

PREPARED FOR SUBMISSION TO JINST

Cosmic ray muon clustering for the MicroBooNE liquid argon time projection chamber using sMask-RCNN

MicroBooNE Collaboration

P. Abratenko^{hh} R. Anⁿ J. Anthony^d L. Arellano^r J. Asaadi^{gg} A. Ashkenazi^{ee}
 S. Balasubramanian^k B. Baller^k C. Barnes^t G. Barr^w J. Barrow^{s,dd} V. Basque^{k,r}
 L. Bathe-Peters^m O. Benevides Rodrigues^{dd} S. Berkman^k A. Bhandari^r A. Bhat^{dd} M. Bishai^b
 A. Blake^p T. Bolton^o J. Y. Book^m L. Camilleriⁱ D. Caratelli^{c,k} I. Caro Terrazas^h F. Cavanna^k
 G. Cerati^k Y. Chen^{a,aa} E. Church^x D. Cianciⁱ J. M. Conrad^s M. Convery^{aa}
 L. Cooper-Troendle^{kk} J. I. Crespo-Anadón^e M. Del Tutto^k S. R. Dennis^d P. Detje^d A. Devitt^p
 R. Diurba^u R. Dorrellⁿ K. Duffy^{k,w} S. Dytman^y B. Eberly^{cc} A. Ereditato^a J. J. Evans^r R. Fine^q
 G. A. Fiorentini Aguirre^{bb} R. S. Fitzpatrick^t B. T. Fleming^{kk} N. Foppiani^{im} D. Franco^{kk}
 A. P. Furmanski^u D. Garcia-Gamez^l S. Gardiner^k G. Geⁱ S. Gollapinni^{ee,q} O. Goodwin^r
 E. Gramellini^k P. Green^r H. Greenlee^k W. Gu^b R. Guenette^m P. Guzowski^r L. Hagaman^{kk}
 O. Hen^s C. Hilgenberg^u G. A. Horton-Smith^o A. Hourlier^s R. Itay^{aa} C. James^k X. Ji^b L. Jiangⁱⁱ
 J. H. Jo^{kk} R. A. Johnson^g Y.-J. Jwaⁱ D. Kalraⁱ N. Kamp^s N. Kaneshige^c G. Karagiorgiⁱ
 W. Ketchum^k M. Kirby^k T. Kobilarcik^k I. Kreslo^a I. Lepetic^z J.-Y. Li^j K. Li^{kk} Y. Li^b K. Lin^q
 B. R. Littlejohnⁿ W. C. Louis^q X. Luo^c K. Manivannan^{dd} C. Marianiⁱⁱ D. Marsden^r
 J. Marshall^{jj} D. A. Martinez Caicedo^{bb} K. Mason^{hh} A. Mastbaum^z N. McConkey^r
 V. Meddage^o T. Mettler^a K. Miller^f J. Mills^{hh} K. Mistry^r A. Mogan^h T. Mohayai^k J. Moon^s
 M. Mooney^h A. F. Moor^d C. D. Moore^k L. Mora Lepin^r J. Mousseau^t S. Mullerlababu^a
 M. Murphyⁱⁱ D. Naples^y A. Navrer-Agasson^r M. Nebot-Guinot^j R. K. Neely^o D. A. Newmark^q
 J. Nowak^p M. Nunes^{dd} O. Palamara^k V. Paolone^y A. Papadopoulou^s V. Papavassiliou^v
 S. F. Pate^v N. Patel^p A. Paudel^o Z. Pavlovic^k E. Piasetzky^{ee} I. D. Ponce-Pinto^{kk} S. Prince^m
 X. Qian^b J. L. Raaf^k V. Radeka^b A. Rafique^o M. Reggiani-Guzzo^r L. Ren^v L. C. J. Rice^y
 L. Rochester^{aa} J. Rodriguez Rondon^{bb} M. Rosenberg^y M. Ross-Lonerganⁱ G. Scanavini^{kk}
 D. W. Schmitz^f A. Schukraft^k W. Seligmanⁱ M. H. Shaevitzⁱ R. Sharankova^{k,gg} J. Shi^d
 J. Sinclair^a A. Smith^d E. L. Snider^k M. Soderberg^{dd} S. Söldner-Rembold^r P. Spentzouris^k
 J. Spitz^t M. Stancari^k J. St. John^k T. Strauss^k K. Suttonⁱ S. Sword-Fehlberg^v A. M. Szelc^j
 N. Tagg^w W. Tang^{f,f} K. Terao^{aa} C. Thorpe^p D. Totani^c M. Touns^k Y.-T. Tsai^{aa} M. A. Uchida^d
 T. Usher^{aa} W. Van De Pontseele^{w,m} B. Viren^b M. Weber^a H. Wei^b Z. Williams^{gg} S. Wolbers^k
 T. Wongjirad^{hh} M. Wospakrik^k K. Wresilo^d N. Wright^s W. Wu^k E. Yandel^c T. Yang^k
 G. Yarbrough^{f,f} L. E. Yates^s F. J. Yu^{hh} H. W. Yu^b G. P. Zeller^k J. Zennaro^k C. Zhang^b

^aUniversität Bern, Bern CH-3012, Switzerland^bBrookhaven National Laboratory (BNL), Upton, NY, 11973, USA^cUniversity of California, Santa Barbara, CA, 93106, USA^dUniversity of Cambridge, Cambridge CB3 0HE, United Kingdom

- ^e *Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid E-28040, Spain*
- ^f *University of Chicago, Chicago, IL, 60637, USA*
- ^g *University of Cincinnati, Cincinnati, OH, 45221, USA*
- ^h *Colorado State University, Fort Collins, CO, 80523, USA*
- ⁱ *Columbia University, New York, NY, 10027, USA*
- ^j *University of Edinburgh, Edinburgh EH9 3FD, United Kingdom*
- ^k *Fermi National Accelerator Laboratory (FNAL), Batavia, IL 60510, USA*
- ^l *Universidad de Granada, E-18071, Granada, Spain*
- ^m *Harvard University, Cambridge, MA 02138, USA*
- ⁿ *Illinois Institute of Technology (IIT), Chicago, IL 60616, USA*
- ^o *Kansas State University (KSU), Manhattan, KS, 66506, USA*
- ^p *Lancaster University, Lancaster LA1 4YW, United Kingdom*
- ^q *Los Alamos National Laboratory (LANL), Los Alamos, NM, 87545, USA*
- ^r *The University of Manchester, Manchester M13 9PL, United Kingdom*
- ^s *Massachusetts Institute of Technology (MIT), Cambridge, MA, 02139, USA*
- ^t *University of Michigan, Ann Arbor, MI, 48109, USA*
- ^u *University of Minnesota, Minneapolis, Mn, 55455, USA*
- ^v *New Mexico State University (NMSU), Las Cruces, NM, 88003, USA*
- ^w *University of Oxford, Oxford OX1 3RH, United Kingdom*
- ^x *Pacific Northwest National Laboratory (PNNL), Richland, WA, 99352, USA*
- ^y *University of Pittsburgh, Pittsburgh, PA, 15260, USA*
- ^z *Rutgers University, Piscataway, NJ, 08854, USA, PA*
- ^{aa} *SLAC National Accelerator Laboratory, Menlo Park, CA, 94025, USA*
- ^{bb} *South Dakota School of Mines and Technology (SDSMT), Rapid City, SD, 57701, USA*
- ^{cc} *University of Southern Maine, Portland, ME, 04104, USA*
- ^{dd} *Syracuse University, Syracuse, NY, 13244, USA*
- ^{ee} *Tel Aviv University, Tel Aviv, Israel, 69978*
- ^{ff} *University of Tennessee, Knoxville, TN, 37996, USA*
- ^{gg} *University of Texas, Arlington, TX, 76019, USA*
- ^{hh} *Tufts University, Medford, MA, 02155, USA*
- ⁱⁱ *Center for Neutrino Physics, Virginia Tech, Blacksburg, VA, 24061, USA*
- ^{jj} *University of Warwick, Coventry CV4 7AL, United Kingdom*
- ^{kk} *Wright Laboratory, Department of Physics, Yale University, New Haven, CT, 06520, USA*

E-mail: microboone_info@fnal.gov

ABSTRACT: In this article, we describe a modified implementation of Mask Region-based Convolutional Neural Networks (Mask-RCNN) for cosmic ray muon clustering in a liquid argon TPC and applied to MicroBooNE neutrino data. Our implementation of this network, called sMask-RCNN, uses sparse submanifold convolutions to increase processing speed on sparse datasets, and is compared to the original dense version in several metrics. The networks are trained to use wire readout images from the MicroBooNE liquid argon time projection chamber as input and produce individually labeled particle interactions within the image. These outputs are identified as either cosmic ray muon or electron neutrino interactions. We find that sMask-RCNN has an average pixel clustering efficiency of 85.9% compared to the dense network's average pixel clustering efficiency of 89.1%. We demonstrate the ability of sMask-RCNN used in conjunction with MicroBooNE's state-of-the-art Wire-Cell cosmic tagger to veto events containing only cosmic ray muons. The addition of sMask-RCNN to the Wire-Cell cosmic tagger removes 70% of the remaining cosmic ray muon background events at the same electron neutrino event signal efficiency. This event veto can provide 99.7% rejection of cosmic ray-only background events while maintaining an electron neutrino event-level signal efficiency of 80.1%. In addition to cosmic ray muon identification, sMask-RCNN could be used to extract features and identify different particle interaction types in other 3D-tracking detectors.

KEYWORDS: Pattern recognition, cluster finding, Particle identification methods, Time projection chambers

ARXIV EPRINT: [ARXIV GOES HERE.](#)

Contents

1	Introduction	1
2	Background	2
2.1	The MicroBooNE LArTPC	2
2.2	Existing cosmic identification tools	4
2.3	Object detection and Mask-RCNN	5
3	Methods	6
3.1	Data preparation	6
3.2	Sparse submanifold convolutions	7
3.3	Sparse Mask-RCNN	8
3.4	Network training	8
4	Comparing dense and sparse performance	10
4.1	Efficiency and purity	11
4.2	Interaction coverage	15
4.3	Network comparisons discussion	17
5	Finding electron neutrinos with sMask-RCNN	17
5.1	Electron neutrino identification	18
5.2	Cosmic-only event veto	20
6	Conclusions	25
7	Bibliography	26

1 Introduction

The MicroBooNE [1] experiment uses a liquid argon time projection chamber (LArTPC) with an active volume of 85 tonnes to study neutrinos from the Fermilab Booster Neutrino Beamline, while also receiving neutrinos from the Neutrinos at the Main Injector (NuMI) beam. The MicroBooNE LArTPC is a near-surface detector that does not utilize any overhead shielding for cosmic background mitigation. This, combined with a long TPC readout time, described in section 2.1, results in a high ratio of cosmic ray muons to the number of neutrinos that interact within the detector. Techniques must be developed to reduce this cosmic ray muon background so that different neutrino interaction channels are measured with high purity.

A cosmic ray interaction can be mistaken for a neutrino interaction regardless of whether there is a neutrino interaction in the readout window. We consider any cosmic ray muon depositing

charge in the detector an ‘interaction’ regardless of whether it is captured in, decays in, or traverses the detector. Background events without a neutrino interaction present are called “cosmic-only” events. An example of a cosmic-only event is depicted in figure 1c. Due to the prevalence of cosmic rays at this near-surface location, any reconstruction tools designed to tag cosmic ray muons should be deployed early in the reconstruction chain to filter such non-neutrino events. This means that our solution to identifying and removing cosmic ray backgrounds must be deployable across the entire MicroBooNE dataset.

In this article, we present an approach to cosmic ray muon tagging using machine learning. Within the MicroBooNE experiment, machine learning techniques have been applied in other areas, such as particle identification, and pixel identification [2–5]. In this article, we make use of a neural network called Mask-RCNN or "Mask Region-based Convolutional Neural Network" [6] to locate, identify, and cluster 2D interactions corresponding to the projections of the LArTPC wire planes. The design of Mask-RCNN is described in section 2.3.

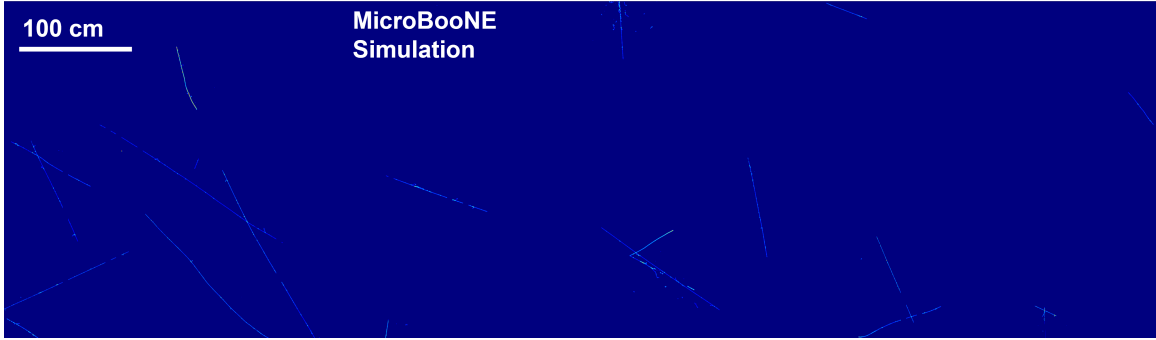
Machine learning algorithms are typically deployed on graphics processing units (GPUs) because their ability to parallelize computations pairs well with the matrix multiplications that are abundant in machine learning code. However, in order to deploy on the full MicroBooNE dataset, our tools need to run on central processing units (CPUs) because the MicroBooNE production chain has access to large amounts of CPUs but not integrated GPUs. While running on GPUs would speed up run time, integrating GPUs would require additional personnel and financial investment that is not presently feasible, therefore operations that use CPUs are required. To solve this problem, we extend Mask-RCNN to use sparse submanifold convolutions [7] which allow for much faster CPU running on sparse datasets by avoiding multiplication when one term is zero. We call this modified version of the network sparse Mask-RCNN or "sMask-RCNN". Section 3.2 contains a brief description of submanifold convolutions, and sMask-RCNN’s modified state is described in 3.3. Several visual examples of sMask-RCNN’s performance are provided in the form of event images in figure 1. For details on how the event images underlying the sMask-RCNN labels are made, see section 3.1. Code detailing our implementation of sMask-RCNN has been made available at: https://github.com/NuTufts/Detectron.pytorch/tree/larcv1_mcc9.

2 Background

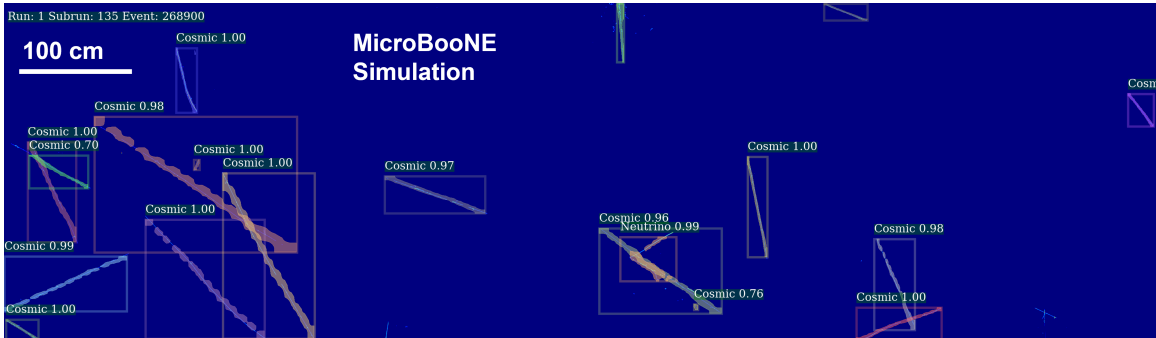
2.1 The MicroBooNE LArTPC

The LArTPC technology is designed to provide precision calorimetry and particle tracking while remaining scalable to larger sizes for future experiments. In MicroBooNE’s LArTPC, a large volume of liquid argon is bounded on six sides within the time-projection chamber (TPC). On one side, the cathode, is a metal plate held at a negative potential of -70 kV. On the other side of the argon, held near ground, is the anode: a collection of three wire planes at progressively higher potentials. Figure 2 shows a diagram of the LArTPC principle.

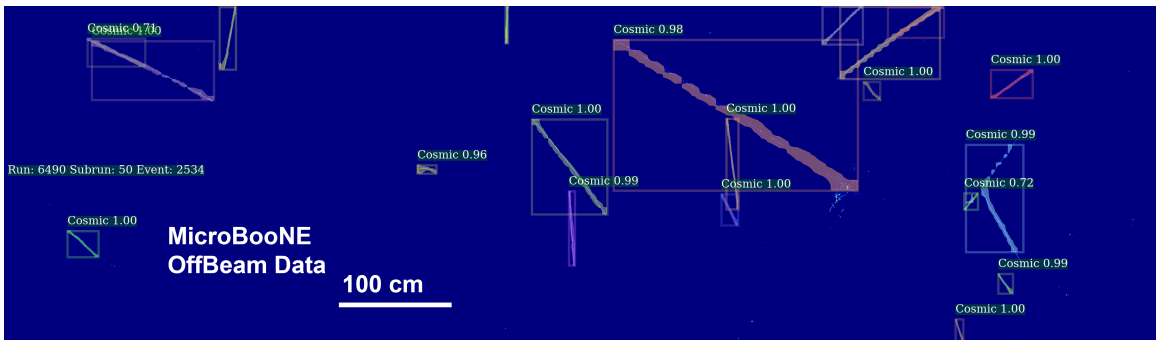
Each wire plane consists of a series of parallel wires spaced every 0.3 cm. The wires in each plane are oriented at an angle of 60° with respect to the wires in the other two planes. When a charged particle passes through the detector, it creates ionization electrons which drift toward the anode wire planes due to the nearly uniform electric field between the cathode and anode. As the



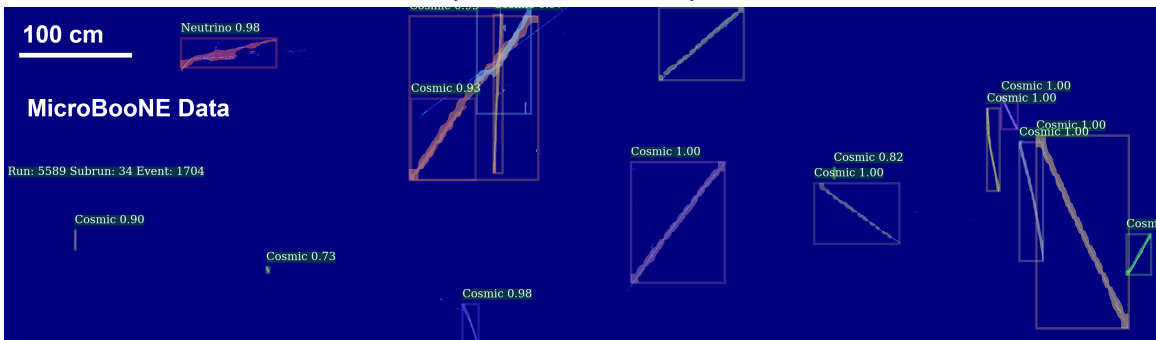
(a) An example of an input image as given to sMask-RCNN to process



(b) A simulated neutrino interaction overlaid on cosmic ray muons from data, labeled by sMask-RCNN



(c) Cosmic-only data event, labeled by sMask-RCNN



(d) Data event containing a neutrino interaction, labeled by sMask-RCNN

Figure 1: Several example event images. The vertical and horizontal scales are the same for all images. Each column of pixels along the x -axis refers to a specific wire readout, and each row along the y -axis refers to a different bin of signal readout time. This is described in greater detail in section 3.1. (a) is an example of an input image given to sMask-RCNN to process, whereas (b) shows the network's subsequent labeling of the same image. (c) shows a cosmic-only data event. (d) shows a data event containing a neutrino interaction that sMask-RCNN correctly identifies with some confidence score, and clusters.

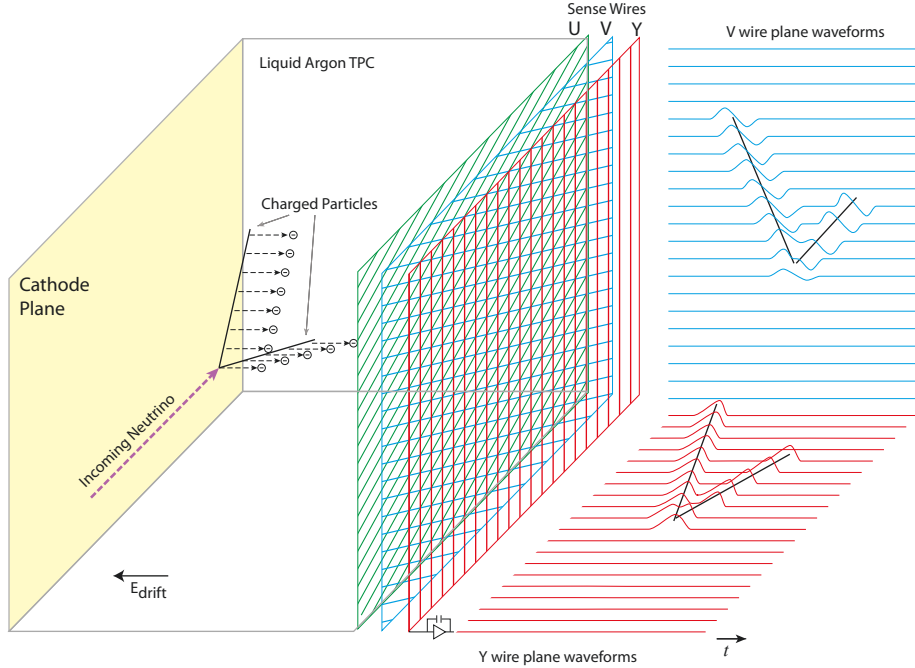


Figure 2: A diagram of the LArTPC principle. The signal formation for the second induction plane (V plane) and the collection plane (Y plane) are shown [1].

electrons pass by the first two wire planes they induce bipolar pulses on the wires before finally arriving at the third and final wire plane. Here they are collected and create a unipolar pulse. Thus this final wire plane is called the "collection plane". The bipolar and unipolar pulses read out from the wires undergo noise filtering and 2D deconvolution described in [8, 9]. This processing removes much of the noise from the wires and transforms the bipolar pulses into unipolar pulses. These wire signals are used to create the input images used by the neural network. The process for creating the input images is described in section 3.1.

2.2 Existing cosmic identification tools

Cosmic ray muon tagging and background removal have previously been performed in MicroBooNE using a variety of methods. One example uses deep learning with semantic segmentation to differentiate cosmic ray muon pixels from neutrino interaction pixels [10]. Mask-RCNN expands on semantic segmentation, further separating each instance of every individual interaction it finds using bounding boxes. This means that each cosmic muon interaction in the detector receives its own labeled and clustered output.

Cosmic ray muon tagging has also been performed with more traditional algorithmic approaches. In the MicroBooNE experiment, one such method is the PandoraCosmic algorithm [11, 12]. This algorithm clusters hits in 2D and then combines these clusters into 3D tracks. It

flags a track as a cosmic ray muon if part of the track is placed outside the detector based on timing information or if the track trajectory begins and ends at a TPC boundary using information related to the track's timing. It provides an exception for through-going trajectories that are parallel to the beam direction. This is designed to eliminate cosmic ray muons as they will appear as tracks originating from outside of the detector, and will be crossing perpendicular relative to the beam direction.

Another method used in MicroBooNE is the Wire-Cell (WC) cosmic tagger [13, 14] which is made up of several event-level requirements, combined with the WC charge-light (Q-L) matching algorithm [15]. This Q-L matching algorithm uses light information detected during the neutrino spill by 32 8-inch cryogenic photomultiplier tubes (PMTs) mounted behind the TPC wire planes. This light information is then spatially matched to charge deposited in the TPC, selecting TPC pulses created during the beam spill. Therefore, both the Q-L algorithm and the full WC cosmic tagger use additional information beyond the wire planes, which is what sMask-RCNN uses as input. In section 5.2 we show results achieved by a combination of sMask-RCNN and WC algorithms to produce a state-of-the-art cosmic ray tagger.

Cosmic ray tagging can also be achieved with hardware solutions. An example in MicroBooNE is the design and construction of the cosmic ray tagger system [16]. This system was introduced partway through MicroBooNE operation, and therefore is only available for part of the MicroBooNE data. It uses plastic scintillation modules to acquire the time and location for particles traversing the TPC. Reconstructed tracks can then be matched to this data and be flagged as pertaining to cosmic ray muons rather than neutrino interactions. Additionally, the cosmic ray tagger system can be used in tandem with software solutions to improve performance.

2.3 Object detection and Mask-RCNN

The original Mask-RCNN network is designed to perform three common tasks in the field of computer vision: object detection, classification, and semantic segmentation. In the field of computer vision, classification is a task commonly performed to label an image as one of some predefined list of classes, for instance an image might be labeled a cat or a dog. Semantic segmentation refers to a labeling performed at pixel level, for example in an image with a cat and a dog, the pixels making up the dog are labeled 'dog' pixels and those making up the cat are labeled 'cat' pixels, while the remaining pixels are given a background label. The network is trained to receive some input image, place bounding boxes around objects of interest, classify these objects within some set of user defined classes, then within each box label each pixel as part of the object or not.

Structurally, Mask-RCNN is comprised of four subnetworks. First is a residual network (ResNet) [17]. This network runs on the input image and creates a feature map for the image. This feature map is then fed into a region proposal network (RPN) [18], which then produces a series of bounding boxes around regions of interest (RoIs) within the image. The bounding boxes are described by a 2D coordinate, a height and a width, and are designed to produce the smallest rectangular box containing the object. The RoIs are aligned in the feature map space via the RoIAlign algorithm, then combined with appropriate features, scaled to a fixed size and fed into the two final subnetworks: a classifier, and a fully convolutional network (FCN) we refer to as the "maskifier". The classifier takes each bounding box and its features and predicts which class of

object it is with some confidence score. The maskifier produces a semantic segmentation mask of all the pixels within the box, determining which pixels correspond to the object and which are background. This semantic segmentation mask is synonymous with a cluster of pixels within the box, though the cluster need not be connected. Figure 3 shows a simplified view of the network architecture.

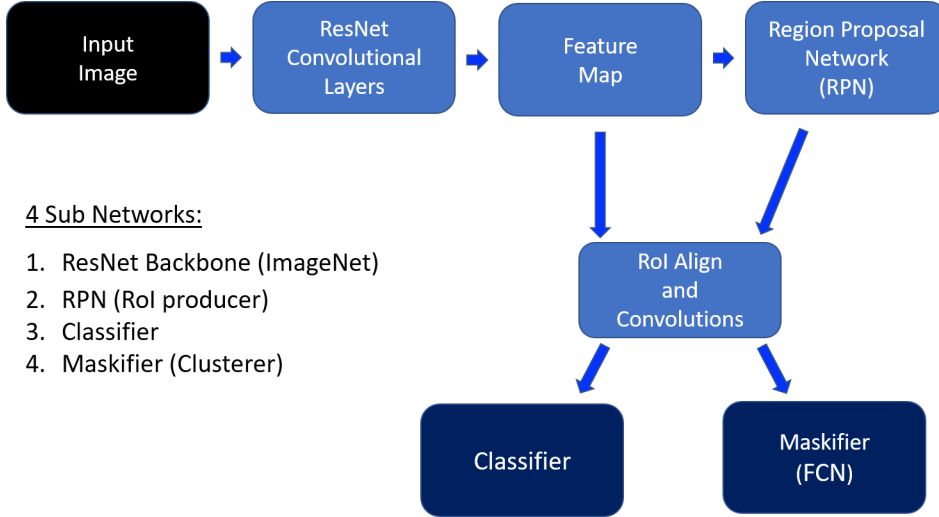


Figure 3: Network Architecture for Mask-RCNN in MicroBooNE.

3 Methods

3.1 Data preparation

To create images for analysis using MicroBooNE LArTPC data, we use the charge readout from the three wire planes at the anode. The neutrino beam window is 1.6 ms, but we record a buffer on either side providing a modified window of about 3 ms. The data taken during this 3 ms comprises an "event". Over the course of this recorded beam window these wires are sampled as a rate of 2 MHz. This equates to 6048 samples per wire.

We create the event images shown in this article by placing the wire number along the x -axis and the sampling time along the y -axis. However, we first downsample the number of time samples per wire by a factor of six, going from 6048 samples per wire to 1008. This downsampling is performed both to reduce image size and to make the drift distance per pixel of an ionization electron roughly 0.3 cm. This distance matches the wire separation within a given wire plane. The value stored in each pixel within an image then is proportional to the charge deposited in the detector. We will refer to this as the "pixel intensity". We then apply a threshold the image by setting any pixel with pixel intensity below 10 (arbitrary units) equal to zero to further reduce noise. In comparison, the pixel intensity distribution from minimum ionizing particles peaks at ~ 40 in these arbitrary units.

While there is one image made for each of the three LArTPC wire planes, for this study we only use the collection plane. We choose the collection plane over the other wire planes because the collection plane does not require signal processing to turn bipolar pulses into unipolar pulses

and therefore the signal is cleaner. The event images then have a dimensionality of the number of wires on the collection plane times the number of samplings per wire, or 3456×1008 .

Lastly, in the MicroBooNE LArTPC a portion of the wires are unresponsive [19]. In the collection plane this happens for about 10% of the wires. This creates an artifact in the images by creating vertical lines of unresponsive pixels. In some cases large groups of adjacent wires are unresponsive, leading to regions of unresponsive pixels.

3.2 Sparse submanifold convolutions

In the event images created from the LArTPC collection plane, about 0.7% of the pixels are nonzero, making the data "sparse". This is because the pixel value comes from the ionized electrons drifting away from the charged particles moving in the detector. Most of the time, the wires are reading out low-level noise that is below the threshold. The resultant low pixel occupancy means that when we apply Mask-RCNN to the event images, there are many computations that involve multiplications by zero. These trivial calculations waste computing resources, particularly if performed in sequence via a CPU and not in parallel via a GPU.

A normal convolution in a neural network takes a convolutional kernel or filter and moves it across the image, multiplying at each location to acquire the convolved value. In comparison, a sparse submanifold convolution only multiplies the kernel against positions centered on nonzero pixels, avoiding computations on zeroed regions of the input. Notably, a submanifold convolution is not mathematically equivalent to a regular convolution. In normal convolutions, kernels centered on zeros in the input image can output a nonzero convolved value if the edge of the filter captures some nonzero input. An example of this is shown in figure 4. This blurs the features coming out of a convolution, so that convolutions which are performed one after another in a deep neural network, such as ResNet, spread information outward.

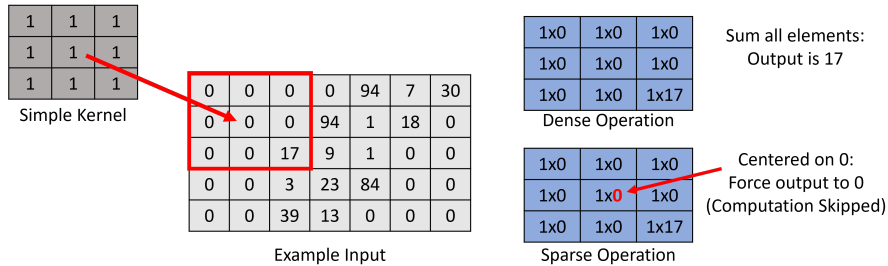


Figure 4: An example of a convolution operation that depends on use of submanifold or normal convolutions. The normal convolution multiplies the kernel against the image at the given position, and outputs a convolved value of 17. Meanwhile, the submanifold convolution does not get computed when centered on a zero. The submanifold convolutions used in sMask-RCNN have a kernel size of 3×3 (same as shown in the figure), with a stride of 1.

3.3 Sparse Mask-RCNN

We utilize Mask-RCNN to locate, identify, and cluster interactions within the 2D event images described in section 3.1. The network places bounding boxes around, classifies, and then clusters pixels corresponding to deposited charge for each interaction it finds within the image. We define an interaction as all of the charge deposited in an image coming from the same ancestor particle. For example, if an electron neutrino interacts with the argon in the LArTPC, and yields a proton and an electron, all of the charge deposited from the proton and electron are combined into one ancestor "electron neutrino" interaction and should be clustered together. A more detailed description of network training is described in section 3.4.

To speed up the network when deployed on CPUs, we swap the ResNet convolutions with sparse submanifold convolutions while maintaining the original network structure. Additional work could be done to make the later subnetworks use sparse convolutions but examinations of the compute times for the individual parts of sMask-RCNN made this unnecessary. For clarity, we will refer to Mask-RCNN without submanifold convolutions as dense Mask-RCNN.

The change to sparse convolutions yields a significant gain in terms of network speed for inference when running on CPUs. The timing information for running the dense and sparse configurations of the network is shown in table 1. These timings were performed on an Intel(R) Core(TM) i9-9820X CPU @ 3.30 GHz and measure wall time. This allows the network to be inserted into production code on CPU farms such as FermiGrid [20] and to scale up how quickly the network is run over large data samples. We note that the Intel CPU we tested on is superior to what is generally available on CPU farms. Brief testing done on available CPU farms shows the ResNet runtime difference is exacerbated. This means the difference between sparse and dense implementations is even greater when older CPUs are used. This further prioritizes shortening the ResNet runtime, as when we deploy on CPU farms, we will use a variety of CPUs with lesser performance than an Intel(R) Core(TM) i9-9820X CPU @ 3.30 GHz. Further, an added benefit of using CPUs is that this technique is scalable to future experiments and studies where more data may be analyzed, and the computing resources cannot scale to a reasonable number of GPUs.

The implementation to sparse ResNet also introduces improvements to training when it comes to memory. For the dense version of the network, due to memory constraints, we could only train the network on 832×512 crops of the event image. We use the word crop to refer to a random cutout of the original image of this new 832×512 size, where the crop must contain a portion of a simulated interaction. However, sMask-RCNN is trained on the full 3456×1008 event images because the memory required on the GPU to store a full image for the network is reduced by roughly a factor of the image's occupancy, as all zero pixels are no longer operated on. Regardless of training size, both the dense and sparse forms of Mask-RCNN could be deployed on the full 3456×1008 event images, as less memory is used if not actively training the network.

All of the event images shown in this article use the sparse implementation of Mask-RCNN. We examine the difference in performance between the dense and sparse networks in section 4.

3.4 Network training

When training, the entire network performs a forward pass on an image, and then the backward pass updates the weights of all the subnetworks based on a combined loss function built from

Table 1: The average inference runtimes per 3456×1008 pixel image on a CPU. The first row is the runtime for just the ResNet portion of Mask-RCNN on the images. The second row is the time to run the entire network on the images. In the case of sparse ResNet, the time spent making the input image into a sparse tensor and the output features into a dense tensor is included in the sparse ResNet module time.

	Dense ResNet	Sparse ResNet
ResNet Runtime	3.172 s	0.1758 s
Full Detection Runtime	8.438 s	5.79 s

the outputs of the maskifier, classifier, and RPN, as described in the original Mask-RCNN article [6]. Rather than train the ResNet from randomized initial weights, we use weights pretrained on the ImageNet dataset [21], which is a publicly available labeled dataset of images of animals and everyday objects. The ImageNet dataset is commonly used in the field of computer vision. We briefly started training ResNet from scratch, with randomly initialized weights, because the ImageNet pretraining is designed to identify animals and everyday objects, but despite this, we found that using the pretrained ResNet gave more useful features for the other components of Mask-RCNN to utilize in this particle physics analysis.

To train both the dense and sparse forms of Mask-RCNN, we use a sample of simulated electron neutrino events featuring simulated cosmic background. This means that every full 3456×1008 image contains a single electron neutrino interaction among many cosmic ray muons. While the dense network was trained on crops containing at least part of a neutrino interaction or cosmic ray muon, they did not always have an example of both within the same crop.

The interactions present in the training data are broken up into six different interaction classes, detailed in table 2. Dense Mask-RCNN was trained on each of these interactions, but as we determined our goal was primarily cosmic ray muon tagging, we only trained the sparse implementation on cosmic ray muons and electron neutrino interactions. Simulated interactions where the simulated ancestor particle was one of the four other classes were still present in the data but the network was told to ignore them. The training data uses CORSIKA [22] for cosmic ray muon simulation. For electron neutrino interactions the GENIE neutrino interaction simulator [23] is used. In both instances, GEANT4 [24] is used to model the detector response.

We also note that the version of simulation used to produce the training data for the sparse network is slightly updated as a newer version of simulation became available. Both the dense and sparse networks have their performance evaluated on the same set of 9400 events from the newer version of simulation, meaning there is a slight discrepancy between the training and testing data for the dense network. However, we expect this change to have minimal effect on the network performance.

As is common practice in machine learning, we split the data into two orthogonal subsets: a training set with 80% of the events and a validation set with 20%. When training the network, we use events in the training set, and whenever we wish to measure the performance of the network, such as calculating performance metrics shown in section 4, we use the validation set. This is a critical part of machine learning because it verifies that the network can generalize and perform its

Table 2: The different class types and number of occurrences in the training sets for the dense and sparse versions of Mask-RCNN. Note that the sparse network only trained on cosmic ray muon and electron neutrino interactions.

Interaction Class	Dense Training		Sparse Training	
	Counts	Percentage	Counts	Percentage
Cosmic Ray Muon	2708730	92.99	786050	95.24
Electron Neutrino	97034	3.33	39296	4.76
Neutron	26072	0.90	-	-
Proton	5738	0.19	-	-
Electron	155	0.005	-	-
Other	75026	2.58	-	-

task on events outside of the training set.

The dense version of the network is trained on a sample of 230,000 crops for 1.75 epochs. In the context of machine learning, an epoch is one training pass through the data. While 1.75 is a low number of epochs, each crop features multiple interactions seen and masked by the network. Then the network is trained on a subset of 30,000 of these crops containing examples of high intersection-over-union (IOU) interactions. IOU between two interactions is defined as the number of pixels present in both interaction bounding boxes, divided by the total number of unique pixels present in either bounding box. These crops featured multiple interactions with overlapping bounding boxes. This fine tuned training is performed due to poor performance by the dense network on overlapping interactions. The training on this subset is performed for 8 epochs so the network can focus on learning these difficult events.

The sparse version of the network is trained on a sample of 40,000 full event images for three epochs. No fine tuning needed to be performed on highly overlapping interactions as the sparse network did not appear to suffer from the same issue as the dense network. It should be noted that while at first glance the training sample sizes of the dense and sparse networks differ by a significant factor, in actuality the dense crops were made at a factor of up to 10 crops per full image, and the training datasets are comparable in terms of interaction sample variance. Further, all performance evaluations for both networks are deployed on the same validation set of 9400 full sized images.

4 Comparing dense and sparse performance

In this section we define several metrics to test the performance of Mask-RCNN at identifying and clustering interactions within MicroBooNE event images. Then we compare the performance between the dense and sparse versions of the network. For this evaluation, we include interactions found by the network with a class score of 0.4 or higher. The class score is a score between 0 and 1.0 indicating how confident the network is that the class label is correct. This threshold is chosen to provide a balance between the purity and efficiency metrics defined and discussed below. Once this threshold is applied, all remaining predicted interactions are treated equally for the purpose of calculating metrics.

All analysis in this section uses the full scale 3456×1008 event images in the validation set of simulated data used to train the sparse network. This refers to the data described in the right two columns of table 2.

4.1 Efficiency and purity

For the purpose of this section, we will refer to interactions either as "true" or "predicted". A true interaction is one that exists in the simulation, whereas a predicted interaction is one that the network claimed it has found and labeled. We define two metrics to measure the pixel level efficiency and purity of the network's ability to find and cluster interactions. The efficiency is a measure of the percentage of pixels in a true interaction that are masked by the network's prediction. Purity is defined for each predicted interaction as the highest fraction of pixels belonging to the predicted interaction and a single true interaction. For example, if 30% of a prediction maps to true interaction A, and 50% maps to true interaction B then the purity is 50% for that predicted interaction. In both of these definitions, only pixels containing charge deposition above the pixel intensity threshold of 10 are considered, as we do not care about clustering empty pixels. Concretely, the efficiency E is defined as

$$E = \frac{\sum_{ij} T_{ij} \cdot W_{ij} \cdot M_{ij}}{\sum_{ij} T_{ij} \cdot W_{ij}}, \quad (4.1)$$

where T , W , and M are matrices representing the truth interaction, wire event image, and predicted interaction mask, respectively, with dimensions of the event image. Meanwhile, i and j are pixels indices. A visual representation of the efficiency calculation is depicted in figure 5. Similarly, the purity P is defined as

$$P = \frac{\sum_{ij} M_{ij} \cdot W_{ij} \cdot T_{ij}}{\sum_{ij} M_{ij} \cdot W_{ij}}. \quad (4.2)$$

The purity calculation is depicted in figure 6. For each of these equations, the values in T are 1 if the pixel belongs to the true interaction, and 0 otherwise, while the values in M are 1 if the pixel belongs to the predicted interaction mask, and 0 otherwise. Finally, the values in W are 1 if the pixel has any deposited charge, and 0 otherwise. After the element-wise multiplication of the matrices, the summations then run over the indices of the matrix. This corresponds to a counting of the pixels corresponding to the union of the given matrices.

A true interaction's efficiency is taken as the best value as calculated for all predicted interactions. A predicted interaction's purity is taken as the best value when calculated for all true interactions.

These definitions mean that a given event image will have one efficiency measurement for each true interaction, and one purity measurement for each predicted interaction. While we are aware that object identification customarily uses panoptic quality [25] or intersection-over-union as evaluation metrics, we choose to use efficiency and purity in better keeping with particle physics analysis language.

For each event, we average the purities and efficiencies for the predicted and true interactions. These averages are of $O(20)$ interactions, where there is a single neutrino interaction and many cosmic ray muon interactions. We reiterate that a cosmic ray muon 'interaction' is just any cosmic ray muon and potential daughter particles that deposit charge in the detector. These event-averaged values are plotted in the 2D histograms shown in figure 7.

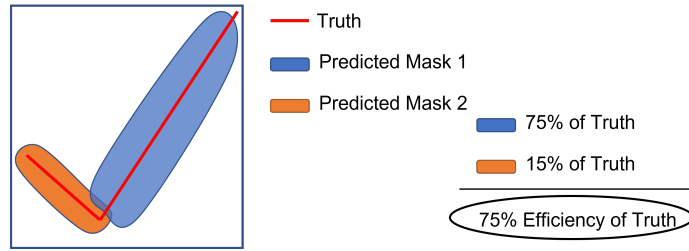


Figure 5: A visual representation of the definition of efficiency. Only nonzero pixels in the event image are counted.

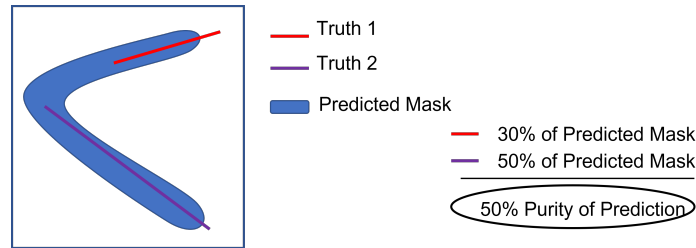


Figure 6: A visual representation of the definition of purity. Only nonzero pixels in the event image are counted.

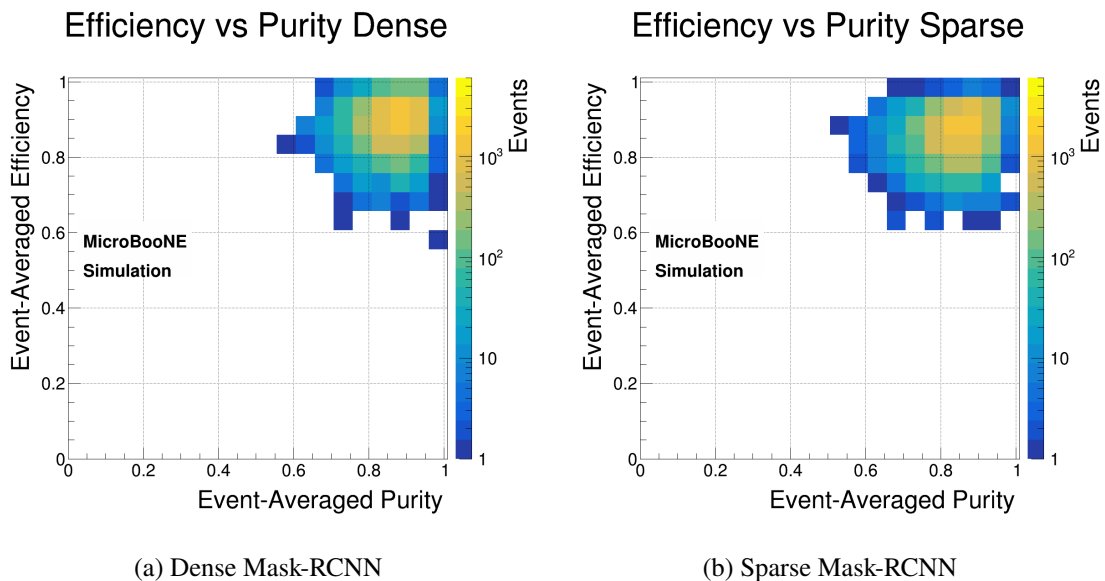


Figure 7: The event-averaged efficiencies and purities for the dense and sparse implementations of Mask-RCNN. The dense network has a mean event-averaged efficiency of 0.89 and a mean event-averaged purity of 0.87. For the sparse network these values are 0.86 and 0.85. Each of these evaluations use the same validation dataset.

Perfect efficiency and purity would yield values of 1.0 for each, so these histograms have targets in the upper right corners. We can see that the event-averaged purity drops from 87% for the dense to 85% for the sparse, while the event-averaged efficiency drops from 89% to 86%. The

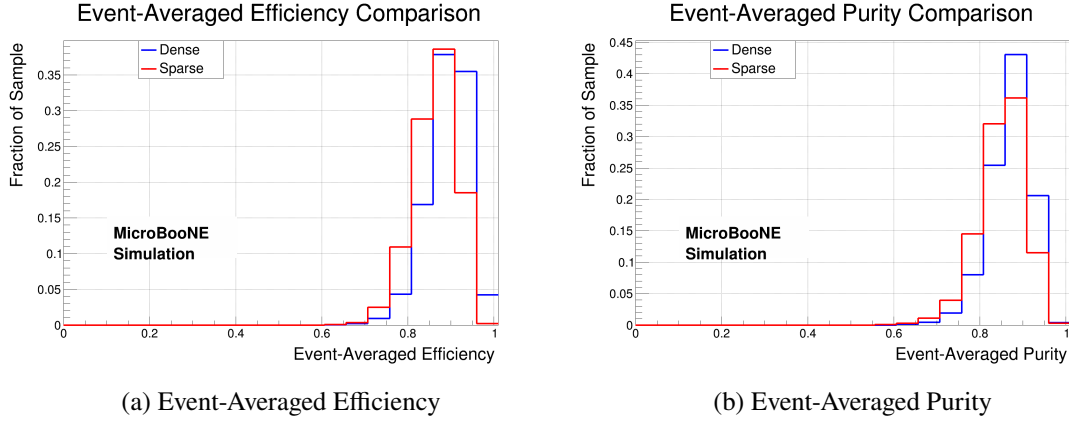


Figure 8: The one-dimensional distributions for the event-averaged efficiency (a) and purity (b) shown in figure 7. Each plot compares the dense and sparse network performances.

one-dimensional projections of figure 7 are shown in figure 8, where the event-averaged efficiency and purity distributions are compared between the dense and sparse networks.

It is also useful to examine the individual interaction efficiencies, rather than the event-averaged versions. The distributions for both the dense and sparse versions of Mask-RCNN are shown in figure 9. Here we can see the sparse (red) distribution is worse than the dense (blue) distribution. Notably the size of the peak at zero is the same for the two versions of the network. A true interaction will have zero efficiency if the network has no prediction that masks part of it. The fact that the two versions have the same sized peak at zero indicates that they each find the same number of interactions, but the dense masks are somewhat more complete.

The efficiency calculation is modified slightly by weighting the pixels within an interaction by their deposited charge. This version of the efficiency we term "charge efficiency" and is shown in figure 10. Here we see a shift to the right for each distribution compared to their pixel-level efficiency in figure 9. This indicates the network's preference for clustering higher value pixels corresponding to larger deposited charge, though this result may be due to the network being more likely to grab the center of tracks in our image, where the higher value pixels lie, compared to the halo of hits along a track's edge. While not surprising, this is a useful feature as the physics quantity we are dealing with is charge, not pixel count.

If we explore the interactions lying within the zero efficiency peaks in these plots then we find two common failure modes. The first is made up of interactions that lie completely or significantly in the unresponsive regions of the event image. Recall that roughly 10% of the MicroBooNE LArTPC wires are unresponsive, corresponding to vertical lines of unresponsive regions in the event images. True interactions within these regions are in the simulation, but have little in the way of signal in the event image for the network to detect. An example of a true interaction simulated in an unresponsive region of the image is shown in figure 11. It is unreasonable to expect the network to be able to label such interactions.

The second failure mode that contributes to the zero efficiency peak are true interactions that tend to be smaller in spatial extent compared to a typical simulated interaction, with less charge across fewer pixels in the event image. An example of this is shown in figure 12. These interactions

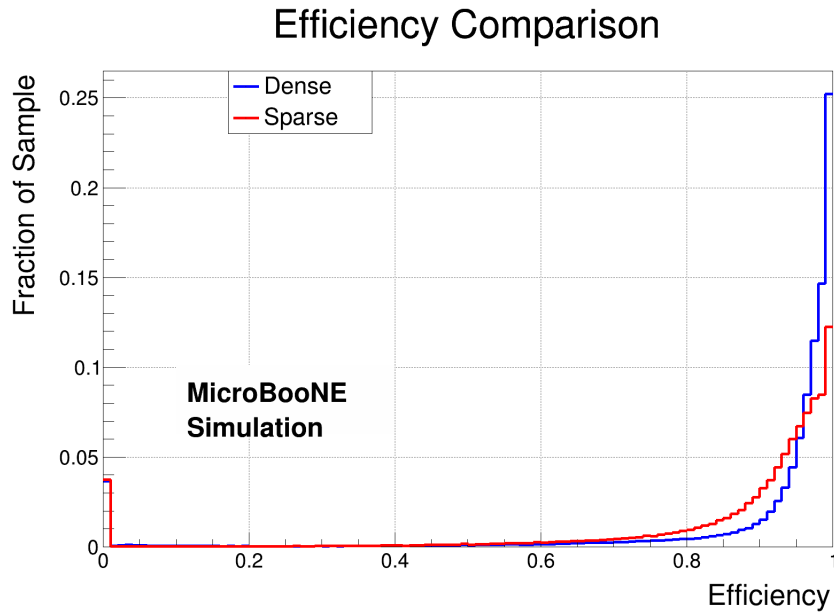


Figure 9: The interaction-level efficiency of the dense and sparse versions of Mask-RCNN as measured on the validation set.

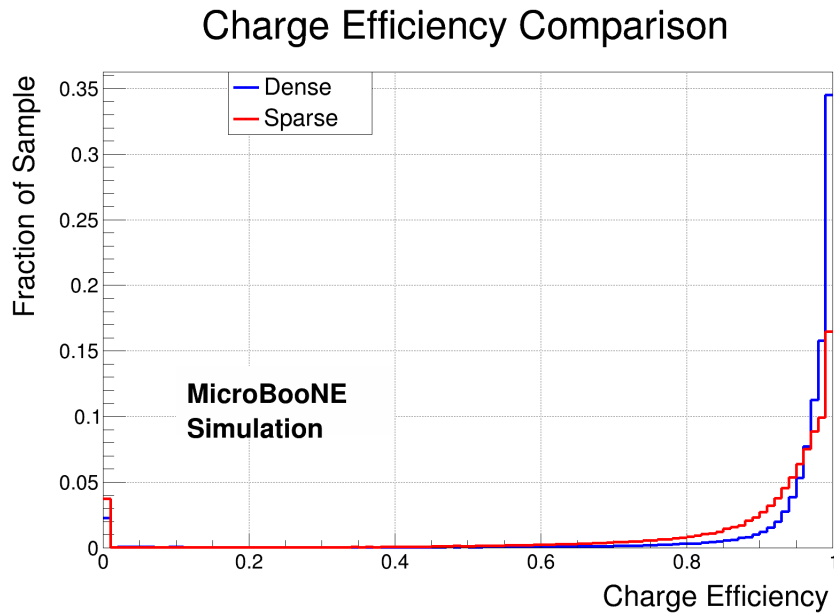


Figure 10: The interaction-level charge efficiency of the dense and sparse versions of Mask-RCNN as measured on the validation set.

are reasonable to expect the network to find as there is nothing to obscure the interaction. However, as they tend to be smaller, they are less likely to overlap with a neutrino interaction in the image or confuse the network, and therefore are a less important part of the background.

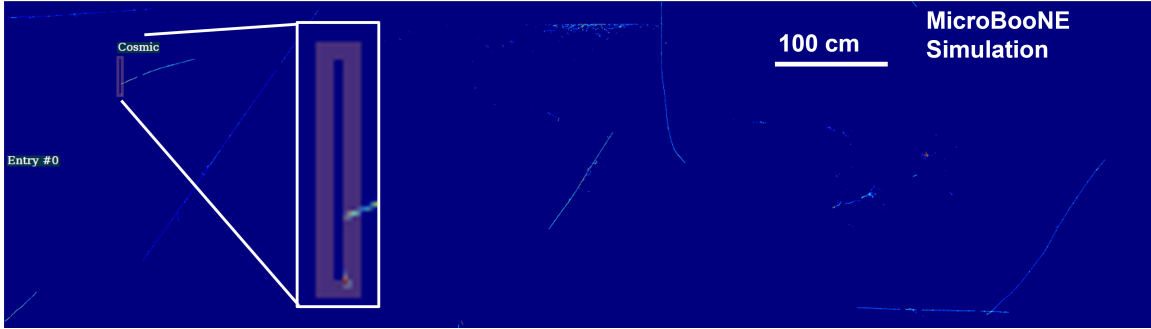


Figure 11: A zero-efficiency true interaction almost entirely in a region of unresponsive wires. The white box shows a zoom-in of the area of interest, and within it, the colored box should contain true neutrino interaction. However, because this interaction falls in an unresponsive region of the detector, no deposited charge is seen inside the colored box.

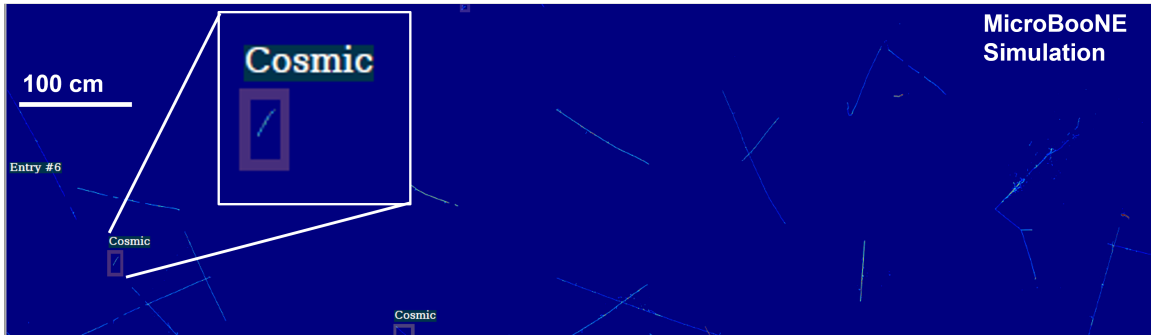


Figure 12: Another failure mode for the zero-efficiency peak. Here the cosmic interaction is relatively small in size compared to others in the event image. The colored box is the true interaction, and the white box shows a zoom-in of the area of interest.

4.2 Interaction coverage

Now that we have explored a pixel-wise efficiency, we next examine interaction coverage within a given event. We define a true interaction as being "covered" if its pixel-level efficiency as defined in section 4.1 is greater than 80%. This means the network has to cluster the majority of the interaction, while still leaving some room for error. Figure 13 compares the fraction of true interactions that are covered in a given event for both the dense and sparse implementations of Mask-RCNN. The dense network has a slight edge over the sparse version, but both networks consistently cover the majority of true interactions within a given event.

It is also useful to examine the performance of Mask-RCNN as a function of the number of true interactions in an event. This investigates whether the performance of the network falls off for 'busier' events with additional particle interactions cluttering up the image. To examine this, we look at the number of covered true interactions as a function of the number of true interactions in the events. Figure 14 shows this measurement for both the dense and sparse networks.

While the ideal network would cover all of the interactions, we see that both versions of the network produce a distribution slightly below the target line $y = x$. The fact that these distributions

Fractional Interaction Coverage

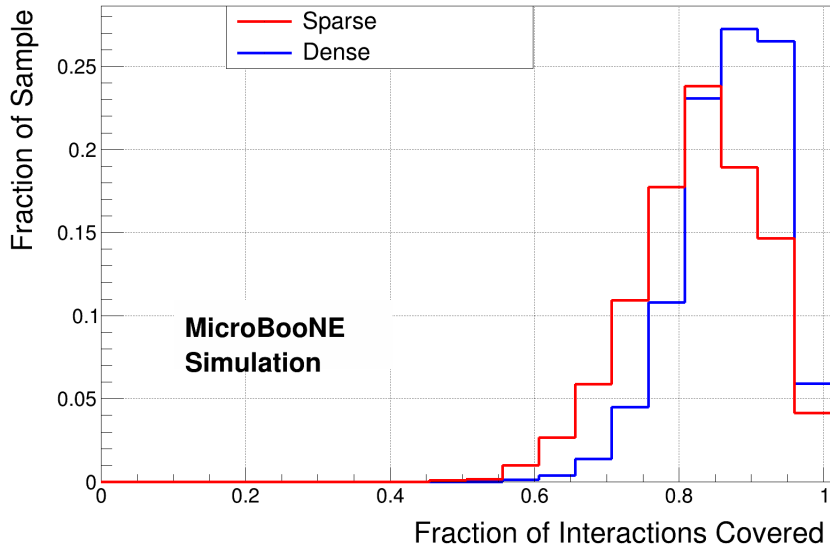
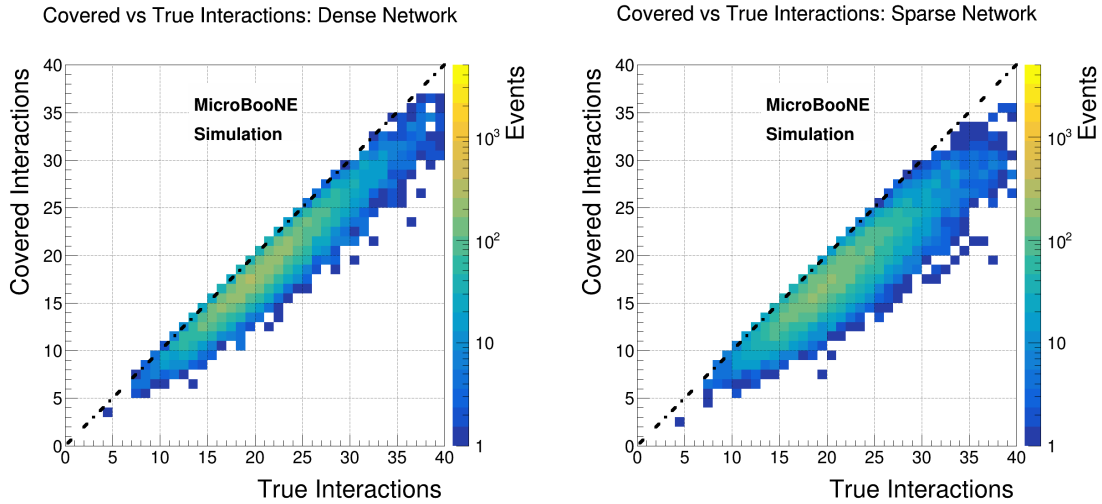


Figure 13: The fraction of true interactions in events that have greater than 80% efficiency as measured on the validation set.



(a) Dense Mask-RCNN

(b) Sparse Mask-RCNN

Figure 14: The number of covered interactions is plotted against the number of true interactions. A dashed line along $y = x$ represents the absolute perfect performance, with all true interactions being covered in each event. Out of an average of 20.8 true interactions per event, the dense network covers an average of 18.2, while the sparse network covers 17.2.

are linear demonstrates that the network performance does not diminish as the number of true interactions in a given event increases.

4.3 Network comparisons discussion

In comparing the dense and sparse versions across these various metrics, we find that the dense implementation performs better than the current version of sMask-RCNN. Viewing the two efficiencies of the networks, the peak at 1.0 for the dense Mask-RCNN is larger, and narrower than the peak in the sMask-RCNN distribution. Similarly the peak at zero-efficiency, representing interactions that are missed, is smaller. The dense Mask-RCNN surpasses sMask-RCNN in average efficiency as well with 89.1% compared to the slightly lower 85.9%. The dense network covers 87.1% of interactions compared to the sparse's 82.7%, where "covered" is defined in section 4.2.

It is important to note that both versions of the network completely miss true interactions with the same frequency. This is shown by observing the peak at 0 efficiency in figure 9. There we see that each network misses about 4% of interactions. This means that while the metrics point to worse performance for the sparse network, it still finds the interactions themselves, and still covers them to largely the same extent. The difference is that it builds less complete masks of the true interactions compared to the dense network, though it still finds part of them.

We note that it is difficult to track the effects of these differences on the training and learning of the networks. As such, we cannot distinguish whether the difference in performance of the two networks is due to the change from dense ResNet to sparse ResNet, the training on crops versus entire event images, or a combination of both. However, we emphasize that sMask-RCNN's ability to cluster interactions is sufficient for us to compare its cosmic tagging ability to current methods deployed in MicroBooNE, particularly given the speedup acquired by moving to submanifold convolutions. Further, while the dense network's performance does slightly outperform the sparse network, MicroBooNE's data processing prioritizes speed and the use of CPUs in order to scale to the size of its datasets. Therefore, deploying the dense network is not a viable option. This means that regardless of the performance of the dense version of the network, it is prohibitively slow to run at the scale MicroBooNE's dataset requires. Therefore for the analysis performed in section 5 we will only use sMask-RCNN.

5 Finding electron neutrinos with sMask-RCNN

In this section, we examine using sMask-RCNN in MicroBooNE to reduce the ratio of cosmic ray background events to electron neutrino events. There are two approaches that we explore, one designed to select neutrino interactions explicitly, and another designed to remove cosmic ray muon interactions. The first, an "identification by positive" approach, would use the neutrino-class output from sMask-RCNN and apply some threshold to select neutrino interactions. This approach is discussed in section 5.1.

The other approach, "identification by negative", applies an event veto, which targets cosmic-only events to flag them for removal. Then the remaining events are those with a neutrino, as well as cosmic ray muons interacting in the detector during the beam window. Finally, we note that while this article specifically targets electron neutrinos, the tools discussed can be adjusted to target muon neutrinos in the same way.

Recall that the cosmic ray background in MicroBooNE is very large, and therefore must be dealt with early on in any chain of reconstruction tools. To demonstrate the scale of this problem, we define three different samples:

Table 3: The expected ratio of the two different neutrino sample events to off-beam background events.

Type of event	Ratio to off-beam sample events
General electron neutrino	1.52×10^{-2}
Low energy electron neutrino	3.85×10^{-4}

1. **General Electron Neutrino Sample:** Events containing a simulated electron neutrino interaction combined with cosmic ray muon background data.
2. **Low Energy Electron Neutrino Sample:** The same as sample 1 but only for electron neutrinos with energy less than 400 MeV.
3. **Off-Beam Sample:** Data taken by the detector in anti-coincidence with the neutrino beam. This means there is no beam neutrino interaction present. This sample represents the cosmic ray-only background events.

The expected ratios of samples 1 and 2 to sample 3 are depicted in table 3. These ratios are instructive, as they indicate the initial signal to cosmic ray background event ratio for analyses that seek to remove cosmic ray-only events. These ratios only depict events. Each event contains $O(20)$ cosmic ray muon interactions, and either zero or one neutrino interaction. So the true ratio of cosmic ray muon interactions to electron neutrino interactions is roughly 20 times higher than the ratio of event types. As the purpose of this article is to develop techniques to reduce the significant cosmic ray muon background, we ignore other backgrounds to an electron neutrino signal, such as muon neutrino events.

While the analysis in section 5.1 uses the validation data used thus far in this article, section 5.2 uses the general and low energy electron neutrino samples and cosmic ray-only sample described above. The validation data contains only simulated interactions, including CORSIKA-simulated cosmic ray muons. However, the three new datasets contain cosmic ray muons from data, rather than simulation. In previous sections, where we need information about the individual cosmic ray muon interactions for the metrics, it is necessary to use simulated cosmic ray muons. However, this is not the case for the event veto described in section 5.2. Therefore it is better to use cosmic ray interactions from data, as there is no reliance on the simulation’s ability to properly emulate a cosmic ray muon.

5.1 Electron neutrino identification

We examine the electron neutrino "identification by positive" approach by looking at the efficiency and purity as defined in section 4.1, broken down by the two class categories: cosmic ray muons and electron neutrinos. Figure 15 shows the efficiency metric (defined in eq. 4.1) for simulated electron neutrino interactions, separated by class. The average efficiency is 76.8% for electron neutrinos, and 86.1% for cosmic ray muons. If the network proposes no neutrino interactions, then the efficiency of that event’s neutrino interaction is 0. Each class has a peak at 0, but the cosmics also have a peak at 100%, whereas the neutrino interactions peak at just over

90%. In the context of neutrino interactions, the interaction has some number of prongs, where a ‘prong’ refers to a shower, or track coming out of the neutrino interaction vertex. It is possible that, for neutrino interactions with at least two prongs, the network fails to mark a shorter prong, or partially masks the track-like portion of an electromagnetic shower with both track- and shower-like topology. Particularly long tracks are also difficult to capture completely due to rescaling within sMask-RCNN, which may lead to the ends of the track getting truncated.

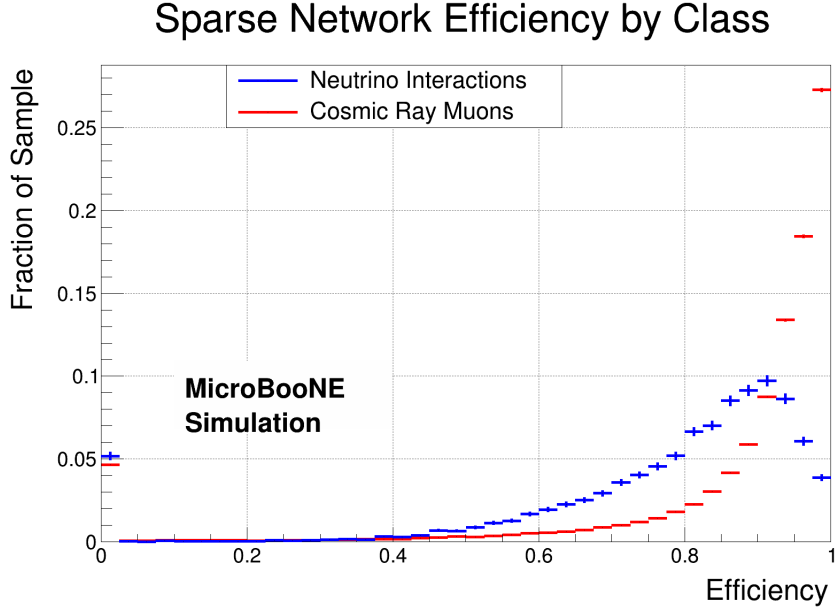


Figure 15: The efficiency of sMask-RCNN broken down by class. The average efficiency is 76.8% for electron neutrinos, and 86.1% for cosmic ray muons. Statistical uncertainty bars are shown.

Figure 16 is also separated by class, but shows the charge efficiency. The average charge efficiency for electron neutrinos is 77.9%, and 86.8% for cosmic ray muons. Both overall and for each class individually the network has a better charge efficiency than standard efficiency, indicating that the interaction masking prioritizes clustering pixels corresponding to larger charge deposition regardless of class.

We see from these two efficiency breakdowns that the network’s ability to find an interaction is not strongly tied to the type of interaction, as the peak at zero efficiency is the same for each class. However the masks for cosmic interactions are more complete than those for neutrino interactions.

The purity of sMask-RCNN predicted interactions (defined in eq. 4.2) is broken down by class in figure 17. The average purity is 64.9% for electron neutrinos and 84.7% for cosmic ray muons. Here we see an issue with using this version of sMask-RCNN in an identification by positive approach. The peak at zero purity for the neutrino class indicates that, in events that contain simulated neutrino interactions, roughly 22% of predicted interactions labeled neutrinos are actually placed on cosmic ray muons. This implies selecting only predicted neutrino interactions yields a ratio of electron neutrinos to cosmic ray muons of 78:22. However, this only applies if the identification were restricted to events that definitely contain a neutrino. When factoring in the significant number of events in the data that contain no neutrino interaction, as indicated by table 3, the number of falsely

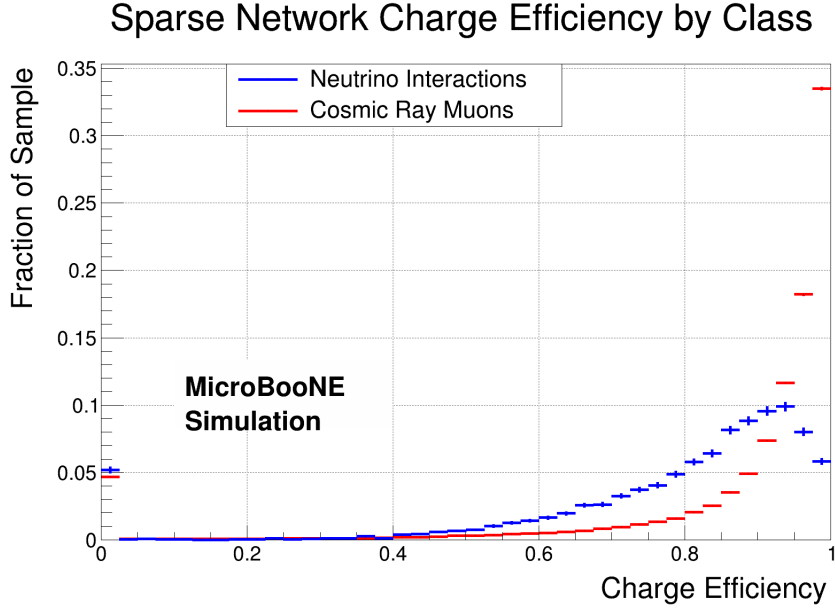


Figure 16: The charge efficiency of sMask-RCNN broken down by class. The average charge efficiency for electron neutrinos is 77.9%, and 86.8% for cosmic ray muons. Statistical uncertainty bars are shown.

identified neutrinos grows much worse.

We can imagine ways to improve this identification by positive approach, from increasing the required confidence score the classifier in sMask-RCNN has in a predicted interaction, to retraining the network with increased penalties for falsely predicting the neutrino class. However, these were set aside in favor of exploring the identification by negative approach in the subsequent section.

5.2 Cosmic-only event veto

The topology of a muon interaction is much more consistent than that of electron neutrino interactions, due to an electron neutrino interaction’s variety of final states. A muon creates a track in the detector, and then will either pass through the detector, capture, or create a Michel electron shower at the end, whereas an electron neutrino can interact in argon in many different ways, creating a variety of different daughter particle scenarios. This is consistent with the improved performance of the network labeling cosmic ray muons compared to electron neutrinos, evident in section 5.1, where the network must learn to recognize the many different patterns and topologies that make up the electron neutrino class label. Therefore, relying on sMask-RCNN’s cosmic ray muon clustering rather than its neutrino clustering may be preferable. As such, the identification by negative approach, which only relies on an understanding of the cosmic ray background, may be more effective.

To study this, we implement an event veto. The goal of this veto is to use sMask-RCNN outputs to separate entire events into those that contain only cosmic ray background, and those that contain an electron neutrino interaction among cosmic rays in the beam window. This task is tested by using this event veto to separate the cosmic ray-only data sample from the general and low energy

Sparse Network Purity by Class

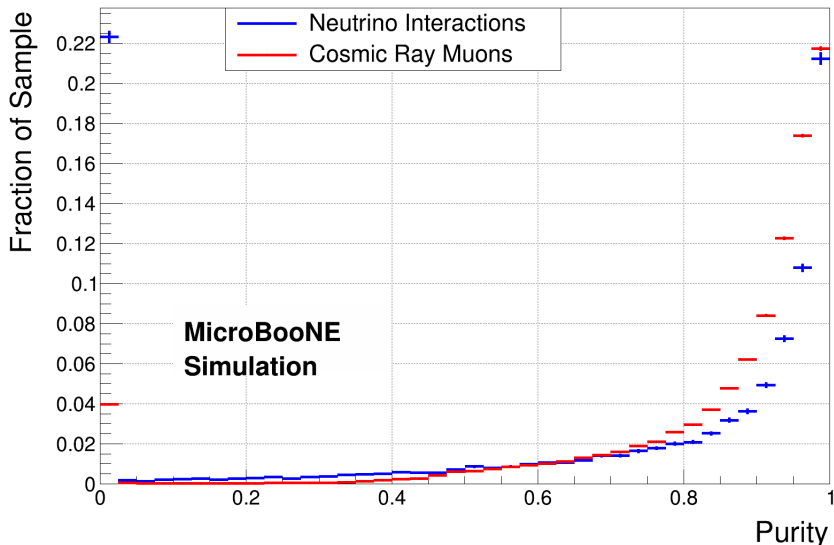


Figure 17: The purity of sMask-RCNN broken down by class. Predicted interactions with zero purity are misclassified, for example a true neutrino labeled as a cosmic ray. The average purity is 64.9% for electron neutrinos and 84.7% for cosmic ray muons. Statistical uncertainty bars are shown.

electron neutrino samples. If this veto were perfect, then the only cosmic ray muons left would be the $O(20)$ interactions per event containing a neutrino interaction. These remaining cosmic ray muons can then be dealt with further down the reconstruction chain.

In order to provide a comparison to current methods used in MicroBooNE, we analyze this event veto using several different versions of cosmic ray tagging. For the first tagger, we include all cosmic ray interaction pixels predicted by sMask-RCNN with a confidence score greater than 0.20. Reducing the confidence score requirement relative to earlier sections of this article allows more cosmic removal at the expense of including multiple overlapping cosmic interaction predictions. Decreasing the confidence score requirement does not significantly impact the electron neutrino efficiency. For individual interaction labeling, shown earlier in the article, this would be problematic. However, for the event veto described below, we are concerned with removing entire events.

For the next tagger, we add the pixels tagged as cosmic ray muons by MicroBooNE’s Wire-Cell Q-L described in section 2.2. This adds information from the PMT light collection system and the two LArTPC induction planes, none of which is used by sMask-RCNN. Therefore by comparing sMask-RCNN alone to this combined tagger, we can see the additional value provided by the light information, as well as demonstrate the effectiveness of sMask-RCNN operating in a regime with less information. Recall WC Q-L matching is a piece of the full WC cosmic tagger. We repeat the following analysis with the full WC cosmic tagger at the end of this section, exploring two additional tagger configurations.

A perfect tagger would tag every pixel containing deposited charge associated with cosmic ray muons. All that would remain in the event image would be pixels holding charge corresponding

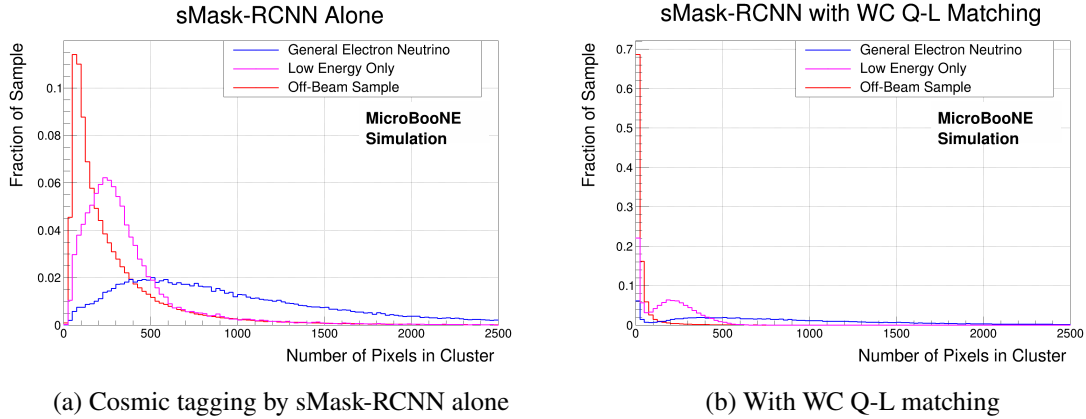


Figure 18: The size of the largest cluster found by DBScan after cosmic tagging.

to a neutrino interaction, if present. However, with the expectation of imperfect performance, we re-cluster the untagged pixels via a "density-based spatial clustering of applications with noise" (DBScan) algorithm [26]. This means that first we perform our cosmic ray muon tagging, remove those pixels from the image, then run DBScan on the resulting image.

DBScan will output clusters of remaining pixels for each event. These pertain to portions of cosmic ray muon interactions not fully tagged, and the neutrino interaction if present. In the case where both neutrino and cosmic ray muon clusters are present, one large cluster will usually represent the neutrino interaction and several smaller clusters represent the untagged parts of muons. This means the size of the largest cluster is a metric that we can use to isolate events containing neutrino interactions. Figure 18 shows the size of the largest of these clusters for sMask-RCNN with and without the WC Q-L matching algorithm. Each figure shows the distribution of the three key samples described above.

Examining the distributions of these three samples, we see a difference between the off-beam sample, which represents the cosmic ray-only event background, and the two electron neutrino samples. Notably, the off-beam sample generally has a smaller number of pixels in the largest cluster found by DBScan. In figure 18a, we see that the sMask-RCNN tagger produces a distinctly shaped distribution for each sample. The off-beam sample has a peak closest to zero, the low energy sample is shifted slightly to the right, and the general electron neutrino sample is shifted further. This reinforces the expectation, as the neutrino pixels remain in the image untagged, allowing for DBScan to find larger clusters. Examining sMask-RCNN with the WC Q-L matching algorithm in figure 18b, we see a shift to the left in all three distributions, placing a strong peak at zero, indicating that the combined tagger frequently labels more pixels as belonging to cosmic ray muon interactions than the sMask-RCNN-based tagger alone. This is particularly notable in the case of the off-beam sample, where the peak at zero accounts for almost 70% of the sample, compared to about 16% of the low energy sample, and about 6% of the general electron neutrino sample.

Using this metric for the size of the largest remaining cluster, we can create a receiver operating characteristic (ROC) curve to demonstrate the efficiency and rejection power of the different cosmic tagging methods when applied as an event veto. A ROC curve is a measure of signal retention or signal efficiency on one axis, and background rejection on the other. A curve is created by

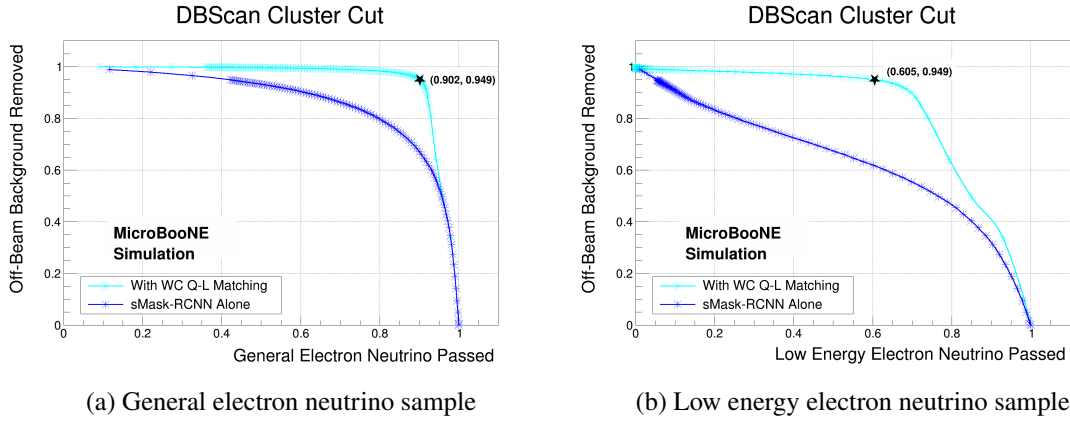


Figure 19: ROC curves for the sMask-RCNN with and without WC Q-L matching based on a requirement on largest cluster size. Curves are shown for the two electron neutrino samples against the off-beam background.

incrementing some requirement, which slowly decreases retention and increases rejection. Ideally the curve has points in the upper right region of the plot such that signal retention and background rejection are both high.

We reiterate that we want to explore this quantity — largest remaining cluster — as an event-level discriminant. As such, it will be the value we increment to create ROC curves. Specifically, these curves are made by applying a requirement to filter events that do not contain a DBScan cluster of size greater than X , where X is incremented from no requirement, to 0, and then incremented by 10 pixels thereafter. The resulting signal retention and background rejection rates give points for the curves. In figure 19, we show the ROC curves for the signal retentions of the general and low energy electron neutrino samples versus the rejection of the off-beam cosmic ray muon background sample. Curves are made for the two different tagging methods. We observe that the combined version of a cosmic tagger using both sMask-RCNN and WC Q-L matching yields a better combination of signal efficiency and background rejection.

For the low energy sample each tagging method performs worse compared to the general sample. However, this is not surprising as the lower energy electron neutrino interactions correspond to less charge deposited in the event image and fewer neutrino pixels in the event. This means that the remaining DBScan clusters related to the lower energy neutrino interactions will be smaller, and harder to isolate from the untagged off-beam cosmic ray muon sample’s distribution.

Examining the ROC curve for sMask-RCNN with WC Q-L matching in figure 19a, we are able to achieve a general electron neutrino signal efficiency of 90.2% while rejecting 94.9% of the off-beam cosmic background if we remove events that do not have a cluster of at least 130 pixels after the taggers are run. For the low energy electron neutrino sample in figure 19b we can achieve a signal efficiency of 60.5% for the same requirement, though a reduced requirement on remaining cluster size could be applied to increase the efficiency at the cost of rejection power, as indicated by the combined curve.

We also examine the effect of adding sMask-RCNN’s cosmic finding to the state-of-the-art complete WC cosmic tagger. This means that we take the event vetoes and cosmic tagging of the

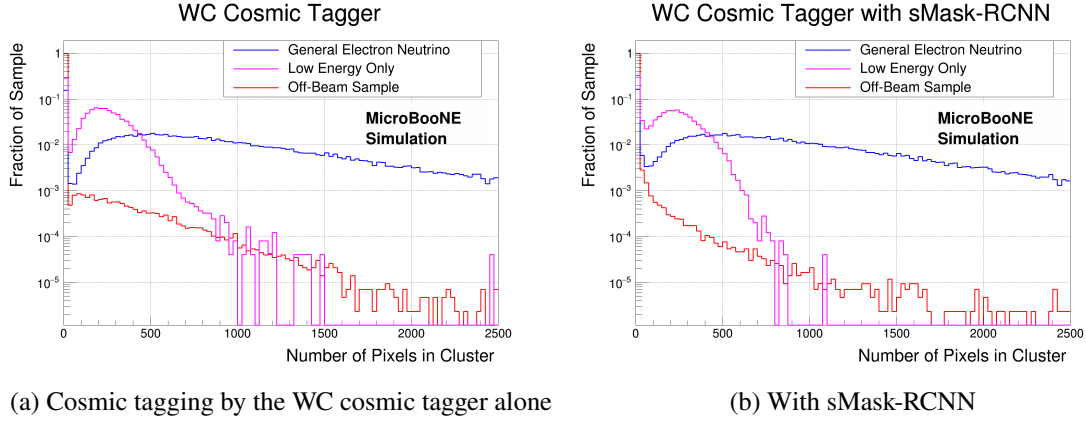


Figure 20: The size of the largest cluster found by DBScan after cosmic tagging. Note the logarithmic scale.

WC cosmic tagger, and add the cosmic tagging of sMask-RCNN to get a combined tagger. In figure 20 we show the distributions of the largest cluster found by DBScan after running the WC cosmic tagger with and without the cosmic ray muons found by sMask-RCNN. For the events that are rejected by one of the WC cosmic tagger event vetos, the largest DBScan cluster is defined to be zero.

The distributions of cluster size before and after adding sMask-RCNN to the WC cosmic tagger show a shift to the left in the shape of the off-beam sample, indicating the added value of sMask-RCNN in cosmic tagging. Without sMask-RCNN, there appears to be a slight peak beyond zero that gets shifted to zero after sMask-RCNN is added. We observe minimal shift in the two electron neutrino samples and each distribution still has a clear second peak separate from zero.

ROC curves for the WC cosmic tagger with and without sMask-RCNN are shown in figure 21. However, as the WC cosmic tagger introduces several of its own event vetos, the point referring to the loosest cut, with the most signal passed, does not allow all events through the veto. Instead it starts with the signal efficiency and background rejection of the WC cosmic tagger, and adjusts as we increase the strength of the DBScan cluster size requirement. We reiterate that the difference between figures 21 and 19 is the additional event vetos added to WC Q-L matching to create the WC cosmic tagger.

Examining these ROC curves we see that including sMask-RCNN on top of the WC cosmic tagger does yield improvement to the background rejection at equivalent signal efficiencies. In the general electron neutrino sample, for a signal efficiency of 80.1% the WC cosmic tagger rejects 99.0% of the background, whereas adding sMask-RCNN rejects 99.7% of the background at the same signal efficiency. This represents a reduction of the remaining background by 70%. For the low energy neutrino sample, a similar effect is seen, albeit at lower signal efficiencies.

In order to evaluate the difference between these two rejection strengths, it is important to recall the imbalance between off-beam background events and electron neutrino signal events described in table 3. The 99.0% rejection provided by the WC cosmic tagger corresponds to a general electron neutrino signal to off-beam background ratio of 1.26, a vast improvement over the starting ratio in the table. However, the addition of sMask-RCNN increases this signal to background

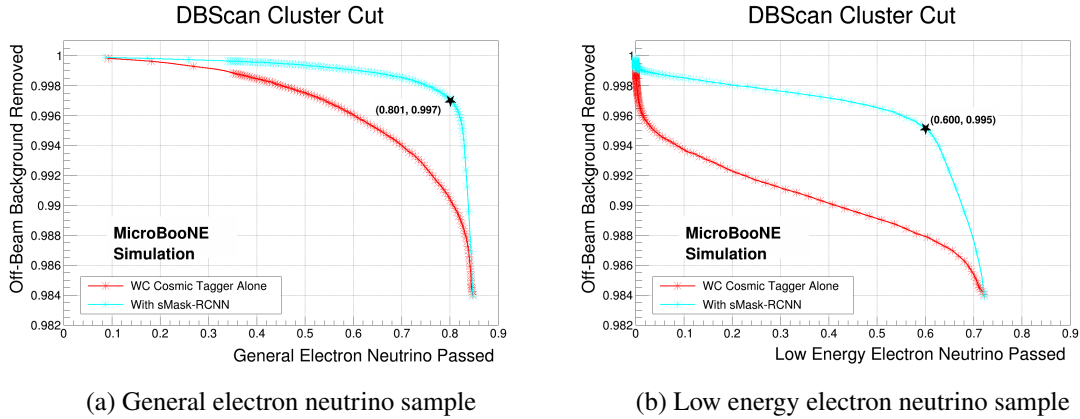


Figure 21: ROC curves for the WC cosmic tagger with and without sMask-RCNN based on a requirement on largest cluster size. Curves are shown for the two electron neutrino samples against the off-beam background. Note the suppressed y-axis shown here demonstrates significant improvement in background removal compared to figure 19.

ratio to 4.14 by improving the rejection power to 99.7%. For this same selection, the low energy signal to background ratio is 0.015 with the WC cosmic tagger, and 0.56 with after the addition of sMask-RCNN.

6 Conclusions

This article demonstrates a novel approach to cosmic ray muon tagging using sMask-RCNN. We demonstrate the ability of this network to locate, identify, and cluster particle interactions in the MicroBooNE LArTPC. We analyze the ability to cluster both the topologically simple cosmic ray muon interactions, as well as highly varied electron neutrino interactions.

We modify the original Mask-RCNN framework by substituting sparse submanifold convolutions in the ResNet portion of the network to create sMask-RCNN. Due to the low pixel occupancy of MicroBooNE event image data this leads to a 20 \times speedup in ResNet processing time on a CPU, as well as decreased runtime memory usage. This improvement is critical in allowing sMask-RCNN to be deployed as a reconstruction tool on CPU farms to scale to high volume data samples that particle physics experiments typically employ.

This analysis also includes several versions of an event veto. The strongest of these demonstrates that adding sMask-RCNN to the state-of-the-art WC cosmic tagger which is currently used in MicroBooNE reduces the cosmic ray-only event background by a further 70% and increases the signal to background ratio of electron neutrino events to cosmic ray-only events by more than a factor of three. This means that application of this technique to future measurements in MicroBooNE will result in improvements over current MicroBooNE reconstruction.

Acknowledgments

This document was prepared by the MicroBooNE collaboration using the resources of the Fermi National Accelerator Laboratory (Fermilab), a U.S. Department of Energy, Office of Science, HEP

User Facility. Fermilab is managed by Fermi Research Alliance, LLC (FRA), acting under Contract No. DE-AC02-07CH11359. MicroBooNE is supported by the following: the U.S. Department of Energy, Office of Science, Offices of High Energy Physics and Nuclear Physics; the U.S. National Science Foundation; the Swiss National Science Foundation; the Science and Technology Facilities Council (STFC), part of the United Kingdom Research and Innovation; the Royal Society (United Kingdom); and The European Union’s Horizon 2020 Marie Skłodowska-Curie Actions. Additional support for the laser calibration system and cosmic ray tagger was provided by the Albert Einstein Center for Fundamental Physics, Bern, Switzerland. We also acknowledge the contributions of technical and scientific staff to the design, construction, and operation of the MicroBooNE detector as well as the contributions of past collaborators to the development of MicroBooNE analyses, without whom this work would not have been possible.

7 Bibliography

References

- [1] R. Acciarri *et al.* (MicroBooNE Collaboration), *Design and Construction of the MicroBooNE Detector*, JINST **12** (2017) P02017, [arXiv:1612.05824]
- [2] R. Acciarri *et al.* (MicroBooNE Collaboration), *Convolutional neural networks applied to neutrino events in a liquid argon time projection chamber*, JINST **12** (2017) P03011, [arXiv:1611.05531]
- [3] C. Adams *et al.* (MicroBooNE Collaboration), *Deep neural network for pixel-level electromagnetic particle identification in the MicroBooNE liquid argon time projection chamber*, Phys. Rev. D **99** (2019) 092001, [arXiv:1808.07269]
- [4] P. Abratenko *et al.* (MicroBooNE Collaboration), *Semantic segmentation with a sparse convolutional neural network for event reconstruction in MicroBooNE*, Phys. Rev. D **103** (2021) 052012, [arXiv:2012.08513]
- [5] P. Abratenko *et al.* (MicroBooNE Collaboration), *Convolutional neural network for multiple particle identification in the MicroBooNE liquid argon time projection chamber*, Phys. Rev. D **103** (2021) 092003, [arXiv:2010.08653]
- [6] K. He, G. Gkioxari, P. Dollár, and R. Girshick, *Mask R-CNN*, IEEE International Conference on Computer Vision **1** (2017) 2980-2988, [arXiv:1703.06870]
- [7] B. Graham, *Spatially-sparse convolutional neural networks*, CoRR (2014) [arXiv:1409.6070]
- [8] C. Adams *et al.* (MicroBooNE Collaboration), *Ionization Electron Signal Processing in Single Phase LArTPCs I. Algorithm Description and Quantitative Evaluation with MicroBooNE Simulation*, JINST **13** (2018) P07006, [arXiv:1802.08709]
- [9] C. Adams *et al.* (MicroBooNE Collaboration), *Ionization Electron Signal Processing in Single Phase LArTPCs II. Data/Simulation Comparison and Performance in MicroBooNE*, JINST **13** (2018) P07007, [arXiv:1804.02583]
- [10] R. Acciarri *et al.* (SBND Collaboration), *Cosmic Background Removal with Deep Neural Networks in SBND*, Frontiers in Artificial Intelligence **4** (2021) 42, [arXiv:2012.01301]
- [11] R. Acciarri *et al.* (MicroBooNE Collaboration), *The Pandora multi-algorithm approach to automated pattern recognition of cosmic-ray muon and neutrino events in the MicroBooNE detector*, Eur. Phys. J. C **78** (2018) 82, [arXiv:1708.03135]

- [12] R. Acciarri *et al.* (MicroBooNE Collaboration), *The Pandora Software Development Kit for Pattern Recognition*, Eur. Phys. J. C **75** (2015) 439, [arXiv:1506.05348]
- [13] P. Abratenko *et al.* (MicroBooNE Collaboration), *Cosmic Ray Background Rejection with Wire-Cell LArTPC Event Reconstruction in the MicroBooNE Detector*, Phys. Rev. Applied **15** (2021) 064071, [arXiv:2101.05076]
- [14] P. Abratenko *et al.* (MicroBooNE Collaboration), *High-performance Generic Neutrino Detection in a LArTPC near the Earth’s Surface with the MicroBooNE Detector*, FERMILAB-PUB-20-666-ND (2020) [arXiv:2012.07928]
- [15] P. Abratenko *et al.* (MicroBooNE Collaboration), *Neutrino Event Selection in the MicroBooNE Liquid Argon Time Projection Chamber using Wire-Cell 3-D Imaging, Clustering, and Charge-Light Matching*, JINST **16** (2021) P06043, [arXiv:2011.01375]
- [16] C. Adams *et al.* (MicroBooNE Collaboration), *Design and construction of the MicroBooNE Cosmic Ray Tagger system*, JINST **14** (2019) P04004, [arXiv:1901.02862]
- [17] K. He, X. Zhang, S. Ren, and J. Sun, *Deep Residual Learning for Image Recognition*, IEEE Conference on Computer Vision and Pattern Recognition **1** (2016) 770–778, [arXiv:1512.03385]
- [18] S. Ren, K. He, R. Girshick, and J. Sun, *Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks*, IEEE Transactions on Pattern Analysis and Machine Intelligence **39** (2017) 1137-1149, [arXiv:1506.01497]
- [19] R. Acciarri *et al.* (MicroBooNE Collaboration), *Noise Characterization and Filtering in the MicroBooNE Liquid Argon TPC*, JINST **12** (2017) P08003, [arXiv:1705.07341]
- [20] Fermilab , *Grid Computing*, <https://www.fnal.gov/pub/science/computing/grid.html>, (2015) Accessed: 2021-09-17 [arXiv:1612.05824]
- [21] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, *ImageNet: A large-scale hierarchical image database*, IEEE Conference on Computer Vision and Pattern Recognition **12** (2009) 248-255
- [22] D. Heck, J. Knapp, J. N. Capdevielle, G. Schatz, and T. Thouw, *CORSIKA: A Monte Carlo code to simulate extensive air showers*, FZKA-6019 (1998)
- [23] C. Andreopoulos, C. Barry, S. Dytman, H. Gallagher, T. Golan, R. Hatcher, G. Perdue, and J. Yarba, *The GENIE Neutrino Monte Carlo Generator: Physics and User Manual* (2015), [arXiv:1510.05494]
- [24] S. Agostinelli *et al.*, *GEANT4—a simulation toolkit*, Nucl. Instrum. Meth. A **506** (2003) 250-303
- [25] A. Kirillov *et al.*, *Panoptic Segmentation*, CVPR (2019) [arXiv:1801.00868]
- [26] M. Ester, H. P. Kriegel, J. Sander, and X. Xiaowei, *A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise*, in KDD’96: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (AAAI Press) (1996) 226–231