica,[6] P. A. Majewski,[3, 2] T. Marley,[2, 3] C. McCabe,[15] L. Millins,[10, 3]
A. F. Mills,[1] M. Nakhostin,[2, 3] R. Nandakumar,[3] T. Neep,[10] F. Neves,[12] K. Nikolopoulos,[10, 16] E. Oliveri,[5]
L. Ropelewski,[5] V. N. Solovov,[12] T. J. Sumner,[2] J. Tarrant,[17] E. Tilly,[1] R. Turnley,[6] and R. Veenhof[5]

(MIGDAL Collaboration)

[1]*Department of Physics and Astronomy, University of New Mexico, Albuquerque, NM, 87131, USA*
[2]*Department of Physics, Blackett Laboratory, Imperial College London, London, SW7 2AZ, UK*
[3]*Particle Physics Department, STFC Rutherford Appleton Laboratory, Didcot, OX11 0QX, UK*
[4]*Luleå University of Technology, 97187 Luleå, Sweden*
[5]*CERN, 1211 Geneva 23, Switzerland*
[6]*ISIS Neutron and Muon Source, STFC Rutherford Appleton Laboratory, Didcot, OX11 0QX, UK*
[7]*Department of Physics, Keble Road, University of Oxford, Oxford, OX1 3RH, UK*
[8]*Helsinki Institute of Physics, University of Helsinki, FI-00014 Helsinki, Finland*
[9]*Department of Physics, Royal Holloway University of London, Egham, TW20 0EX, UK*
[10]*School of Physics and Astronomy, University of Birmingham, Birmingham, B15 2TT, UK*
[11]*Department of Physics and Astronomy, University of Sheffield, Sheffield, S3 7RH, UK*
[12]*LIP – Laboratório de Instrumentação e Física Experimental de Partículas,*
University of Coimbra, P-3004-516 Coimbra, Portugal
[13]*Université Paris Saclay, F-91191 Gif-sur-Yvette, France*
[14]*Departamento de Fisica Teorica, Universidad Autonoma de Madrid, 28049 Madrid, Spain*
[15]*Department of Physics, King's College London, London, WC2R 2LS, UK*
[16]*University of Hamburg, 22767, Hamburg, Germany*
[17]*Technology Department, STFC Rutherford Appleton Laboratory, Didcot, OX11 0QX, UK*

Deep learning-based object detection algorithms enable the simultaneous classification and localization of any number of objects in image data. Many of these algorithms are capable of operating in real-time on high resolution images, attributing to their widespread usage across many fields. We present an end-to-end object detection pipeline designed for real-time rare event searches for the Migdal effect, using high-resolution image data from a state-of-the-art scientific CMOS camera in the MIGDAL experiment. The Migdal effect in nuclear scattering, crucial for sub-GeV dark matter searches, has yet to be experimentally confirmed, making its detection a primary goal of the MIGDAL experiment. Our pipeline employs the YOLOv8 object detection algorithm and is trained on real data to enhance the detection efficiency of nuclear and electronic recoils, particularly those exhibiting overlapping tracks that are indicative of the Migdal effect. When deployed online on the MIGDAL readout PC, we demonstrate our pipeline to process and perform the rare event search on 2D image data faster than the peak 120 frame per second acquisition rate of the CMOS camera. Applying these same steps offline, we demonstrate that we can reduce a sample of 20 million camera frames to around 1000 frames while maintaining nearly all signal that YOLOv8 is able to detect, thereby transforming a rare search into a much more manageable search. Our studies highlight the potential of pipelines similar to ours significantly improving the detection capabilities of experiments requiring rapid and precise object identification in high-throughput data environments.

I. INTRODUCTION

Convolutional neural networks (CNNs) as backbones for computer vision systems have found remarkable success in extracting meaningful information from image and video data. AlexNet[1] was one of the first major breakthroughs in CNN-based computer vision, where it achieved a Top-5 image classification error rate that was more than 10 percentage points lower than its closest competition in the ImageNet[2] 2012 contest. This result brought deep learning and CNNs to the forefront of modern computer vision research. Since then, CNNs have enabled a host of other computer vision applications including regression predictions of image inputs, object detection, key point detection, and instance segmentation as exemplified with the image data from our experiment shown in Fig. 1.

Image classification and regression are among the simplest computer vision applications, where images are passed through an algorithm (often a CNN) and mapped to discrete and continuous sets of outputs, respectively. Common examples of classification and regression tasks
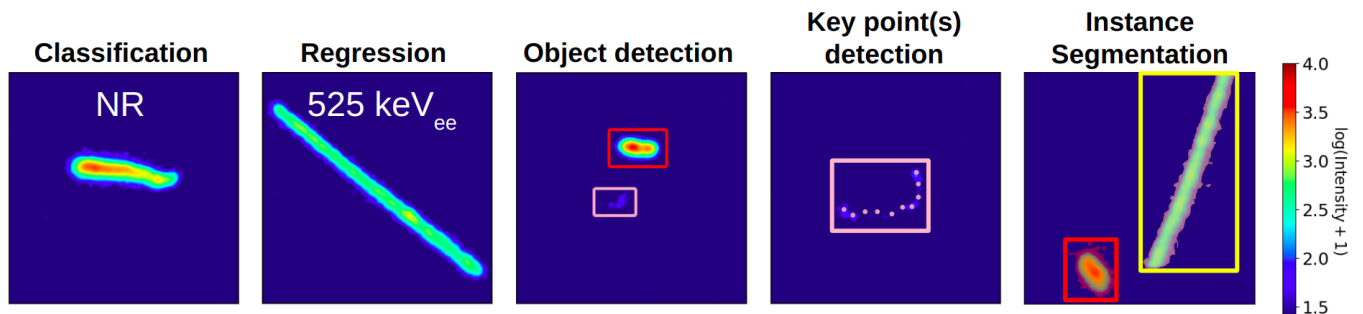
* jschueler1@unm.edu

FIG. 1. Example outputs of common CNN-based computer vision tasks illustrated on snippets of image data recorded in the MIGDAL detector. Referring to each frame from left to right: (i) Classification maps an image input to a discrete set of outputs. The example shown here is for an image classifier trained for particle identification, where the classifier predicts a nuclear recoil (NR) from neutron scattering in the input image. (ii) Regression maps the input to a continuous set of outputs. The example illustrated here is for a model trained to reconstruct energies, so the regression model reconstructs the energy present in the input image as 525 keV of visible energy. (iii) Object detection algorithms simultaneously classify and localize any number of objects in a single image. The two bounding boxes shown were predicted by our trained YOLOv8 pipeline (see Section III) and indicate that the algorithm detected a NR track (red box) and an ER track (pink box). (iv) Key point detection takes object detection a step further and identifies key points within bounding boxes; shown here is an example of particle trajectory fitting with key points. (v) Instance segmentation is arguably the most complicated of these tasks, as these algorithms assign classes to each individual pixel and further specify the instances of each assigned class. This example shows the specific pixels of a NR and proton within bounding boxes of red and yellow respectively designating the instance.

in high energy physics are particle identification [3] and energy reconstruction [4], both of which can be performed simultaneously with suitable choice of loss function and model architecture [5].

Object detection is more complicated and involves the simultaneous classification and localization of any number of objects in an input image. The output of an object detection algorithm consists of bounding boxes surrounding each identified object (center panel of Fig. 1), where each bounding box has an associated classification prediction. The Regions with CNN features (R-CNN) algorithm [6] demonstrated the first usage of CNNs for object detection in 2014, and since then, many refinements and new approaches to deep learning-based object detection have been introduced [7–10]. Due to their balance of speed and accuracy, the You Only Look Once (YOLO) [11] family of algorithms are among the most popular object detection algorithms for real-time applications. YOLO has been continuously improved upon (see Ref. [12] for a comprehensive history) with YOLOv8 [13], the version used in this work, being among the current state of the art in fast object detection.

Object detection is used widely in applications spanning many fields [14–18], and is beginning to see more usage in the physical sciences. In astronomy, YOLO in particular has found success in galaxy detection and identification [19], and the recently developed YOLO-CIANNA [20] outperformed the winner of the Square Kilometre Array Science Data Challenge 1 [21], demonstrating YOLO's efficacy in analyzing large astronomical datasets. Object detection and semantic segmentation – the class-assignment of individual pixels in an image – have garnered recent interest in neutrino physics,

with MicroBooNE [22–26] using semantic segmentation for track reconstruction in their search for the anomalous low energy excess observed by MiniBooNE [27, 28]. YOLO has also been proposed as a way to improve long distance supernova burst trigger performance in DUNE [29]. Other novel object detection approaches have also been proposed for particle physics applications outside of neutrino experiments [30].

In this work, we expand on the existing body of deep learning applications in nuclear and high energy physics by introducing a YOLOv8-based object detection pipeline capable of performing an end-to-end rare event search analysis on high resolution images online and in real-time on consumer desktop hardware. This pipeline is developed for the Migdal In Galactic Dark mAtter expLoration (MIGDAL) experiment [31], which is a neutron scattering experiment with the goal to detect and measure the rare Migdal effect [32, 33], where a recoiling nucleus is accompanied by the low-probability emission of an atomic electron. We train YOLOv8 on 2048×1152 pixel images recorded by a Hamamatsu ORCA-Quest qCMOS camera (OQC) [34] in the MIGDAL detector. The combination of readout speed and resolution (up to 1.1 gigapixel/s for 16-bit readout), high quantum efficiency ($\sim$70% at 620 nm), and low readout noise (0.43e RMS at 120 frames per second (fps)), make the OQC a state-of-the-art instrument for high resolution, high dynamic range particle imaging. As a result the OQC has been applied to work in low light microscopy [35], quantum optics [36], directional particle detection [37], and ground-based telescope imaging [38]. Our ability to analyze data from a state-of-the-art CMOS sensor in real-time, with modest compute requirements, is enabled by

the use of deep learning-based object detection, and is therefore of broader interest to applications beyond rare event search experiments.

We structure the rest of this work as follows: Section II A sets the stage by introducing the Migdal effect and the MIGDAL experiment. Section II B then discusses some important considerations in our rare event search for the Migdal effect, making it clear that the 2D OQC readout alone is not sufficient for unambiguous detection of the Migdal effect. Instead, it is used to select candidates that will require additional information from other detector subsystems for confirmation. Section III then introduces YOLOv8, our procedures for labeling data and training YOLOv8, and all steps of the raw data processing and analysis procedure that are automated by our pipeline. We also present benchmark studies testing the pipeline's end-to-end processing and analysis speeds both on typical and high occupancy data to verify that our pipeline is capable of performing a Migdal effect search on OQC data in real-time. Section IV presents several simulation studies that quantify YOLOv8's single and multiple track detection performance. These include single electron recoil (ER) and nuclear recoil (NR) identification and localization studies, as well as studies of Migdal detection efficiencies versus ER energy, NR energy, and geometrical overlap between ERs and NRs. This section concludes by quantifying YOLOv8's background rejection and signal retention when applying Migdal search criteria on simulation. Section V highlights quantities computed and reported by our pipeline online and in real-time, including those relevant to our rare event search. The section concludes with applying the same Migdal search criteria used on simulation in Section IV C 5 to a large sample of offline-stored images, demonstrating the scale of the data reduction we achieve when using YOLOv8 to search for Migdal effect candidates in real data. Finally, we summarize our key findings in Section VI.

II. OVERVIEW OF THE MIGDAL EXPERIMENT

Although the Migdal effect was predicted decades ago [32, 33], its relevance for the dark matter (DM) search community was firmly established only in 2017 – with the derivation of an explicit relationship between the Migdal electron probability and ER and NR energies in Ref. [39]. While the Migdal effect is a very rare process, this derivation elevated the Migdal effect in nuclear scattering to an attractive process for enhancing the sensitivity of sub-GeV DM searches. In this DM mass regime, detected energies from Migdal electrons can far exceed those from DM-induced NRs, enabling the indirect detection of sub-threshold NRs via the Migdal effect. Several experiments have since published DM scattering limits exploiting the Migdal effect to enhance their sensitivity to light DM [40–52] but, as of this writing, there has been

no experimental confirmation of the Migdal effect in nuclear scattering [53].

A. The MIGDAL experiment

The primary goal of the MIGDAL experiment is to make the first direct detection and measurement of the Migdal effect in nuclear scattering, which will be used to test theoretical predictions in Refs. [39, 54]. To achieve sufficient statistics, the experiment uses a commercial deuterium-deuterium (D-D) fusion generator from Adelphi Technology Inc. The generator provides an approximately monoenergetic (2.5 MeV) source of neutrons with a nominal isotropic rate of 10^9 neutrons/s incident on a Time Projection Chamber (TPC) filled with 50 torr CF_4 gas. Particle interactions with the CF_4 in the active volume produce a primary (S1) scintillation signal, and ionization amplification through a double glass-Gas Electron Multiplier (GEM) [55] layer with a gain of $\mathcal{O}(10^5)$, produce a secondary (S2) scintillation signal. Both signals are recorded by a photomultiplier tube (PMT) readout, while the latter signal is imaged by the OQC. The generator and TPC are located at NILE/ISIS at the STFC Rutherford Appleton Laboratory in the UK. To date, the MIGDAL experiment has collected D-D-generator data over two dedicated science runs spanning several weeks. We briefly introduce the MIGDAL TPC and its primary readout components, with emphasis on the OQC readout, since our pipeline is designed for this readout. More detailed information about the MIGDAL experiment and detector operations can be found in Ref. [31] and Ref. [56], respectively.

MIGDAL detector At the core of the experiment is a 110 cm^3 optical TPC with combined optical and electronic readouts, allowing for full 3D reconstructions of particle tracks [57, 58]. To minimize diffusion while maintaining reasonable neutron interaction rates, the drift region of the detector spans 3.0 cm, with a 200 V/cm electric field applied between a cathode mesh and the first GEM. The GEM active area is 10×10 cm^2; however to improve scintillation light yield, the camera was brought closer and images an 8.0×4.5 cm^2 region in the transverse x-y plane, leading to an observable active volume of $8.0\,cm \times 4.5\,cm \times 3.0\,cm$. A collimator and shield separate the D-D generator from the TPC, with the neutrons incident in the $+x$ direction of the TPC. An external 80 MBq ^{55}Fe X-ray source is attached to the TPC on a movable shaft and can be remotely deployed [31] for energy calibrations. This electron capture source emits 5.9 keV K-α and 6.5 keV K-β X-rays from ^{55}Mn fluorescence, which irradiate the active volume with relative intensities of around 87% and 13%.

PMT A 3-inch Hamamatsu R11410 PMT serves the dual purpose of measuring S1 and S2, as well as triggering the DAQ system. For events with a detectable S1 signal, the absolute z position of the recoil is determined from the time difference between the S1 and S2 signals and
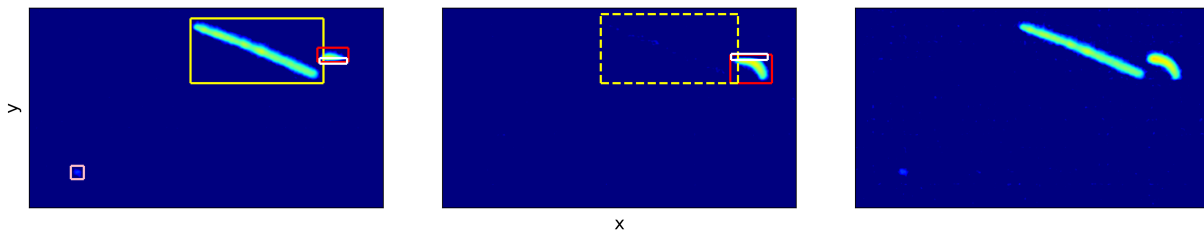
FIG. 2. Two consecutive camera frames at 20 ms exposure (left and middle panels) and their sum (right). The pink, yellow, and red bounding boxes represent YOLOv8's bounding box predictions for an ER, a proton, and a NR, respectively (see Section III for YOLO pipeline details). The white bounding boxes in the left and middle panels show YOLOv8's prediction that the rolling shutter clipped the NR and also estimate where the track was clipped. Summing these two frames together recovers the clipped NR at the expense of signal to noise. The faint long track inside the yellow dashed bounding box shows an example of a proton ghost.

the drift velocity.

ITO anode An Indium Tin Oxide (ITO) anode plane is segmented into 120 resistive strips of 833 μm pitch beyond the double-GEM stack, and provides x-z projections of events [57]. Pairs of strips spaced 60 strips (5 cm) apart are connected to 60 charge amplifiers and digitizer channels. The 500 MHz sampling rate of the digitizers provides sufficient z resolution in the ITO waveforms, given that the 3.0 cm fiducial depth of the MIGDAL detector corresponds to about 230 ns at our operating drift field in 50 torr CF_4.

Camera readout The OQC images the S2 light from the output of the second GEM through an EHD-25085-C F0.85 lens [59]. As such, throughout this work, we reconstruct track energies using this light signal and report "electron-equivalent" energies, which we denote with an ee-subscript (e.g. keV_{ee}). The OQC sensor measures 4096×2304 pixels, corresponding to a 20 μm pixel scale projected onto our field of view on the second GEM. The sensor is Peltier-cooled to $-20°C$ to limit dark current to negligible levels relative to readout noise. The camera operates with a continuous rolling-shutter with two acquisition modes designated as "Standard" and "Ultra Quiet". With Hamamatsu's proprietary CoaXPress cable to interface with the readout PC, images are recorded at up to 120 fps in Standard mode at the expense of signal to noise (0.43e RMS), while the Ultra Quiet mode provides ultra-low noise (0.27e RMS) at the expense of readout speed (5 fps). Both acquisition modes are used in the experiment, with Standard being used for recording neutrons from the D-D generator, and Ultra Quiet primarily used for energy calibrations with the ^{55}Fe source.

An FPGA pulse register records triggers from both the camera and other subsystems, providing the means for synchronization between the camera images and the rest of the DAQ system. Depending on the position of the rolling shutter relative to the location of a PMT trigger, there is ambiguity as to whether the event will be present in camera frame n or $n + 1$. Additionally, it is possible that the event is clipped by the rolling shutter, with a portion of the event in frame n and the remainder of the event in frame $n + 1$. In these cases, tracks can be recovered by stitching the two camera frames together as shown in Fig. 2. Image lag or "ghosting"– where a dimmed version of an event in a given frame persists into subsequent frames – is a separate effect that is well known to occur in CMOS cameras [60] and important to identify in our analyses. Moving forward, we will call image lagged events "ghosts".

To reduce data volume while retaining good spatial resolution,[1] we recorded data using an on chip 2×2 binning in both science runs, reducing the image size to 2048×1152 16-bit pixels. With this reduction, the raw image data collection rate at 120 fps acquisition is about 2 TB/hour. Logistical issues during Science Run 1 required the use of USB 3.1 to interface the camera with the readout PC, limiting acquisition rates to 50 fps over the entire run. In Science Run 2 we recorded neutron data at the full 120 fps using the CoaXPress cable.

B. Considerations for Migdal effect searches

A key goal of this work is to demonstrate YOLOv8's performance in selecting Migdal-like event topologies in the 2D OQC images, which is a critical step in our search. We emphasize, however, that the OQC alone is not sufficient for confirming the Migdal effect and that information from the ITO and PMT subsystems is needed to provide the full event reconstruction necessary to make a confirmation. Besides providing the 3rd dimension of the event, the 2 ns timing of the ITO is crucial for ruling out coincident ER-NR pairs unrelated to the Migdal effect, which can occur during the 8.33 ms camera exposure (Fig. 3). Therefore, selection criteria based on information integrated from all detector subsystems will be required to define the final set of Migdal candidates.

Ref. [31] sets conservative region of interest (ROI) thresholds for the MIGDAL experiment's Migdal effect

---

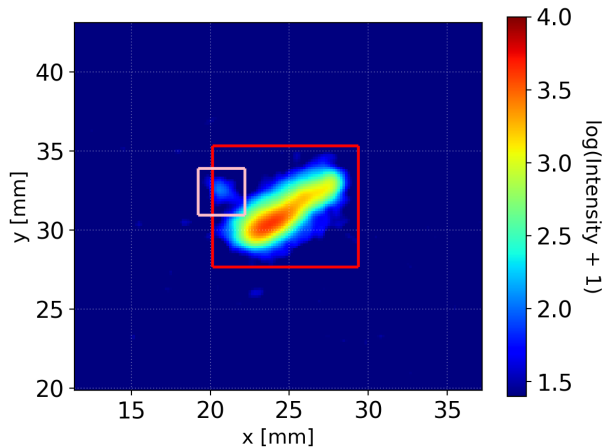[1] A 2x2 binned pixel images a 40 μm region projected on the GEMs, which have a hole pitch of 270 μm.

FIG. 3. OQC image post-processed with 4×4 pixel binning and Gaussian smoothing recorded in the presence of both an ^{55}Fe calibration X-ray source and neutrons from the D-D generator, with YOLOv8's bounding box predictions shown. This event mimics the characteristic 2D Migdal effect topology with a NR (red box; $E \sim 310\,\mathrm{keV_{ee}}$) and ER (pink box; $E \sim 6.0\,\mathrm{keV_{ee}}$) plausibly sharing the same vertex. The 2.0 ns timing resolution of the ITO strips rules this event out as an accidental coincidence (e.g. see Ref. [56]) where a D-D-induced NR and an ER from ^{55}Fe X-ray conversion happened to spatially overlap in 2D within the 8.33 ms camera exposure. Despite the two tracks overlapping without any gaps between them, YOLOv8 was able to correctly detect both tracks.

search of $15\,\mathrm{keV_{ee}} \geq E_{\mathrm{ER}} \geq 5\,\mathrm{keV_{ee}}$, $L_{\mathrm{ER}} \geq 4\,\mathrm{mm}$, and $E_{\mathrm{NR}} \geq 60\,\mathrm{keV_{ee}}$.[2] Here L_{ER} is the 2D length of the ER track. These thresholds are dictated by ensuring (1) that ERs are sufficiently long to resolve the ER head outside of the NR penumbra (like in Fig. 3), and (2) that the NR energy is high enough that there is no ambiguity in its particle identification. As will be discussed in Section IV B, YOLOv8 does an excellent job detecting NRs down to $20\,\mathrm{keV_{ee}}$ and ERs down to about $3.6\,\mathrm{keV_{ee}}$, so we are sensitive to detections well below this conservative threshold.

III. THE END-TO-END PIPELINE

We are interested in using YOLOv8 as a tool to identify topologies consistent with the Migdal effect in 2D OQC data. One approach to achieve this goal would be to train YOLOv8 to directly identify Migdal effect topologies on simulated images. This approach comes with the notable downside of the often-observed poor-performance generalization when applying machine learning models trained on simulation to real data. To circumvent the

---

[2] This is an approximation of the nuclear recoil threshold after ionization quenching, as Ref. [31] quotes the threshold in terms of recoil energy: $E_{\mathrm{NR}} \geq 100\,\mathrm{keV_r}$.

so-called Sim2Real gap [61], we instead employ a fully data-driven approach to searching for the Migdal effect in the OQC where we re-frame the Migdal search as a search for pairs of ERs and NRs within close proximity of one-another, including those that spatially overlap. Framing the Migdal search in this way allows us to train YOLOv8 on an abundance of measured ER and NR tracks observed over the course of the two science runs.

Here we detail all steps of the automated image processing and rare event search analysis pipeline. We begin with a more comprehensive overview of YOLOv8 and also detail our procedures for labeling data and training YOLOv8. After this, we detail the steps of our pipeline to process raw image data and use the extracted information to search for Migdal effect candidates. We conclude this section with a benchmark study demonstrating the end-to-end processing and analysis speeds of our pipeline on our readout PC.

A. YOLOv8

YOLOv8 is a Pytorch-based [62] open-source model released by Ultralytics in January 2023. Here we briefly describe important details of YOLOv8's loss function and the various YOLO models present in the Ultralytics YOLOv8 package. Specific details on YOLOv8's architecture can be found in Ref. [12].

A useful metric for quantifying the spatial overlap between bounding boxes is known as intersection-over-union (IoU). Given two bounding boxes B_1 and B_2, their IoU is computed as

$$\mathrm{IoU}(B_1, B_2) \equiv \frac{|B_1 \cap B_2|}{|B_1 \cup B_2|}. \tag{1}$$

Fig. 4 shows Eq. (1) diagrammatically with $\mathrm{IoU}(B_1, B_2)$ being equal to the ratio of the area of the red shaded region, to the sum of the areas of the red and blue shaded regions. Intuitively, IoU seems like a good quantity to
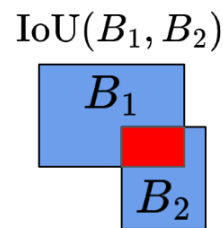

FIG. 4. Diagrammatic representation of $\mathrm{IoU}(B_1, B_2)$. B_1 and B_2 are distinct bounding boxes with their overlap region shaded in red.

optimize when training an object detection algorithm, as it is scale invariant; however, it has the problem of always being 0 when there is no overlap between a predicted bounding box and the ground truth bounding box. The introduction of Generalized intersection-over-union

(GIoU) [63] solved this problem by adding an additional penalty term to Eq. (1) that is related to the distance between the two bounding boxes, making GIoU a viable metric for a bounding box loss function that can be optimized when training. Complete intersection-over-union IoU (CIoU) was later introduced and takes into account the aspect ratio of the bounding boxes, leading to a substantial improvement in average precision scores on the PASCAL VOC 2007 dataset compared to GIoU [64].

All together, the YOLOv8 loss function consists of three terms: (1) CIoU loss; (2) Distribution Focal Loss (DFL) [65]; and (3) Binary cross entropy loss with a sigmoid function applied to the class prediction associated with a bounding box. DFL also aims to optimize bounding box regression but, unlike CIoU, DFL predicts the distribution of possible bounding box offsets, thereby reducing the uncertainty in bounding box location. Terms (1) and (2) together, then, optimize YOLOv8's localization performance, while term (3) optimizes bounding box classification. For a given image, YOLOv8's loss function achieves a minimum when all predicted bounding boxes and their associated classifications agree exactly with the ground truth labels for that image. The exact functional form for YOLOv8's loss function can be found in Appendix B.

Out of the box, the Ultralytics YOLOv8 package contains five model architectures labeled in order of smallest to largest as "n", "s", "m", "l", and "x". When evaluated on the MS COCO 2017 test set [66], a trained model "x" outperformed YOLOv5, YOLOv6, and YOLOv7 [13]. In addition to object detection, the YOLOv8 package also contains models for image classification (that can be easily adapted to regression), key point detection, and instance segmentation, so the YOLOv8 package is capable of performing all of the tasks highlighted in Fig. 1. Hereafter, we drop the "v8" designation and refer to YOLOv8 simply as YOLO.

B. Data labeling and training YOLO

Of the five previously mentioned YOLO model architectures, we train model "m" from scratch using measured OQC data, as we find this model to strike the best balance between object detection performance and inference speed for our real-time application. Naturally, training on measurement with human-labeling comes with the trade-off of model performance being limited to human-level performance. Ref. [67] shows that human-level performance limitations can potentially be overcome by augmenting real data training samples with high quality labeled simulation. We test this later (Sections IV C 2–IV C 5) and find that augmenting our real data training set with simulation improves performance in several metrics. Human labeling also comes with the downside of the labeling process being labor intensive, however once a model is sufficiently trained, we can use automated pre-annotations to significantly speed up the data labeling

process. Biases in drawing bounding boxes, as well as assigning class labels to events with ambiguous particle identification in the 2D OQC images, such as short-length sub-10 keV$_{ee}$ ER and NR tracks, can also be problematic. In our case, all training data were labeled by a single user, which may reduce variation in the geometries of labeled bounding boxes, but may also amplify biases in assigning class labels to ambiguous tracks. Using physically motivated metrics to assign track labels can help reduce bias in these scenarios.

We label data using an open-source data labeling platform called Label Studio [68], that provides data labeling templates for many machine learning applications, including several computer vision tasks. We use an object detection template that provides a click-and-drag interface for bounding box labeling. After annotating an image with all appropriate bounding boxes, the template generates a label text file for the image in the normalized $xywh$ format – a commonly used format for bounding box labels. Specifically, each ground truth bounding box, B_t, in the labeled image corresponds to a single line in the label text file that contains the following contents in this order: (i) the class label y_t; (ii) and (iii) the coordinates of the x and y centroids of B_t: $\bar{b}_{x,t}$ and $\bar{b}_{y,t}$, respectively; and (iv) and (v) the width and height of B_t: w_t and h_t, respectively. We denote bounding boxes predicted by YOLO as B_p.

Label Studio has functionality for automated pre-annotations through machine learning-assisted labeling,

TABLE I. Class name and total number of instances (number of bounding boxes) of each of the nine classes in our labeled training data. All labeled data were extracted from measurements recorded during the two science runs. Frames containing multiple instances of tracks (including those of the same class) are included in this sample.

Class Label	$N_{\text{instances}}$
Electron recoil or NR ghost[a]	4396
Hot pixel	130
Nuclear recoil	3840
Proton or alpha	300
Proton ghost or alpha ghost	63
Rolling shutter clip	539
Spark	135
Spark ghost	172
High occupancy shower[b]	58
Total	9633

[a] Due to the similar dE/dx signatures of NR ghost tracks and ER tracks, we label NR ghosts as ERs and train YOLO to initially classify NR ghosts as ERs. NR ghosts are flagged at a later stage in our pipeline (see Section III C).

[b] These events consist of dense pileups of ER tracks that cover significant portions of an image frame. We are unable to resolve individual tracks among the pileup, so we train YOLO to identify the presence of these showers so we can reject them from our analysis. These showers are relatively uncommon, typically occurring in fewer than one in one thousand frames at 120 fps acquisition.
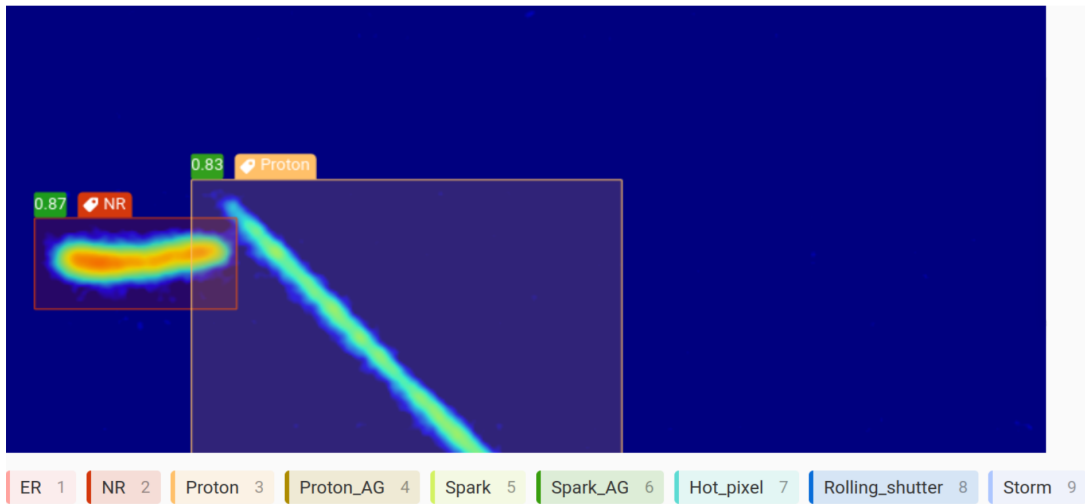
FIG. 5. Snippet of the Label Studio front-end interface showing pre-annotated bounding boxes with classification labels and scores, produced by a YOLO model pre-trained on the training data from Table I.

which we utilize to streamline the image labeling process. After YOLO has been sufficiently trained, we inject the trained model weights to a custom-written machine learning back-end plugin that interacts with Label Studio's object detection template. This plugin evaluates our trained YOLO model to pre-annotate uploaded images with editable bounding boxes in the Label Studio front end. Fig. 5 shows an example of the automated pre-annotations placed on a NR candidate and a proton candidate in the Label Studio interface. Generally speaking, as model performance improves with more training, fewer manual adjustments are required on pre-annotated data, thereby streamlining the labeling process.

Table I shows the class breakdown of human-labeled data used to train YOLO. Since Migdal searches are our application of interest, the overwhelming majority of tracks we train on are ERs and NRs, as these are the constituents of Migdal effect topologies. We split the set of training frames with their associated labels into a 70%/30% training/validation split. In each training epoch, we optimize YOLO's model weights using stochastic gradient descent with momentum and weight decay [69] and employ early stopping [70] where we terminate training after 25 successive epochs with no improvement in the mAP@50:95:5 metric (mean average precision over IoU thresholds of 0.5 to 0.95 in steps of 0.05) [71]. When training YOLO using a training set that is augmented with additional simulation (as in Section IV C 2), we follow this same training procedure.

C. YOLO-based data processing and analysis pipeline

Next, we describe the automated pipeline for processing and analyzing images batch-by-batch. This pipeline is integrated with our DAQ software and begins processing batches of 200 camera frames as they are written to disk. The steps of our pipeline are as follows:

1. **Dark subtraction**: We pre-generate two 2048×1152 master dark frames: one that is used for Ultra Quiet data acquisition and the other for Standard acquisition. These frames are generated from two dedicated dark runs where, in each run, 800 camera frames are recorded with the rest of the detector powered off. Master dark frames are created from the mean intensity of each pixel over the 800 frames, with sigma clipping at a 5σ tolerance applied.[3] During a run, the appropriate master dark frame is subtracted from each batch of recorded images.

2. **Downsample and Gaussian smoothing**: Each batch of images is downsampled using 4×4 binning, yielding 200 frames with 512×288 pixels. A 3×3 Gaussian smoothing kernel with $\sigma_x = \sigma_y = 4/3$ is then applied to the stack of frames. For performance, we use Pytorch to perform both the downsampling and Gaussian smoothing on a GPU.

3. **Convert images to PNG format**: Each camera frame is converted into a PNG file to be passed as input into YOLO. To best capture the large dynamic range required for simultaneously elucidating ERs and NRs, we use an empirically-determined fixed logarithmic intensity scale with a

---

[3] Sigma clipping is used to remove anomalous pixel intensities when creating master dark frames. First, we compute the mean and standard deviation intensity of each image-pixel over the 800 dark frames. Then, we mask all pixels with intensities greater than 5σ of the mean and repeat the process until there are no remaining pixels to mask.

minimum of $\log_{10}(I + 1) = 1.4$ and maximum of $\log_{10}(I+1) = 4$, where I is the intensity of a 4×4-binned, Gaussian-smoothed pixel.

4. **Pass the PNG images into YOLO and store the coordinates and classifications of each bounding box**: The trained version of YOLO simultaneously localizes and classifies (using the nine class labels from Table I) all objects in each PNG image. The bounding box pixel coordinates on the PNG images are mapped back to the original 200 downsampled and Gaussian smoothed 512×288 frames. YOLO runs on a GPU for performance.

5. **Extract physical information from bounding box contents**: A key benefit of YOLO classifying and localizing events at the same time is that these classifications can immediately be used to decide which bounding boxes to extract physical content from. We do not want to waste computational resources performing fits on a spark event, for example. Given this, for each bounding box identified as an ER, NR, or proton/alpha, we compute the following quantities: (i) track intensity/energy, (ii) track length, (iii) axial angle, (iv) for ERs and NRs only: head/tail asymmetries and vector angles (Section V A). Computations of track energies require intensity corrections due to vignetting (Appendix A), which is done using data from ^{55}Fe calibration runs.

6. **Flag and reject NR ghosts**: NR ghosts have the potential to be a significant background with dE/dx signatures that mimic ERs, so they must be flagged and rejected to reduce false positives in a Migdal search. Since YOLO is trained to initially identify NR ghosts as ERs, we flag NR ghosts by first computing the IoU overlap between each ER in a given frame and each NR in the previous frame. Then, if an ER in the given frame and NR in the previous frame have IoU overlap greater than 0 and the peak pixel intensity of the 4×4 binned Gaussian smoothed NR is greater than an empirically determined threshold of $150\,\mathrm{ADU}$, we flag that initially identified-ER as a NR ghost.

7. **Identify and analyze all unique fiducialized ER-NR pairs**: After rejecting NR ghosts, we identify all frames containing fiducialized pairs of ERs and NRs. For each unique pair, we compute $\mathrm{IoU}(B_\mathrm{p}^\mathrm{ER}, B_\mathrm{p}^\mathrm{NR})$, and the distance between the centroids of B_p^ER and B_p^NR. This is a critical step in identifying Migdal effect candidates that will be discussed in more detail in Sections IV C and V B.

8. **Save extracted information**: For each image batch, we write an output file with its contents indexed by bounding-box, thereby converting frame-indexed data to event-indexed data. In addition to the physical information extracted from each bounding box, we also store the coordinates of the corners of the bounding box, classifications with associated confidence scores, and metadata such as image timestamps, run IDs, file and frame indices, and, optionally, the coordinates and intensities of each pixel within each bounding box. Pairs of ERs and NRs showing up on the same frame that satisfy user-defined Migdal search criteria are flagged as candidate events and are saved separately alongside of the event-indexed data. Candidate event files are indexed by ER-NR pair and include pair-specific quantities such as the IoU overlap between the ER and NR in the pair, as well as the distance between the centroids of the ER and NR bounding boxes.

During a run, after the pipeline performs the above steps, each event-indexed processed file is then added to a temporary database that reports event rates, track energy versus length distributions of several event species, and NR energy spectra in a live display that updates in real-time (see Fig. 16 in Section V). Counts of candidate ER-NR pairs are also enumerated in the live display. At the end of the run, all processed files are transferred to an offline computing server.

D. Benchmarking end-to-end processing speed

To verify our claim that this pipeline is capable of processing and performing a rare event search on OQC data online and in real-time, with modest hardware requirements, we developed a benchmark script available at: https://github.com/jschuel/migYOLO/.
This script can be run on any PC but its configuration parameters have been optimized for the MIGDAL readout PC which uses an Intel Core i9-10900X CPU and a single slot NVIDIA RTX A4000 graphics card. On the MIGDAL readout PC, the script runs two processes:

Process 1: This process is run as three parallel subprocesses that each read in a batch of 200 raw images, perform steps 1-3 of the pipeline, and save the processed image batch and 200 PNG images. Each iteration of Process 1 therefore processes 600 raw images.

Process 2: This process performs steps 4-8 of the pipeline using the outputs of Process 1 as input. When this process completes, it deletes the processed image batch and PNG images generated in Process 1.

To compute the end-to-end pipeline processing time, timestamps are saved for each image batch at the beginning of Process 1 and at the end of Process 2. Each raw batch of images has a file size of around $900\,\mathrm{MB}$, so to avoid data storage burdens when benchmarking the processing time of the pipeline, we run the benchmark script repeatedly on a single batch of 200 images. We perform

two benchmarks: one using a randomly selected batch of 200 images and the other using a custom high occupancy batch formed with 200 copies of the frame shown in Fig. 6.
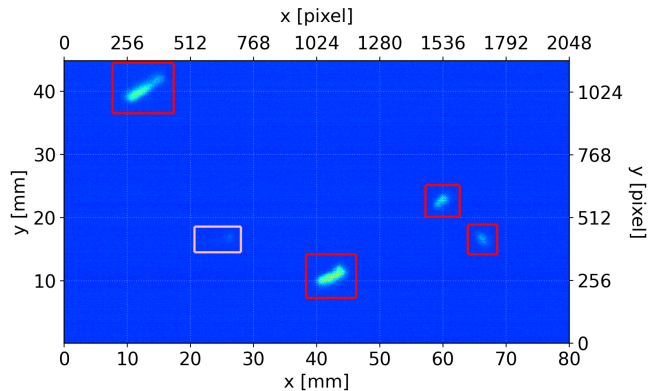

FIG. 6. Unprocessed frame used for the high occupancy batch with YOLO's bounding box predictions shown. The pink bounding box is an ER prediction and the red bounding boxes are NR predictions.
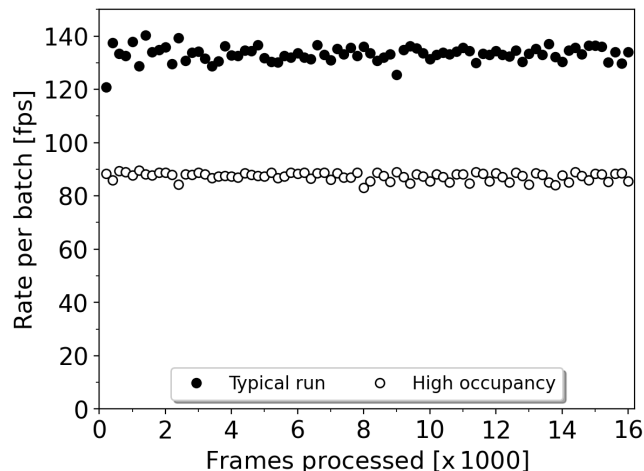

FIG. 7. Pipeline processing rate as a function of number of frames processed for a batch with events corresponding to a typical 20 ms exposure run (filled points), and our custom high occupancy batch (unfilled points). Downsampling, which we parallelize into three subprocesses, is the performance bottleneck for the typical run sample, leading to a larger variance in when the inputs of Process 2 are generated, hence the larger spread in overall processing rates for this sample.

Fig. 7 shows the results of this benchmark study evaluated over 16000 images (80 batches). The filled points show the processing speed of a randomly selected batch from a typical 20 ms exposure (50 fps) run in the presence of neutrons from the D-D generator. The event occupancy within frames in such a run is, on average, 2.4 times higher than an identical run with 120 fps acquisition. Nevertheless, our benchmark shows that we

consistently process and analyze these frames faster than 120 fps. While our benchmark does not achieve 120 fps on the high occupancy batch, we achieve an average processing rate near 90 fps, indicating that our pipeline is able to process and analyze tracks at rates close to 450 Hz in a higher occupancy environment. These benchmark studies not only show that we are able to process and analyze typical data online and in real-time, but also that our pipeline is capable of processing significantly higher track rates than the designed peak NR rate of MIGDAL [31].

IV. PIPELINE PERFORMANCE

With the end-to-end speed of our pipeline established, we next evaluate the object detection performance of our implementation of YOLO. While we train YOLO on measurement, we need simulation to quantitatively evaluate YOLO's detection and localization performance. We therefore perform several simulation studies of increasing complexity: In Section IV A, we detail the process of producing realistic optical image simulation using simulated ^{55}Fe X-ray tracks. We start with ^{55}Fe both because its spectrum covers a relatively small dynamic range, and because it allows us to tune and compare effective gains in simulation and measurement. Beginning with the procedure described in Ref. [31] to produce digitized optical pixel intensities of ^{55}Fe tracks, we apply additional intensity scalings and position-dependent vignetting, and finally add noise from randomly selected dark frames. We verify both this procedure and YOLO's ER identification performance by comparing the simulated ^{55}Fe intensity spectrum to an ^{55}Fe intensity spectrum reconstructed from a measured sample of ^{55}Fe X-rays in the presence of ERs from the D-D generator. Next, Section IV B studies YOLO's ER and NR track-identification performance more broadly, using frames containing only single tracks. These frames contain either simulated ERs drawn from a discrete uniform energy spectrum or simulated NRs drawn from a continuous energy spectrum. Finally, in Section IV C we provide detailed analyses of YOLO's performance in identifying multiple tracks in a single frame, with an emphasis on Migdal candidate detection and accidental coincidence background rejection (frames containing independent ERs and NRs that are not consistent with the Migdal effect). We perform these two-track analyses primarily on "hybrid" simulations, which consist of measured NR frames that are stitched together with simulated ERs to form either simulated Migdal events or simulated coincidences.

A. Simulated ^{55}Fe generation and identification

Adequate low-energy electron identification (EID) and localization is essential for detecting Migdal effect candidates. To assess YOLO's EID performance and validate our simulation of detector effects, we begin by evaluat-

ing YOLO on simulated ^{55}Fe tracks. Since this version of YOLO was trained exclusively on measurement, and there are inevitable differences between simulated and observed tracks, we expect the results shown here to underestimate the EID performance of YOLO when evaluated on measurement.

We generate 71432 frames, each containing a single ^{55}Fe track from either the 5.9 keV K-α or 6.5 keV K-β decay mode (statistical breakdown in Table II) following the general optical readout simulation procedure described in Ref. [31], but binning to the dimensions of the OQC. We then perform the following procedure to incorporate further effects in the optical system:

1. **Perform gain scaling:** Scale simulated tracks' pixel intensities using an empirically determined multiplicative factor with resolution smearing.

2. **Apply vignetting scaling:** Vignetting in the optical readout causes pixel intensities to decrease with distance from the center of the readout. We simulate this effect by first randomizing the location of simulated tracks along the readout (shifting using random uniform distributions in x and y) and then applying the following correction to each pixel in the track

$$I(x,y) = I_0(x,y)\frac{(a-r)^2}{a^2}. \qquad (2)$$

Here, $I_0(x,y)$ is the gain-scaled pixel intensity of the pixel at location (x,y), r is the distance between the pixel at (x,y) and the center pixel of the readout $(\overline{x},\overline{y})$, and $I(x,y)$ is the resulting intensity of the pixel at (x,y) after applying vignetting scaling. We empirically assign a to be 95 mm to achieve good agreement between this model of vignetting and the vignetting we observe in a typical ^{55}Fe run (see Appendix A).

TABLE II. Electron identification performance summary on simulated frames containing a single ^{55}Fe track. Columns from left to right: (1) Truth decay mode, (2) total number of frames where YOLO identified a single ^{55}Fe track (true positive detections), (3) number of frames where YOLO identified >1 ^{55}Fe track, (4) number of frames where YOLO did not identify any tracks, and (5) total number of frames. Frames where YOLO did not identify any tracks mostly consist of tracks located near the edge of the readout, so vignetting heavily suppresses their intensities to below threshold. On the other hand, most instances of YOLO identifying >1 ER bounding box were cases where YOLO drew multiple boxes around the same track. In a few instances, YOLO drew a bounding box around noise.

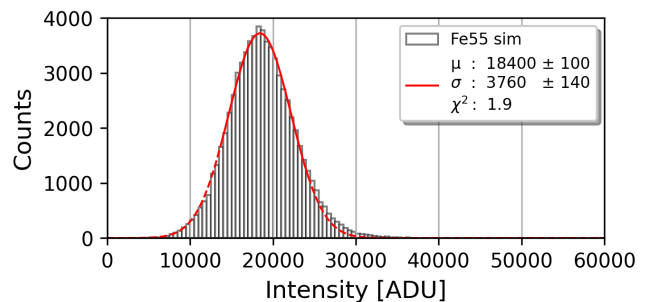| Mode	1 ER	>1 ER	No tracks	Total
K-α	60920 (98.1%)	456 (0.7%)	720 (1.2%)	62096
K-β	9113 (97.6%)	115 (1.2%)	108 (1.2%)	9336

FIG. 8. Intensity spectrum of simulated ^{55}Fe tracks identified by YOLO after applying gain scaling, vignetting, and adding noise. Data-driven vignetting corrections are computed following the procedure outlined in Appendix A to achieve the shown spectrum. The solid line-portion of the curve fit is our fit-region for a single-peak Gaussian.

3. **Apply noise:** Randomly select a measured dark-subtracted dark frame from a sample of 200 such frames and add it to the simulated ^{55}Fe frame.

Performing all steps of our pipeline on the resulting images, including vignetting corrections (Appendix A), we obtain the intensity spectrum shown in Fig. 8. While contributions of K-β X-rays are not completely negligible, we opt to perform a single-peak Gaussian fit, so our reported resolution of $\sigma/\mu = 20.4\%$ is a slight overestimate of the true resolution of the 5.9 keV ER peak. Table II shows that YOLO trained on measurement is excellent at detecting simulated ^{55}Fe tracks, with YOLO correctly predicting one and only one ER bounding box, $B_{\rm p}^{\rm ER}$, that has nonzero overlap with the ground truth bounding box of the frame ($\mathrm{IoU}(B_{\rm p}^{\rm ER}, B_{\rm t}^{\rm ER}) > 0$) in 98.1% of K-$\alpha$ frames and 97.6% of K-β frames. We note that the difference in false positive ER identification rates between the K-β and K-α samples is small but significant. Further study is required to understand this difference, but we hypothesize the difference to be the result of K-β tracks having slightly longer low-ionization tails on average than K-α tracks. If portions of these tails fall below the intensity threshold of the PNG images passed into YOLO, then YOLO may be more likely to assign multiple bounding boxes to the track. We expect training on a larger sample of higher energy ERs would resolve this discrepancy.

We next compare the simulated ^{55}Fe spectrum fit in Fig. 8 to a reconstructed ^{55}Fe spectrum on data recorded in the presence of both the D-D generator and an ^{55}Fe X-ray source. Performing such a comparison allows us to simultaneously validate our simulation of detector effects while qualitatively validating YOLO's EID performance on measurement through recovering the ^{55}Fe peak in a mixed-field sample of data with a continuous background ER spectrum. Fig. 9 shows the results of this study with the top panel showing the ER intensity spectrum during the run with both the D-D-generator and ^{55}Fe source present (labeled in legend as "DD+^{55}Fe"), as well as
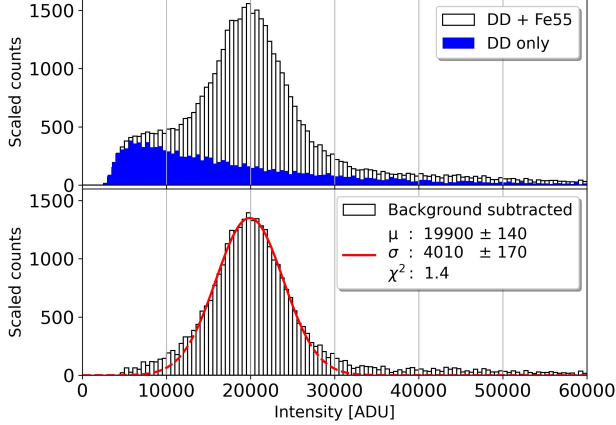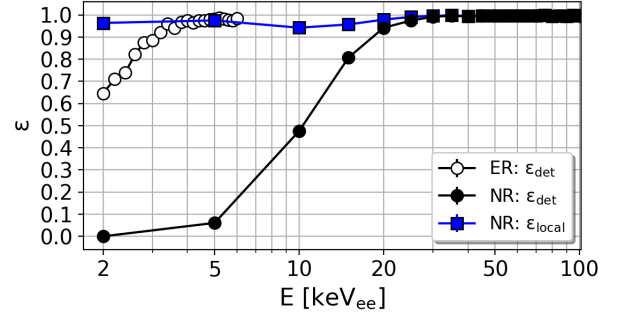
FIG. 10. Localization and detection efficiencies versus truth energy (SRIM quenching factors applied to NR energies) for frames containing single simulated tracks. For ERs, $\varepsilon_{\mathrm{det}} = \varepsilon_{\mathrm{local}}$, so we do not include a separate $\varepsilon_{\mathrm{local}}$ trace for ERs.

FIG. 9. Top: Electron recoil intensity spectrum for a D-D run with the ^{55}Fe source present ("DD+^{55}Fe"; black bars) and a D-D run without the ^{55}Fe source present ("DD only"; blue shaded) scaled to the equivalent elapsed time of the D-D+^{55}Fe run. Bottom: Recovered ^{55}Fe spectrum after background subtracting the time-scaled D-D run spectrum from the D-D+^{55}Fe signal spectrum. The solid line-portion of the curve fit is our fit-region for a single-peak Gaussian.

the "background" spectrum consisting of ERs recorded with only the D-D-generator ("DD only" in legend). NR ghosts are rejected from both spectra and the background spectrum counts are normalized to the elapsed time of the D-D+^{55}Fe run. Subtracting this normalized background spectrum from the signal D-D+^{55}Fe data, we obtain the resulting ^{55}Fe spectrum in the bottom panel of the figure. As with Fig. 8, we fit a single-peak Gaussian to this spectrum, assuming the peak at 5.9 keV. We observe a peak of 19 900 ADU and 20.2% energy resolution, both of which are consistent with simulation when accounting for the ~15% systematic gain variation over the course of an operating day in the presence of neutrons from the D-D generator (details Ref. [56]), indicating that our simulation procedure properly models gain and ^{55}Fe track light yield.

B. General ER and NR identification on simulation

Next, we evaluate YOLO's ER and NR identification performance on frames containing a single simulated ER or NR of varying energy. We assess performance using localization efficiency, $\varepsilon_{\mathrm{local}}$, and detection efficiency, $\varepsilon_{\mathrm{det}}$ defined as

$$\varepsilon_{\mathrm{local}} \equiv \frac{N(\mathrm{IoU}(B_{\mathrm{p}}, B_{\mathrm{t}}) > 0)}{N} \tag{3}$$

$$\varepsilon_{\mathrm{det}} \equiv \frac{N\left((\mathrm{IoU}(B_{\mathrm{p}}, B_{\mathrm{t}}) > 0) \wedge (y_{\mathrm{p}} = y_{\mathrm{t}})\right)}{N}, \tag{4}$$

where $\wedge$ is the logical "and" symbol, and B_{p} and B_{t} are YOLO's bounding box prediction and the ground truth

bounding box with associated classifications y_{p} and y_{t}, respectively. The numerator of Eq. (3) is the total number of frames where YOLO identified exactly one bounding box, B_{p}, and B_{p} has nontrivial overlap with B_{t}. The numerator of Eq. (4) includes the additional restriction that the prediction associated with the bounding box agrees with the ground truth class of the event. The denominator of both of these equations is the total number of frames in the sample.

We evaluate $\varepsilon_{\mathrm{local}}$ and $\varepsilon_{\mathrm{det}}$ using a sample of 10500 simulated ERs and 46585 simulated NRs (36779 fluorine recoils and 9806 carbon recoils). The ER sample has a discrete uniform energy distribution varying in steps of 0.2 keV between 2.0 keV and 6.0 keV inclusive, and the NR sample has a continuous energy spectrum mimicking the expected energy distribution of D-D-induced nuclear recoils in the MIGDAL detector. Fig. 10 shows YOLO's identification performance for all ERs and for NRs up to 100 keV$_{\mathrm{ee}}$. Energies in this figure are reported in keV$_{\mathrm{ee}}$, where SRIM [72] quenching factors have been applied to the ground truth simulated NR energies. Additionally, we find that for the truth ER samples, all bounding boxes identified are also predicted to be ERs, so $\varepsilon_{\mathrm{det}} = \varepsilon_{\mathrm{local}}$ for ERs. Given this, we only plot a single ER trace for ERs in the figure. Comparing the three traces, we make the following observations

1. YOLO's detection efficiency exceeds 90% for ERs down to 3.2 keV$_{\mathrm{ee}}$ – well below our 5.0 keV$_{\mathrm{ee}}$ threshold – and 70% down to 2.2 keV$_{\mathrm{ee}}$.

2. The localization efficiency of 2 keV$_{\mathrm{ee}}$ NRs far exceeds that of 2 keV$_{\mathrm{ee}}$ ERs. This is likely due to the comparatively higher $\mathrm{d}E/\mathrm{d}x$ near the center of NR tracks, leading to regions of NRs with higher signal to noise ratios than ERs of comparable track energy.

3. Below 10 keV$_{\mathrm{ee}}$, YOLO misidentifies most truth NRs as ERs. Although troubling at first glance, we emphasize that our aim here is to efficiently select Migdal candidates in the 2D OQC while

maximizing background rejection, so that identified candidates can be analyzed in all detector subsystems. There is a small probability of recording OQC frames where two independent NRs spatially coincide in 2D with one falling in the NR ROI energy range ($E_{NR} \geq 60\,\mathrm{keV_{ee}}$) and the other in the ER ROI energy range ($5\,\mathrm{keV_{ee}} \leq E \leq 15\,\mathrm{keV_{ee}}$) for Migdal searches, making it possible for a false positive Migdal identification to occur in such a frame. Analyses of backgrounds in Ref. [31], however, show essentially perfect discrimination between ERs and NRs down to $5\,\mathrm{keV_{ee}}$ for reconstructed tracks with 3D lengths above $4\,\mathrm{mm}$, so such false positives will not be an issue in our search ROI when integrating detector subsystems to reconstruct events in 3D. For this reason, the preponderance of misidentifying true NRs as ERs is actually preferential for our analysis, as we expect essentially zero false negative ER identifications, thereby maximizing ER detection efficiencies when searching for Migdal-like topologies. As we will see in Sections IV C 5 and V B, the number of Migdal candidate frames remaining after a search is small enough to be analyzed in detail in all subdetectors.

C. Migdal and coincidence identification performance

With YOLO's single track identification performance established, we next study its performance on frames containing two tracks. An analysis solely using OQC data is limited by the relatively long $8.33\,\mathrm{ms}$ exposure time, giving rise to accidental coincidences where independent ERs and NRs can be recorded in the same camera frame. The much faster $2\,\mathrm{ns}$ sampling of the ITO can and is used to reject these coincidences, but we can also use the fact that Migdal-like topologies consist of an ER and NR in close proximity to one another to reject the majority of accidental coincidences in OQC frames. Coincidences can also arise from background physical processes that generate correlated ER-NR pairs, however the background studies performed in Ref. [31] suggest that correlated coincidences satisfying our search ROI are rarer than the Migdal effect within our ROI, so we ignore these in our OQC studies. In this section, we quantify both YOLO's Migdal identification and accidental coincidence background rejection performance on simulation. Moving forward we simply refer to accidental coincidences as "coincidences".

1. Description of samples

We study Migdal and coincidence identification using both a purely simulated sample and a "hybrid" simulation, where we start with *measured* frames (real data) containing a single NR and add a single simulated ER to those frames – recalling that the Migdal probability is so small that it is unlikely we will bias this study due to the NR sample containing real Migdal events. Specifically, we create the following four samples, where each frame consists of exactly one ground truth ER and one NR:

1. **Pure simulation Migdal events**: Using the same sample of pure simulated ERs and pure simulated NRs described in Section IV B, we select around 27000 NRs with $E_{NR} > 60\,\mathrm{keV_{ee}}$. Then, for each NR frame, we shift the NR track to a random position on the readout, ensuring that no NR pixel falls within 20 pixels of any edge of the frame. After this, we select a random simulated ER track (from a sample of 10500 ERs, so a given ER can be assigned to more than one NR) and translate the ER track so its truth vertex sits on the same pixel as the point of highest intensity of the NR.[4] Once both tracks are aligned, we perform gain scaling, vignetting scaling, and then add noise using a randomly selected dark-subtracted dark frame from a sample of 800 such frames.

2. **Pure simulation coincidence events**: Selecting from the same set of pure simulated NRs and ERs, we repeat the procedure above except instead of aligning the ER truth vertex with the point of highest intensity of the NR, we separate the ER and NR vertices in both x and y drawing from separate random uniform distributions for each coordinate.

3. **Hybrid Migdal events**: We form these starting with around 27000 measured frames, where YOLO identified a single NR with a reconstructed energy of more than $60\,\mathrm{keV_{ee}}$ and nothing else. For each of these frames, a random ER from the sample of 10500 simulated ERs previously described is selected and translated so its ground truth vertex matches the point of highest intensity of the measured NR. Gain scaling and vignetting scaling are then applied to the ER. Since the NR frame is from measurement, it already contains noise, so we do not apply any noise to the simulated ER when stitching it to the NR frame. Furthermore, we do not apply any translations to the NRs, as randomly selected measured NRs should mimic the expected spatial distribution of NRs in the MIGDAL TPC. Fig. 11 shows visual examples of hybrid Migdal construction.

4. **Hybrid coincidence events**: Using the same sample of around 27000 NR frames from measurement, simulated ERs are redrawn at random and placed in a random location on the measured NR frame, ensuring that no ER pixel is within 20 pixels

----

[4] We approximate the location of NR vertices as the location of the pixel with highest intensity.
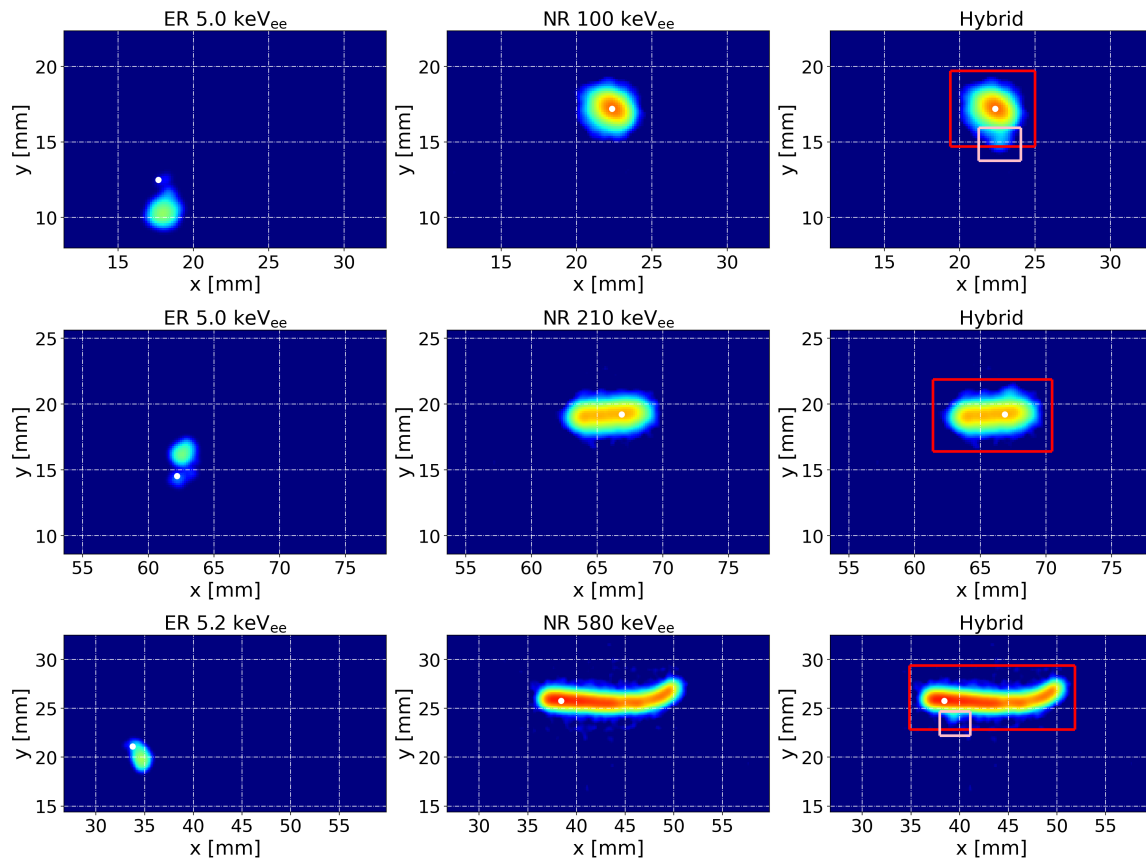
FIG. 11. Each row shows an example of stitching together the vertex of a simulated ER with the point of highest intensity of a measured NR to generate a hybrid Migdal event. Left column: Simulated ER tracks with gain scaling but no noise; truth ER vertices are shown in white. Middle column: Measured NRs with their point of highest intensity shown in white. Right column: hybrid Migdal formed by translating the ER so its vertex aligns with the point of highest intensity of the NR; vignetting scaling is applied to the ER once it's aligned with the NR. The bounding box predictions of YOLO trained on the Augment sample (Section IV C 2) are shown for the three hybrids with red (pink) boxes denoting NR (ER) predictions. The top and bottom rows show examples of positive detections. YOLO did not identify the ER in the middle-row hybrid indicating a negative detection. The fractions of ER significant pixels, f_{sigpix} (see Section IV C 3), in the top, middle, and bottom-row hybrids are 0.42, 0.27, and 0.10, respectively.

of any edge of the readout. After ERs are placed, gain scaling and vignetting corrections are applied to the ER.

All together we use around 27000 frames of each of the two pure simulation classes and a similar number of frames of each of the two hybrid simulation classes.

In terms of evaluating YOLO's Migdal detection performance, one would ideally evaluate YOLO's performance on Migdal samples generated from both measured ERs and measured NRs; however this is not practical for the following reasons:

1. Vertices of ERs are challenging to correctly locate. Unlike NRs, where the location of the most intense pixel in the image is a reasonable approximation for the vertex, tails of ER tracks tend to be regions with relatively low light yield due to their characteristic Bragg curves. Furthermore, ERs tend to form tracks with more meandering paths than

NRs, further complicating efforts to accurately detect measured ER vertices.

2. Measured ERs also have noise, thus in regions where the ER is stitched to the NR frame, noise would be double counted, which could bias deep learning-based methods of track identification.

There are clear advantages to considering hybrid simulation samples over pure simulation. Most notably, we observe larger halos forming around the outer edges of measured NR tracks that are not present in simulation [56]. These halos are especially apparent in higher energy NRs and result in an increased area of observed NR tracks compared to their simulated counterparts, further obscuring Migdal electrons. Additionally, simulating Migdal events with measured NRs has the advantage of ensuring the correct gains, energy distributions, and spatial distributions of NRs when assessing YOLO's Migdal detection performance. To keep these
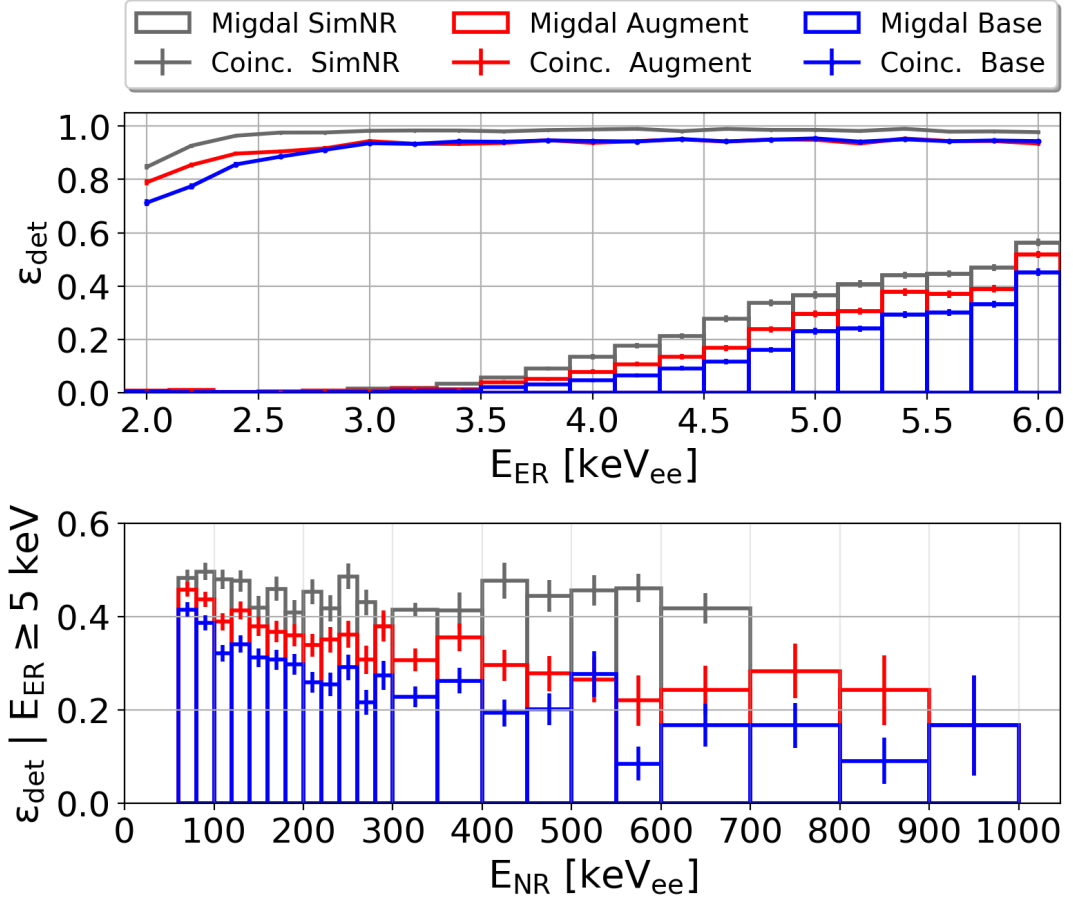
FIG. 12. Top: Simulated Migdal (bars) and coincidence (solid curves) detection efficiency as defined in Eq. (5) versus ER energy, integrated over all NR energies. Bottom: Migdal detection efficiency versus NR energy for pure simulated and hybrid Migdal events with $E_{ER} \geq 5.0 \, keV_{ee}$. Recall that simulated NR energies are ground truth energies with SRIM quenching factors applied, which is why the pure simulation histogram terminates at a lower NR energy than the hybrid simulation histograms. The training and evaluation samples for each color plotted in this figure are summarized in Table III.

performance studies general, all ERs considered in these studies have isotropic angular distributions.

2. Detection results: Simulated Migdal efficiencies versus energy

Quantifying Migdal detection efficiencies as a function of ER energy based on 2D track information from the OQC is an essential step toward deriving the total Migdal efficiency of the experiment, which will involve information from other detector subsystems such as the ITO and PMT. For each of the simulated Migdal and coincidence samples, we define YOLO's detection efficiency as

$$\varepsilon_{det} \equiv \frac{N_{det}}{N}, \quad (5)$$

where N is the total number of frames in the sample and N_{det} is the number of frames with a positive detection. A positive detection satisfies the following criteria: (1)

YOLO identified exactly one ER and one NR, and (2) YOLO correctly localized its ER and NR bounding boxes in that $IoU(B_p^{ER}, B_t^{ER}) > 0$ and $IoU(B_p^{NR}, B_t^{NR}) > 0$.

Fig. 12 summarizes YOLO's Migdal and coincidence detection efficiency performance versus energy. The legend of this figure shows results for two distinct YOLO training campaigns and different evaluation sets which are summarized in Table III.

TABLE III. Companion table for Fig. 12 showing the training and evaluation samples associated with each color in the figure. The Base training sample includes only the real data shown in Table I. The Augment training sample includes all training data from the Base sample, as well as an additional 3500 pure simulated Migdal events and 3500 pure simulated coincidences.

Legend color	Training sample	Evaluation sample
Gray	Augment	Pure simulation
Red	Augment	Hybrid
Blue	Base	Hybrid

The top panel of Fig. 12 shows ε_{det} versus ER energy over all NR energies in our sample, with bars representing Migdal detection efficiencies on the Midgal sample and solid curves representing coincidence detection efficiencies on the coincidence sample. We find that YOLO trained on the Augment sample and evaluated on hybrid simulation (red) improves on low-energy ER detection in the coincidence sample and also achieves significantly higher Migdal detection efficiencies compared to YOLO trained only on the Base sample (blue). As expected, the Migdal detection efficiencies are highest on the pure simulation sample, suggesting that less of the ER is, on average, obscured by the NR, since simulated NRs do not have the additional halo surrounding the track. This is further supported by the bottom panel of Fig. 12, which shows Migdal detection efficiencies as a function of E_{NR} for the subsets of pure and hybrid-simulated Migdal events with $E_{\text{ER}} \geq 5.0\,\text{keV}_{\text{ee}}$. The hybrid subset has roughly a factor of two decrease in ε_{det} between the lowest NR energy bins and higher energy ($\gtrsim 300\,\text{keV}_{\text{ee}}$) bins that is not present in the pure simulation sample. Evidently, the halos surrounding NR tracks are more extreme for higher energy NRs, further obscuring Migdal electrons. Besides the smaller halos present in low energy NRs, their topology also becomes more spherical, leading to ε_{det} becoming independent of the angle between the NR and ER track directions. This increase in the angular phase space for detection leads to the rise in ε_{det} at lower E_{NR} that is observed in both the hybrid and pure simulation datasets.

Integrating the hybrid Migdal distributions from the bottom panel of Fig. 12 over all NR energies, we find ε_{det} for the subset of hybrid Migdal events with ERs satisfying our ROI threshold ($E_{\text{ER}} \in \{5.0, 5.2, 5.4, 5.6, 5.8, 6.0\}\,\text{keV}_{\text{ee}}$) to be 0.29 and 0.35, respectively. Because these efficiencies were derived using a flat energy spectrum for ERs, we expect them to be overestimates since the predicted Migdal ER emission probabilities fall exponentially with ER energy [54]. More relevant, however, is the efficiency, ε_{det}, integrated over the hybrid Migdal NR spectrum and shown as a function of E_{ER} in the top panel of Fig. 12, as measurements expressed in this way can be used to compare with theory. YOLO's ability to detect ERs in Migdal events was integral in deriving these ER energy-dependent efficiencies, so it is natural to consider more broadly how well YOLO performs this task. We address this next where we describe a method to quantify YOLO's performance in detecting overlapping tracks.

3. Detection results: YOLO performance observables

To address YOLO's overlapping track identification performance more generally, we define a detectability criterion where we call a truth ER pixel in a simulated Migdal event significant if at least 1/3 of its intensity comes from the truth ER track.[5] We can then define n_{sigpix} as the number of significant pixels in an event. We define a related quantity, f_{sigpix}, as the fraction of ground truth ER pixels in an event that are significant

$$f_{\text{sigpix}} \equiv \frac{n_{\text{sigpix}}}{n_{\text{ERpix}}}, \tag{6}$$

where n_{ERpix} is the number of ground truth ER pixels. For convenience, we summarize definitions of n_{sigpix}, n_{ERpix}, and f_{sigpix} in Table IV.

Evaluating Migdal detection efficiency as a function of f_{sigpix} (or related quantities) is a more intrinsic measure of YOLO's performance. Our ROI for Migdal searches covers a broad dynamic range, with peak NR intensities orders of magnitude more intense than peak ER intensities, so in events where a Migdal ER is fully embedded within the NR penumbra, we expect n_{sigpix} to be nearly 0. While comparing YOLO's Migdal detection performance in pure simulation samples to hybrid simulation was useful in the previous study to show the adverse effects NR halos have on Migdal detection, the presence of halos should merely shift the distribution of n_{sigpix} closer to zero. We therefore opt to only consider hybrid simulation when evaluating ε_{det} versus f_{sigpix}. We evaluate this relationship using YOLO trained on the Augment training sample (Table III), as YOLO trained on this sample outperformed YOLO trained on the Base sample in Fig. 12.

Fig. 13 shows the results of this study. We find that 71.9% of hybrid Migdal candidates have $f_{\text{sigpix}} < 0.05$ and are essentially undetectable ($\varepsilon_{\text{det}} = 0.004$), with YOLO mostly identifying these events as single NRs. Beyond this, we observe hybrid Migdal detection efficiency steadily rising and eventually leveling off around 90% for $f_{\text{sigpix}} \geq 0.45$ (our sample statistics are small for $f_{\text{sigpix}} \geq 0.70$).

We previously found a Migdal detection efficiency of 0.35 for the subset of hybrid Migdal events with ER energies within our Migdal search ROI. ε_{det} nearly doubles to 0.67 for the subset of this sample that also satisfies

TABLE IV. Description of quantities related to our detectability criterion.

Quantity	Definition
n_{sigpix}	Number of truth ER pixels where at least 1/3 of the total pixel intensity comes from the ER.
n_{ERpix}	Total number of truth ER pixels.
f_{sigpix}	$n_{\text{sigpix}} / n_{\text{ERpix}}$

––––––

[5] To compute the fractional composition of intensity for a given pixel, we use the truth ER's intensity after applying gain scaling, vignetting scaling, and noise. We do not perform any processing to the NR intensities since they come from measurement.
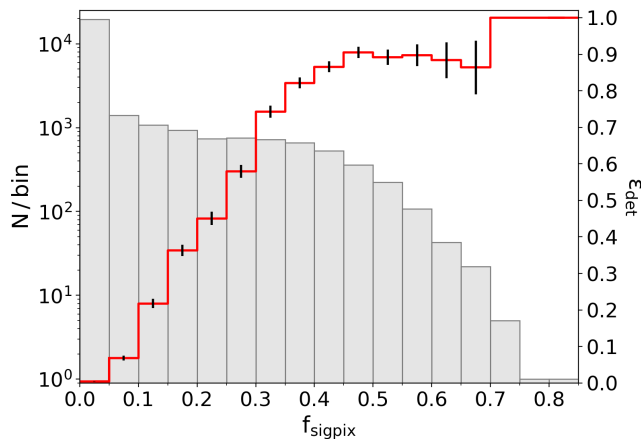
FIG. 13. Distribution of f_{sigpix} (gray bars; left vertical axis) and Migdal detection efficiency versus f_{sigpix} (red histogram; right vertical axis) of the hybrid Migdal simulation set trained on the Augment training sample.

$f_{sigpix} > 0.1$, demonstrating that YOLO is capable of detecting the majority of Migdal events within our search ROI provided that at least 10% of ER pixels are significant. This demonstrates the significance of mitigating optical halo formation and, more broadly, reducing diffusion through the usage of a negative ion drift gas mixture, as reducing diffusion should significantly reduce the likelihood of a Migdal event having zero significant pixels.

4. *Localization performance*

Our results indicate that YOLO performs very well for detecting overlapping ER and NR tracks in the hybrid Migdal sample, provided there are enough significant ER pixels. In terms of localization, our definition of a positive detection only tells us that $IoU(B_p^{NR}, B_t^{NR})$ and $IoU(B_p^{ER}, B_t^{ER})$ are each greater than 0. To quantify localization performance, we therefore plot these IoU distributions for positive detections in both the hybrid Migdal and hybrid coincidence samples, which are shown in black and red in Fig. 14, respectively. Overall, we observe excellent localization performance for NRs with mean IoUs between prediction and ground truth of 0.92 and 0.95 for positive detections in the hybrid Migdal and hybrid coincidence samples, respectively. ER localization performance is also very good, with a mean IoU of 0.61 for hybrid coincidences. Due to the larger variance in ER trajectory and lower pixel intensities compared to NRs, it is not surprising that ERs are not localized as well as NRs in general. The fact that there is a significant drop in the localization performance in the Migdal sample when compared to the coincidence sample is also not a surprise, as a significant fraction of ERs are obscured by the NR in the former sample.
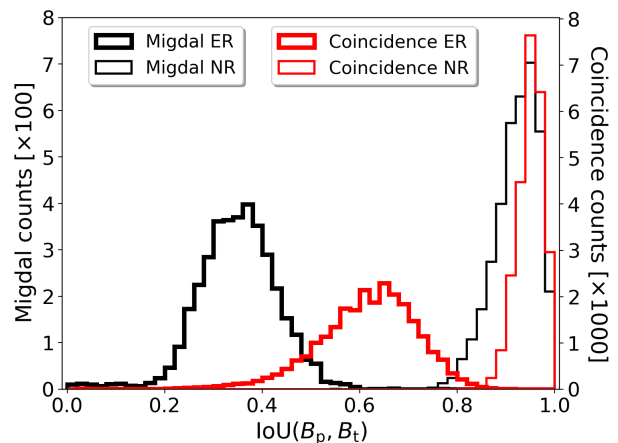

FIG. 14. Comparison of bounding box overlap between ground truth and YOLO's predictions for positive detections in the hybrid Migdal (black) and hybrid coincidence simulation (red) samples. The bolded curves represent ER IoU overlap ($IoU(B_p^{ER}, B_t^{ER})$), while the unbold curves represent NR IoU overlap ($IoU(B_p^{NR}, B_t^{NR})$).

5. *Background rejection study*

We next perform an optimization study on coincidence background rejection and Migdal detection efficiency. Since ERs and NRs share a vertex in Migdal events, their bounding boxes will be in close proximity to one another, whereas for coincidence backgrounds, the two bounding boxes should be randomly distributed. We can therefore use the Euclidean distance between the centroids of the ER and NR bounding boxes, $d(\bar{b}_{ER}, \bar{b}_{NR})$, to maximize the signal to background. The background sample for this study consists of the roughly 25000 coincidence frames with positive detections from YOLO trained on the Augment sample. The signal sample is our entire hybrid Migdal sample.

We evaluate our background rejection and signal efficiency at six different thresholds of f_{sigpix}. When referring specifically to a threshold value of f_{sigpix}, we denote it as $\widehat{f}_{sigpix}$. For a given $\widehat{f}_{sigpix}$ threshold, we evaluate the signal efficiency $\varepsilon_{sig}(d_{step}, \widehat{f}_{sigpix})$ at maximum centroid distance steps $d_{step} \in \{85, 84, \ldots, 3, 2\}$ mm. With this notation established, the efficiencies shown in Fig. 15 are evaluated as

$$\varepsilon_{sig} = \frac{N_{det}(f_{sigpix} \geq \widehat{f}_{sigpix} \wedge d \leq d_{step})}{N(f_{sigpix} \geq \widehat{f}_{sigpix} \wedge d \leq d_{step})}. \qquad (7)$$

From this figure, it is evident that with suitable choice of maximum centroid distance between the ER and NR bounding boxes, we can reject orders of magnitude of background candidates while retaining nearly all detected signal candidates. We optimize this maximum centroid distance selection to maximize the product of signal retention to background rejection. Considering the
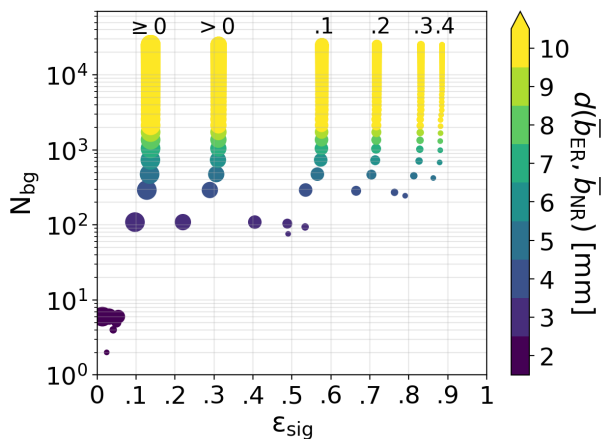
FIG. 15. Surviving coincidence-background frame counts, N_{bg}, versus Migdal signal efficiency as a function of distance between the bounding box centroid of YOLO's ER and NR prediction, $d(\bar{b}_{ER}, \bar{b}_{NR})$, evaluated on around 25000 detected hybrid coincidence frames and about 27000 hybrid Migdal frames. The number above the each curve represents the threshold fraction of significant ER pixels, $\hat{f}_{sigpix}$, for that curve. For example, the leftmost curve represents $f_{sigpix} \geq 0$ and the next curve represents $f_{sigpix} > 0$. For all other curves we use selections greater than or equal to the number shown, so the rightmost curve represents $f_{sigpix} \geq 0.4$. Point sizes are different between the six curves for visual clarity.

$\hat{f}_{sigpix} \geq 0$ slice, we find the optimal integral centroid distance to be $d_{step} = 6$ mm. Selecting frames satisfying $d \leq 6$ mm retains 99.4% of detected Migdal events while rejecting 97.1% of coincident backgrounds. Expressed in terms of raw counts, this selection reduces coincident background frames from 25154 to 735, while retaining 3709 of the 3731 signal frames.

We have found that using the distance between the centroids of YOLO's predicted bounding box leads to excellent coincidence background rejection with essentially no loss in detected signal. What remains to be tested is YOLO's false positive identification rate. We test this by evaluating our pipeline (with YOLO trained on the Augment sample) on a sample of 5000 randomly selected frames containing a single simulated NR satisfying $E_{NR} \geq 60$ keV$_{ee}$ and nothing else. The simulated frames are processed following the procedure detailed in Section IV A. Of these 5000 frames, YOLO falsely identifies 7 NRs as an ER-NR pair, with the remaining 4993 correctly identified as a single NR, leading to a false positive rate of 0.14%. While these false positive events could pass YOLO's initial Migdal search and be flagged as candidates in an online analysis, the combination of more detailed later stage offline image analyses and analyses with other subsystems should rule out these false positives.

V. REAL-TIME ANALYSIS AND MIGDAL SEARCHES

Our earlier benchmark showed that our pipeline is capable of achieving real-time speeds when performing an OQC end-to-end Migdal effect search on the MIGDAL readout PC. With this established, we begin this section by highlighting in more detail the online deliverables provided and displayed in real-time by our pipeline. We then conclude this section with the application of the steps of our online Migdal search to a very large sample of OQC data to illustrate the scale of the data reduction achieved when performing a Migdal effect search.

A. Online deliverables

Recall that our pipeline converts image-indexed data to track-indexed data using YOLO's bounding box assignments, allowing for efficient extraction of physics information on a track-by-track basis. For each track-index, we compute the following quantities online using the pixel content within the bounding box evaluated on 4×4-binned, Gaussian smoothed images:

1. Intensity/energy: Sum up all pixels above threshold to compute the intensity of the track. We correct the intensity for vignetting (Appendix A) and use appropriate ^{55}Fe calibration-run data to convert the corrected intensity into energy in units of keV$_{ee}$.

2. 2D length: Use a singular value decomposition [73] to identify the track's principal axis in the OQC readout plane and compute the length of the track along that axis. Generally speaking, this method works better for NRs than ERs since NR topologies are typically better modeled by straight lines than ER topologies.

3. Head/tail identification (for ERs and NRs only): We split the track in half along its principal axis and count the intensity in each half of the track to identify the track's head and tail [74]. If YOLO identified the track as a NR (ER), we assign the side of the track with less intensity as the head (tail). Since this method is reliant on principal axis identification, it works best in cases where track trajectories are modeled well by straight lines.

4. 2D vector angle (ERs and NRs only): Use the assigned head/tail direction to compute the angle of the track's principal vector with the $+x$ axis.

5. 2D axial angle (protons and alphas only): Compute the angle between the track's principal axis with the $+x$ axis.

While we could substantially improve our 2D track fitting by training YOLO to identify key points along
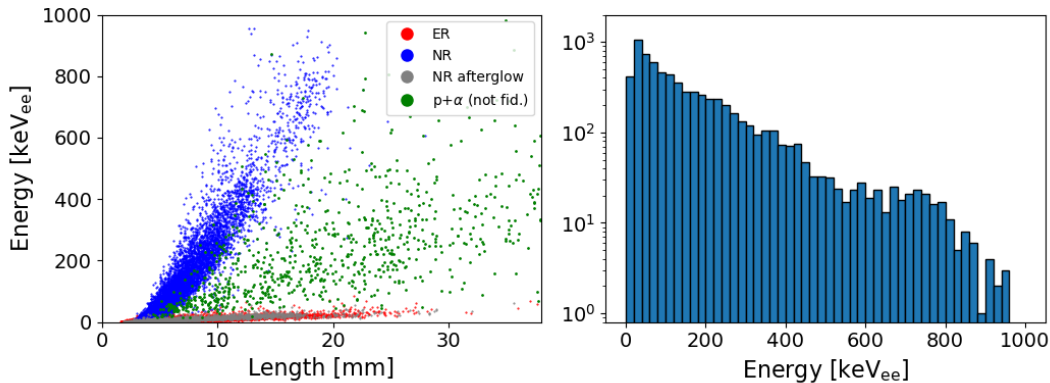
FIG. 16. Snapshot of a live online display that updates with 600 images worth of data every five seconds. The top banner shows updating counters of objects of interest accumulated over the course of the run. The left plot shows energy versus 2D length of tracks with color representing YOLO's classification assignments of track species of interest. The right plot shows the energy spectrum of NRs that YOLO identified.

a track's trajectory (see Fig. 1), which would in turn improve head/tail identification, angular reconstruction performance, and likely still be implementable in real-time online analyses, this is beyond the scope of this work. Improving track fitting has no-bearing on the coincidence background rejection steps described in Section IV C 5, and is therefore not necessary for our pipeline's OQC-driven Migdal effect search, so we leave this for future work.

Some of the physical track quantities extracted by our pipeline are implemented into live online displays showing relevant event statistics during D-D runs. Fig. 16 shows an example of such a display. The event counters and both plots refresh with new processed data from three image batches (600 frames) every five seconds. We operate this display on our readout PC during data acquisition, providing us with mixed-field particle identification, dE/dx distributions of select classes of events, NR energy spectra, and event-class counts that include candidates satisfying user-input Migdal search criteria, all at real-time rates (120 fps) with at most a five second lag behind image acquisition. To calibrate NR energies online, we use the peak of a Gaussian fit to the most-recently recorded ^{55}Fe spectrum (at the same nominal voltage across both GEMs). These ^{55}Fe calibration fits are performed on online-processed data.

The count of four candidates shown in the banner of Fig. 16 represent ER-NR pairs satisfying Migdal search criteria of $E_{NR} \geq 60\,keV_{ee}$ and $d(\bar{b}_{ER}, \bar{b}_{NR}) \leq 6\,mm$. Recall that each time a candidate is recorded, a separate file containing the ER-NR pair-specific information and associated metadata is saved, allowing for immediate analysis of candidate-events.

In addition to performing Migdal candidate searches, our pipeline's online reporting is used alongside ITO detector data to provide immediate feedback on effective gains during ^{55}Fe calibration runs. At a constant voltage

across the GEMs, the presence of D-D neutrons causes a degradation in effective gain over the course of a day [56], so this feedback is crucial for making suitable adjustments to GEM voltages that ensure reasonably stable long-term gain. Online-processed data has also been useful for performing data-informed D-D generator and collimator alignments. All of these examples are made possible by the level of detail of the information extracted by our pipeline coupled with YOLO's real-time inference speed.

B. An offline search for the Migdal effect

Our online analysis identifies candidates with 2D topologies that are consistent with the Migdal effect in real-time over the course of a run. It is therefore only necessary to perform an offline search for the Migdal effect if we want to integrate other detector subsystems with the optical system. That being said, it is illustrative to highlight the scale of the data reduction achieved

TABLE V. Summary showing the number of camera frames remaining after increasingly restrictive selections throughout a Migdal candidate search over frames with 8.33 ms exposure. Each row in the table incorporates the selections from all previous rows. Employing a $60\,keV_{ee}$ NR energy restriction and a 6 mm maximum separation distance between ER and NR bounding box centroids reduces our sample from about 20 million frames to around 1000 frames.

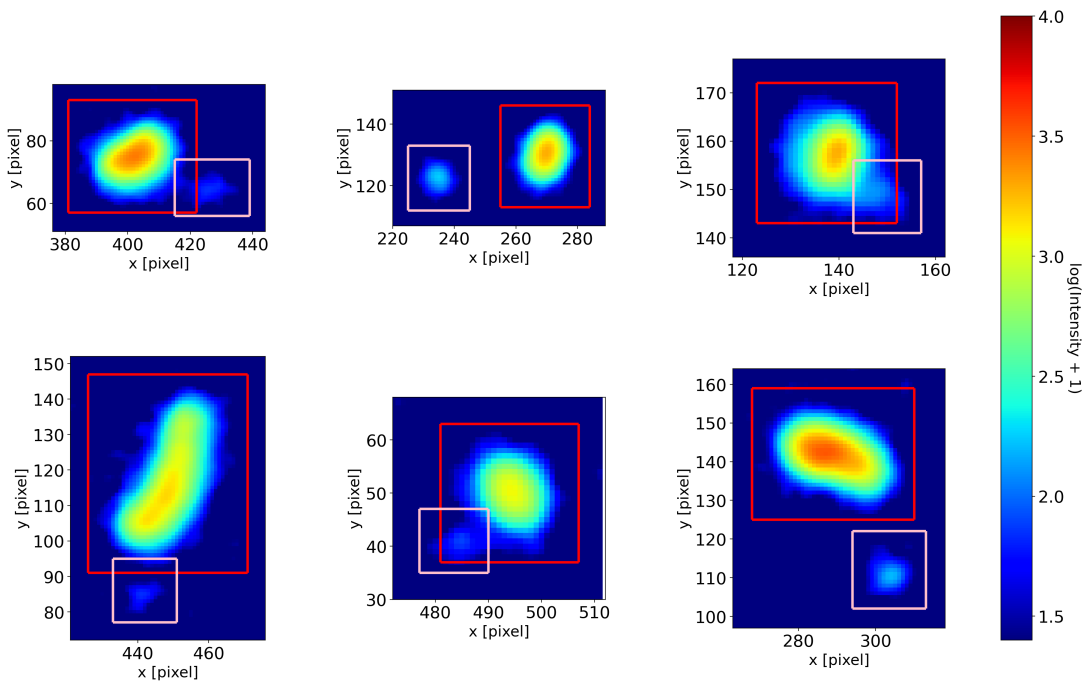| Selection	N_{frames}
None (all frames analyzed by YOLO)	19996200
At least one fiducial ER-NR pair	25105
$E_{NR} \geq 60\,keV_{ee}$	15121
$d(\bar{b}_{ER}, \bar{b}_{NR}) \leq 6\,mm$	1130

FIG. 17. Event displays with YOLO's predicted bounding boxes for six randomly selected frames in our Migdal search sample drawn from events satisfying all criteria applied in Table V. The frames are 4×4 binned with Gaussian smoothing applied, leading to a 3.1 mm per twenty pixel conversion. NR bounding boxes are colored red and ER bounding boxes are colored pink. Clockwise starting from the upper left frame, the reconstructed energies $\{E_{\mathrm{NR}}, E_{\mathrm{ER}}\}$, in units of keV$_{\mathrm{ee}}$, within YOLO's predicted bounding boxes are (i) $\{140, 5.8\}$, (ii) $\{61, 4.9\}$, (iii) $\{64, 5.6\}$, (iv) $\{170, 4.3\}$, (v) $\{82, 5.4\}$, and (vi) $\{290, 4.1\}$.

when applying our online Migdal search criteria to a large sample of OQC images, so as an exercise, we apply these search criteria to an offline sample consisting of nearly 20 million images (90 TB raw, uncompressed images) recorded at the peak 120 fps acquisition rate of the camera. Evaluating this sample using YOLO trained on the Base training set (Table I) and applying the search criteria of $E_{\mathrm{NR}} \geq 60 \, \mathrm{keV_{ee}}$ and $d(\bar{b}_{\mathrm{ER}}, \bar{b}_{\mathrm{NR}}) \leq 6 \, \mathrm{mm}$ to frames containing at least one fiducial ER-NR pair (Table V) ultimately reduces the sample from 20 million frames to 1130 frames.

Walking through Table V step-by-step, the selections listed in each row are applied successively. We first select frames containing at least one fiducial ER-NR pair satisfying (1) the ER is not flagged to be a NR ghost, (2) the NR is not clipped by the rolling shutter, and (3) the IoU overlap between the ER and NR is not greater than 0.95.[6] Applying this selection reduces our sample from 20 million frames to around 25000 frames. In each of these 25000 frames, we next link all unique ER-NR pairs together so we can conveniently make selections on an

ER-NR track-pair basis. We then apply a $60 \, \mathrm{keV_{ee}}$ NR threshold to each track-pair, further reducing the sample to around 15000 frames. In Section IV C 5 we determined $d(\bar{b}_{\mathrm{ER}}, \bar{b}_{\mathrm{NR}}) \leq 6 \, \mathrm{mm}$ to be the optimal selection for simultaneously reducing coincidence backgrounds with minimal effect on signal acceptance. Applying this criterion to the remaining 15121 frames leaves us with 1130 frames.

Fig. 17 shows six randomly selected frames satisfying all selections in Table V. The orientations and separations of the identified ERs and NRs in frames (i), (iii), (v), and (vi) (labeled clockwise starting from the upper left frame) suggest that the events shown in these frames are consistent with a 2D Migdal effect topology hypothesis. As a reminder, the OQC alone is not sufficient for confirming the Migdal effect, so these events need to be further analyzed with the remaining detector subsystems to test whether or not they are consistent with the Migdal effect in 3D. Still, this figure illustrates that applying our search criteria on real data extracts the classes of events we are interested in. Moreover, on simulation we found the $d(\bar{b}_{\mathrm{ER}}, \bar{b}_{\mathrm{NR}}) \leq 6 \, \mathrm{mm}$ selection rejected 97.1% of coincidence backgrounds while retaining 99.4% of the detected signal, making it unlikely that these selections rejected any Migdal candidates identified by YOLO in the original 20 million frames. Through this exercise, we have demonstrated that applying our online search criteria on a sample of 20 million frames reduces the sample to 1130 frames with simulation suggesting essentially no

---

[6] The IoU ≤ 0.95 restriction is to ensure that YOLO did not draw two nearly identical bounding boxes around the same track. This happens occasionally for low energy NRs where YOLO makes two plausible predictions; one being an ER and the other being a NR with nearly identical bounding boxes.

loss of detectable signal. This reduction transforms our rare event search into a much more manageable search, enabling detailed analyses of 2D candidates in all detector subsystems.

VI. SUMMARY

In summary, we have developed a YOLOv8-based image processing pipeline capable of performing a rare event search for the Migdal effect, in real-time, using 2D Orca-Quest qCMOS camera images recorded by the MIGDAL experiment. By reframing a Migdal effect search as one for pairs of ERs and NRs in sufficiently close proximity, we were able to train YOLO on an abundance of measured data (Table I), rather than exclusively on a simulated rare-event signal. Training with a combination of real data and simulation and evaluating YOLO's performance on simulated Migdal effect events formed by stitching together measured NRs with simulated ERs, we derived a detection efficiency of 35% for events with ER energies between 5 keV and 6 keV. While this efficiency appears low, more than 25% of events with ER energies falling in this range have essentially zero significant ER pixels (see Section IV C 3 for the definition of a significant pixel). This efficiency nearly doubles to 67% when considering events in this sample where at least 10% of the truth ER pixels are significant. Regardless of ER energy, YOLO on average achieves detection efficiencies above 80% for simulated Migdal events where at least 1/3 of truth ER pixels are significant. These results demonstrate the importance of mitigating optical halo formation around NRs in our detector, as these halos can rapidly envelop Migdal ERs, rendering them much more difficult to detect. More broadly, mitigating diffusion with negative ion drift gas mixtures also has the potential to significantly improve detection efficiencies.

In the Orca-Quest, accidental coincidence backgrounds – where independent ERs and NRs that are not consistent with the Migdal effect spatially coincide within the exposure time of a frame – are the dominant source of ER-NR pairs occupying a single frame. On simulated frames containing exactly one ER-NR pair, we found that selecting frames where the Euclidean distance between the centroids of the ER and NR bounding boxes, $d(\bar{b}_{\mathrm{ER}}, \bar{b}_{\mathrm{NR}})$, was within 6 mm, is optimal for coincidence background rejection. This selection led to a rejection of 97.1% of coincidence backgrounds with a 99.4% acceptance of Migdal events detected by YOLO. We found that applying this selection and a 60 keV$_{\mathrm{ee}}$ NR energy threshold to a large sample of 20 million measured Orca-Quest camera frames, reduced that sample to 1130 frames, thereby transforming our rare event search into a significantly more contained search.

Benchmarking the end-to-end processing speed of our pipeline, we found it to exceed the peak 120 fps acquisition rate of the Orca-Quest when processing typical data for our experiment, and additionally found it to analyze tracks at rates near 450 Hz on a sample of frames with high track occupancy. We also demonstrated the pipeline's real-time processing speed to allow us to perform the full camera analysis chain online while acquiring data. The ability to read out data at the maximum frame rate of a state-of-the-art CMOS camera system while simultaneously processing and analyzing it in real time on a single PC was made possible through the use of a fast object detection algorithm like YOLOv8. The fast inference speed of YOLO, coupled with its robust performance in challenging scenarios like those involving detection of overlapping tracks, makes pipelines such as ours highly attractive for other applications and experiments that require efficient object detection in data streams at rates of $\mathcal{O}(1\,\mathrm{gigapixel/s})$.
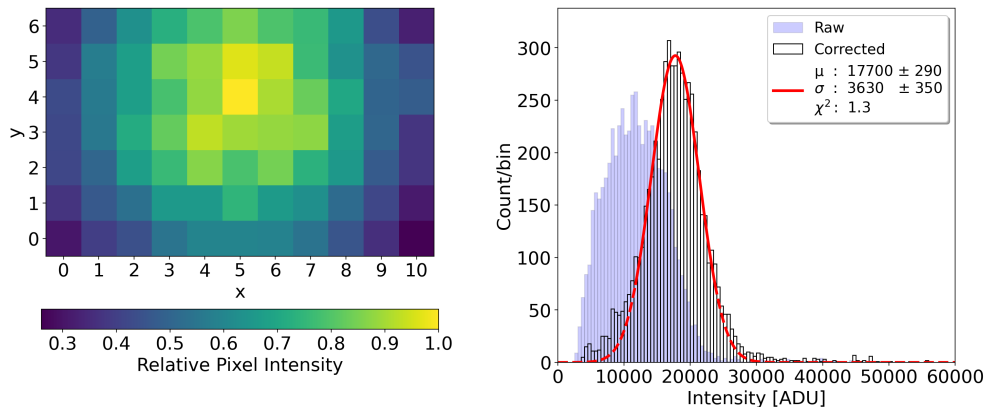
FIG. 18. Left: Effective intensity map generated from ^{55}Fe X-rays. Right: ^{55}Fe spectrum before (pale blue bars) and after (black bars with fit) applying the vignetting correction map.

Appendix A: Vignetting corrections and energy calibrations

In the presence of mono-energetic events, observed intensities in the optical readout diminish due to vignetting as we move radially away from the center of the optical axis. While K-α and K-β modes both contribute to the ^{55}Fe X-ray spectra, the 5.9 keV K-α mode contributes to nearly 90% of detectable X-ray emissions, so we treat the spectrum as having a mono-energetic 5.9 keV peak and log the intensities of ^{55}Fe X-rays in regions across the optical readout to generate an effective intensity map across the camera field. We use this map to flatfield the intensity spectrum across the readout, which is what we call vignetting corrections.

To create such a map, we discretize the camera plane into an 11×7 grid of logical pixels and perform the following procedure at the end of each ^{55}Fe run:

1. Select only fiducial ^{55}Fe tracks.

2. Let P_{ij} : $i \in [0, 10]$, $j \in [0, 6]$ represent a logical pixel in our 11×7 grid. For each ^{55}Fe track, we count the number of pixels in the track falling inside each logical pixel. We assign the track to the logical pixel, P_{mn} that contains more of the ^{55}Fe track's pixels than any other P_{ij}.

3. For each logical pixel, compute the mean intensity, I_{ij} of the subset of ^{55}Fe tracks that are *fully* con-

tained in P_{ij}. Normalize to the logical pixel with the highest mean intensity.

The left panel of Fig. 18 shows the effective intensity map generated using this procedure on an ^{55}Fe run from Science Run 1. For a given ^{55}Fe track with raw intensity I, that was assigned to logical pixel P_{mn}, its vignetting-corrected intensity I_{cor} is given by

$$I_{\text{cor}} = \frac{I_{mn}}{\max_{ij} I_{ij}} I. \tag{A1}$$

The black outlined histogram in the right panel of Fig. 18 shows I_{cor} for each fiducial ^{55}Fe track in the run. Contrast this to the raw intensity spectrum shown in pale blue. We calibrate energy by assigning 5.9 keV to the value of I_{cor} corresponding to the peak of the single-peak Gaussian fit to the I_{cor} spectrum.

We generated effective intensity maps for each ^{55}Fe run recorded during Science Runs 1 and 2 and performed vignetting corrections when computing the energies for all tracks during D-D runs. We found that at a given voltage across the GEMs, the ^{55}Fe peak varies over the course of the day. Thus we need to choose which ^{55}Fe run to use when assigning energy calibrations and vignetting corrections for D-D runs. The D-D runs analyzed in this article use calibration information (including the effective intensity map) from the nearest-in-time ^{55}Fe calibration run that had the same voltage across the GEMs as the given D-D run. For the given D-D run with associated ^{55}Fe calibration information, we compute the energy of each track in the D-D run using the following procedure:

1. Assign to the track the logical pixel, P_{mn} that contains more of the track's pixels than any other logical pixel.

2. Use Eq. (A1) to compute I_{cor}.

3. Let I'_{cor} be the vignetting-corrected intensity of the ^{55}Fe calibration associated with the current D-D

run. The energy of the track is computed as

$$E = \frac{5.9\,\text{keV}}{I'_\text{cor}} I_\text{cor}. \tag{A2}$$

We acknowledge that the following improvements could be made to improve energy estimates during D-D runs

1. Effective gain tends to decay with increasing D-D exposure at a given voltage across the GEMs. This means that interpolating the ^{55}Fe peak intensities versus time to assign ^{55}Fe peak intensities at all times during a D-D run would be advantageous compared to our current approach of assigning the nearest-in-time ^{55}Fe calibration performed with the same voltage across the GEMs as the D-D run.

2. Rather than assigning logical pixels on a track-by-track basis to create effective intensity maps, we could instead assign a logical pixel to each pixel in a track and perform vignetting corrections for each individual pixel. This change may substantially improve energy estimates for long tracks (especially alphas and protons) that span many logical pixels.

While our current energy calibration procedure provides sufficient energy resolutions for resolving the fluorine and carbon endpoints from 2.5 MeV neutrons produced from the D-D-generator (Fig. 16), these further improvements to our calibrations may sharpen these endpoints in the NR energy spectrum.

Appendix B: YOLOv8 loss function

In Section III A we briefly introduced the three terms composing YOLOv8's loss function. Here, we go through the functional form in more detail.

For a given ground truth bounding box, B_t, with centroid, $\bar{b}_\text{t}$, and associated class label, y_t, we can write YOLOv8's loss function, $\mathcal{L}$, for a predicted bounding box, B_p, with centroid, $\bar{b}_\text{p}$, and class prediction, y_p, as

$$\mathcal{L} = \omega_\text{CIoU}\ell_\text{CIoU} + \omega_\text{DFL}\ell_\text{DFL} + \omega_\text{cls}\ell_\text{cls}, \tag{B1}$$

where ω_CIoU, ω_DFL, and ω_cls are weights associated with each loss term. We use the default values for these weights present in the Ultralytics YOLOv8 package, which as of this writing are 7.5, 1.5, and 0.5, respectively, however we could treat these three weights as hyperparameters. The loss terms can be written as

$$\ell_\text{CIoU} = 1 - \text{IoU}(B_\text{p}, B_\text{t}) - \frac{\left|\bar{b}_\text{t} - \bar{b}_\text{p}\right|^2}{d^2} + \alpha\nu \tag{B2}$$

$$\ell_\text{DFL} = -(\text{IoU}(B_R, B_\text{t}) - \text{IoU}(B_\text{p}, B_\text{t}))\log(p_L)$$
$$+ (\text{IoU}(B_\text{p}, B_\text{t}) - \text{IoU}(B_L, B_\text{t}))\log(p_R) \tag{B3}$$

$$-\ell_\text{cls} = \sum_{c \in \text{cls}} \left(\sigma(y_{\text{p},c})\log(y_{\text{t},c}) + (1 - y_{\text{t},c})\log(1 - \sigma(y_{\text{p},c}))\right), \tag{B4}$$

where

$$\alpha \equiv \frac{\nu}{1 + \text{IoU}(B_\text{p}, B_\text{t}) + \nu} \tag{B5}$$

$$\nu \equiv \frac{4}{\pi^2}\left(\arctan\left(\frac{w_\text{t}}{h_\text{t}}\right) - \arctan\left(\frac{w_\text{p}}{h_\text{p}}\right)\right). \tag{B6}$$

Here, d is the diagonal distance of the smallest box fully enclosing B_p and B_t, w_j and h_j are the width and height of bounding box B_j, B_L and B_R, are the left and right nearest bounding box hypotheses to B_p, and $\sigma(\cdot)$ is the sigmoid function. Finally, p_L and p_R represent the softmax predictions (probabilities) for the locations of B_L and B_R. The global minimum of this loss function is 0 and corresponds to exact agreement between the predicted bounding box and classification with its ground truth counterpart.

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks, in *Advances in Neural Information Processing Systems*, Vol. 25, edited by F. Pereira, C. Burges, L. Bottou, and K. Weinberger (Curran Associates, Inc., 2012).

[2] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, ImageNet: A large-scale hierarchical image database, in *2009 IEEE Conference on Computer Vision and Pattern Recognition* (2009) pp. 248–255.

[3] A. Aurisano, A. Radovic, D. Rocco, A. Himmel, M. Messier, E. Niner, G. Pawloski, F. Psihas, A. Sousa, and P. Vahle, A convolutional neural network neutrino event classifier, Journal of Instrumentation **11** (09), P09001.

[4] P. Baldi, J. Bian, L. Hertel, and L. Li, Improved Energy Reconstruction in NOvA with Regression Convolutional Neural Networks, Phys. Rev. D **99**, 012011 (2019),

arXiv:1811.04557 [physics.ins-det].

[5] D. Belayneh *et al.*, Calorimetry with deep learning: particle simulation and reconstruction for collider physics, Eur. Phys. J. C **80**, 688 (2020), arXiv:1912.06794 [physics.ins-det].

[6] R. Girshick, J. Donahue, T. Darrell, and J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2014) pp. 580–587.

[7] S. Ren, K. He, R. Girshick, and J. Sun, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, arXiv e-prints , arXiv:1506.01497 (2015), arXiv:1506.01497 [cs.CV].

[8] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, Ssd: Single shot multibox detector, in *Computer Vision–ECCV 2016: 14th European*

Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14 (Springer, 2016) pp. 21–37.

[9] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, Focal loss for dense object detection, in *Proceedings of the IEEE international conference on computer vision* (2017) pp. 2980–2988.

[10] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, End-to-end object detection with transformers, in *European conference on computer vision* (Springer, 2020) pp. 213–229.

[11] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, You only look once: Unified, real-time object detection, in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016) pp. 779–788.

[12] J. Terven and D. Cordova-Esparza, A comprehensive review of YOLO: From YOLOv1 to YOLOv8 and beyond, arXiv preprint arXiv:2304.00501 (2023).

[13] G. Jocher, A. Chaurasia, and J. Qiu, Ultralytics YOLOv8 (2023).

[14] Y. Ghasemi, H. Jeong, S. H. Choi, K.-B. Park, and J. Y. Lee, Deep learning-based object detection in augmented reality: A systematic review, Computers in Industry **139**, 103661 (2022).

[15] A. Sarda, S. Dixit, and A. Bhan, Object Detection for Autonomous Driving using YOLO algorithm, in *2021 2nd International Conference on Intelligent Engineering and Management (ICIEM)* (2021) pp. 447–451.

[16] R. Yang and Y. Yu, Artificial Convolutional Neural Network in Object Detection and Semantic Segmentation for Medical Imaging Analysis, Frontiers in Oncology **11**, 638182 (2021).

[17] Z. Wu, X. Chen, Y. Gao, and Y. Li, Rapid Target Detection in High Resolution Remote Sensing Images Using Yolo Model, ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences **42.3**, 1915 (2018).

[18] H. H. Nguyen, T. N. Ta, N. C. Nguyen, V. T. Bui, H. M. Pham, and D. M. Nguyen, YOLO Based Real-Time Human Detection for Smart Video Surveillance at the Edge, in *2020 IEEE Eighth International Conference on Communications and Electronics (ICCE)* (2021) pp. 439–444.

[19] R. González, R. Muñoz, and C. Hernández, Galaxy detection and identification using deep learning and data augmentation, Astronomy and Computing **25**, 103 (2018).

[20] D. Cornu, P. Salomé, B. Semelin, A. Marchal, J. Freundlich, S. Aicardi, X. Lu, G. Sainton, F. Mertens, F. Combes, and C. Tasse, YOLO-CIANNA: Galaxy detection with deep learning in radio data. I. A new YOLO-inspired source detection method applied to the SKAO SDC1, arXiv e-prints , arXiv:2402.05925 (2024), arXiv:2402.05925 [astro-ph.IM].

[21] A. Bonaldi *et al.*, Square Kilometre Array Science Data Challenge 1: analysis and results, Mon. Not. Roy. Astron. Soc. **500**, 3821 (2020), arXiv:2009.13346 [astro-ph.IM].

[22] R. Acciarri *et al.* (MicroBooNE), Design and Construction of the MicroBooNE Detector, JINST **12** (02), P02017, arXiv:1612.05824 [physics.ins-det].

[23] R. Acciarri *et al.* (MicroBooNE), Convolutional Neural Networks Applied to Neutrino Events in a Liquid Argon Time Projection Chamber, JINST **12** (03), P03011, arXiv:1611.05531 [physics.ins-det].

[24] P. Abratenko *et al.* (MicroBooNE), Semantic segmentation with a sparse convolutional neural network for event reconstruction in MicroBooNE, Phys. Rev. D **103**, 052012 (2021), arXiv:2012.08513 [physics.ins-det].

[25] P. Abratenko *et al.* (MicroBooNE), Convolutional neural network for multiple particle identification in the MicroBooNE liquid argon time projection chamber, Phys. Rev. D **103**, 092003 (2021), arXiv:2010.08653 [hep-ex].

[26] P. Abratenko *et al.* (MicroBooNE), Cosmic ray muon clustering for the MicroBooNE liquid argon time projection chamber using sMask-RCNN, JINST **17** (09), P09015, arXiv:2201.05705 [hep-ex].

[27] P. Abratenko *et al.* (MicroBooNE), Search for an anomalous excess of charged-current quasielastic νe interactions with the MicroBooNE experiment using Deep-Learning-based reconstruction, Phys. Rev. D **105**, 112003 (2022), arXiv:2110.14080 [hep-ex].

[28] A. A. Aguilar-Arevalo *et al.* (MiniBooNE Collaboration), Updated MiniBooNE neutrino oscillation results with increased data and new background studies, Phys. Rev. D **103**, 052002 (2021).

[29] Greer, Joel Thomas, *Neutrino Interaction Identification for the DUNE Trigger*, Ph.D. thesis (2022).

[30] J. Kieseler, Object condensation: one-stage grid-free multi-object reconstruction in physics detectors, graph and image data, Eur. Phys. J. C **80**, 886 (2020), arXiv:2002.03605 [physics.data-an].

[31] H. M. Araújo *et al.*, The MIGDAL experiment: Measuring a rare atomic process to aid the search for dark matter, Astropart. Phys. **151**, 102853 (2023), arXiv:2207.08284 [hep-ex].

[32] A. Migdal, Ionizatsiya atomov pri yadernykh reaktsiyakh (Ionization of atoms in nuclear reactions), ZhETF **9**, 1163 (1939).

[33] A. Migdal, *Qualitative Methods In Quantum Theory*, 1st ed. (CRC Press, 1977).

[34] *ORCA-Quest qCMOS camera C15550-20UP Technical note*, Tech. Rep. (Hamamatsu Photonics, 2022).

[35] Morales-Curiel *et al.*, Volumetric imaging of fast cellular dynamics with deep learning enhanced bioluminescence microscopy, Communications Biology **5** (2022).

[36] O. Wolley, T. Gregory, S. Beer, T. Higuchi, and M. Padgett, Quantum imaging with a photon counting camera, Scientific Reports **12**, 8286 (2022).

[37] F. D. Amaro *et al.*, The CYGNO/INITIUM experiment, SciPost Phys. Proc. , 019 (2023).

[38] I. A. Strakhov, B. S. Safonov, and D. V. Cheryasov, Speckle Interferometry with CMOS Detector, Astrophysical Bulletin **78**, 234 (2023), arXiv:2305.00451 [astro-ph.IM].

[39] M. Ibe, W. Nakano, Y. Shoji, and K. Suzuki, Migdal Effect in Dark Matter Direct Detection Experiments, JHEP **03**, 194, arXiv:1707.07258 [hep-ph].

[40] D. S. Akerib *et al.* (LUX), Results of a Search for Sub-GeV Dark Matter Using 2013 LUX Data, Phys. Rev. Lett. **122**, 131301 (2019), arXiv:1811.11241 [astro-ph.CO].

[41] E. Armengaud *et al.* (EDELWEISS), Searching for low-mass dark matter particles with a massive Ge bolometer operated above-ground, Phys. Rev. D **99**, 082003 (2019), arXiv:1901.03588 [astro-ph.GA].

[42] Z. Z. Liu *et al.* (CDEX), Constraints on Spin-Independent Nucleus Scattering with sub-GeV Weakly Interacting Massive Particle Dark Matter from the CDEX-1B Experiment at the China Jinping Underground Laboratory, Phys. Rev. Lett. **123**, 161301 (2019),

arXiv:1905.00354 [hep-ex].

[43] E. Aprile *et al.* (XENON), Search for Light Dark Matter Interactions Enhanced by the Migdal Effect or Bremsstrahlung in XENON1T, Phys. Rev. Lett. **123**, 241803 (2019), arXiv:1907.12771 [hep-ex].

[44] L. Barak *et al.* (SENSEI), SENSEI: Direct-Detection Results on sub-GeV Dark Matter from a New Skipper-CCD, Phys. Rev. Lett. **125**, 171802 (2020), arXiv:2004.11378 [astro-ph.CO].

[45] G. Adhikari *et al.* (COSINE-100), Searching for low-mass dark matter via the Migdal effect in COSINE-100, Phys. Rev. D **105**, 042006 (2022), arXiv:2110.05806 [hep-ex].

[46] E. Armengaud *et al.* (EDELWEISS), Search for sub-GeV dark matter via the Migdal effect with an EDELWEISS germanium detector with NbSi transition-edge sensors, Phys. Rev. D **106**, 062004 (2022), arXiv:2203.03993 [astro-ph.GA].

[47] P. Agnes *et al.* (DarkSide), Search for Dark-Matter–Nucleon Interactions via Migdal Effect with DarkSide-50, Phys. Rev. Lett. **130**, 101001 (2023), arXiv:2207.11967 [hep-ex].

[48] K. Abe *et al.* (XMASS), Direct dark matter searches with the full data set of XMASS-I, Phys. Rev. D **108**, 083022 (2023), arXiv:2211.06204 [astro-ph.CO].

[49] J. Aalbers *et al.* (LZ), Search for new physics in low-energy electron recoils from the first LZ exposure, Phys. Rev. D **108**, 072006 (2023), arXiv:2307.15753 [hep-ex].

[50] D. Huang *et al.* (PandaX), Search for Dark-Matter–Nucleon Interactions with a Dark Mediator in PandaX-4T, Phys. Rev. Lett. **131**, 191002 (2023), arXiv:2308.01540 [hep-ex].

[51] M. F. Albakry *et al.* (SuperCDMS), Search for low-mass dark matter via bremsstrahlung radiation and the Migdal effect in SuperCDMS, Phys. Rev. D **107**, 112013 (2023), arXiv:2302.09115 [hep-ex].

[52] P. Adari *et al.* (SENSEI), SENSEI: First Direct-Detection Results on sub-GeV Dark Matter from SENSEI at SNOLAB, (2023), arXiv:2312.13342 [astro-ph.CO].

[53] J. Xu *et al.*, Search for the Migdal effect in liquid xenon with keV-level nuclear recoils, Phys. Rev. D **109**, L051101 (2024), arXiv:2307.12952 [hep-ex].

[54] P. Cox, M. J. Dolan, C. McCabe, and H. M. Quiney, Precise predictions and new insights for atomic ionization from the Migdal effect, Phys. Rev. D **107**, 035032 (2023), arXiv:2208.12222 [hep-ph].

[55] H. Takahashi, Y. Mitsuya, T. Fujiwara, and T. Fushie, Development of a glass GEM, Nucl. Instrum. Meth. A **724**, 1 (2013).

[56] MIGDAL Collaboration, First Operations and Performance of the Optical Time Projection Chamber of the MIGDAL Experiment [In Progress], (2024).

[57] F. Brunbauer, F. García, T. Korkalainen, A. Lugstein, M. Lupberger, E. Oliveri, D. Pfeiffer, L. Ropelewski, P. Thuiner, and M. Schinnerl, Combined Optical and Electronic Readout For Event Reconstruction in a GEM-based TPC, IEEE Transactions on Nuclear Science **65**, 913 (2018).

[58] E. Tilly and M. Handley (MIGDAL), 3D track reconstruction of low-energy electrons in the MIGDAL low pressure optical time projection chamber, JINST **18** (07), C07013, arXiv:2307.10477 [hep-ex].

[59] *EHD F0.85 Lenses*, EHD Imaging GmbH, Zum Rennplatz 15, D-49401 Damme, Germany (2020), available at https://www.ehd.de/wp-content/uploads/2024/05/EHD-F085-Lenses.pdf.

[60] E. R. Fossum *et al.*, Charge transfer noise and lag in cmos active pixel sensors, in *Proc. 2003 IEEE Workshop on CCDs and Advanced Image Sensors, Elmau, Bavaria, Germany* (Citeseer, 2003) pp. 11–13.

[61] S. Höfer *et al.*, Sim2real in robotics and automation: Applications and challenges, IEEE Transactions on Automation Science and Engineering **18**, 398 (2021).

[62] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, *et al.*, Pytorch: An imperative style, high-performance deep learning library, Advances in neural information processing systems **32** (2019).

[63] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, Generalized intersection over union: A metric and a loss for bounding box regression, in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2019) pp. 658–666.

[64] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, Distance-IoU loss: Faster and better learning for bounding box regression, in *Proceedings of the AAAI conference on artificial intelligence*, Vol. 34 (2020) pp. 12993–13000.

[65] X. Li, W. Wang, L. Wu, S. Chen, X. Hu, J. Li, J. Tang, and J. Yang, Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection, Advances in Neural Information Processing Systems **33**, 21002 (2020).

[66] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, Microsoft coco: Common objects in context, in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13* (Springer, 2014) pp. 740–755.

[67] C. Tang, S. Vishwakarma, W. Li, R. Adve, S. Julier, and K. Chetty, Augmenting experimental data with simulations to improve activity classification in healthcare monitoring, in *2021 IEEE Radar Conference (RadarConf21)* (2021) pp. 1–6.

[68] M. Tkachenko, M. Malyuk, A. Holmanyuk, and N. Liubimov, Label Studio: Data labeling software (2020-2022), open source software available from https://github.com/heartexlabs/label-studio.

[69] I. Loshchilov and F. Hutter, Decoupled weight decay regularization, arXiv preprint arXiv:1711.05101 (2017).

[70] L. Prechelt, Early Stopping — But When?, in *Neural Networks: Tricks of the Trade: Second Edition*, edited by G. Montavon, G. B. Orr, and K.-R. Müller (Springer Berlin Heidelberg, Berlin, Heidelberg, 2012) pp. 53–67.

[71] R. Padilla, S. Netto, and E. da Silva, A survey on performance metrics for object-detection algorithms (2020).

[72] J. F. Ziegler, M. D. Ziegler, and J. P. Biersack, SRIM - The stopping and range of ions in matter (2010), Nuclear Instruments and Methods in Physics Research B **268**, 1818 (2010).

[73] V. Klema and A. Laub, The singular value decomposition: Its computation and some applications, IEEE Transactions on Automatic Control **25**, 164 (1980).

[74] M. T. Hedges, S. E. Vahsen, I. Jaegle, P. M. Lewis, H. Nakayama, J. Schueler, and T. N. Thorpe, First 3D vector tracking of helium recoils for fast neutron measurements at SuperKEKB, Nucl. Instrum. Meth. A **1026**, 166066 (2022), arXiv:2106.13079 [physics.ins-det].