
Faculty of Science

Faculty Publications

This is a post-print version of the following article:

Energy reconstruction in a liquid argon calorimeter cell using convolutional neural networks

Luke Polson, Leonid Kurchaninov, Michel Lefebvre

2022

The final publication is available at:

<https://doi.org/10.1088/1748-0221/17/01/P01002>

Citation for this paper:

Ponson, L., Kurchaninov, L., and Lefebvre, M. (2022). Energy reconstruction in a liquid argon calorimeter cell using convolutional neural networks. *Journal of Instrumentation* vol. 17. <https://doi.org/10.1088/1748-0221/17/01/P01002>

Energy reconstruction in a liquid argon calorimeter cell using convolutional neural networks

L. Polson,^a L. Kurchaninov^b and M. Lefebvre^a

^a*University of Victoria,
3800 Finnerty Road, Victoria, Canada*

^b*TRIUMF,
4004 Wesbrook Mall, Vancouver, Canada*

E-mail: lpolson@uvic.ca

ABSTRACT: The liquid argon ionization current in a sampling calorimeter cell can be analyzed to determine the energy of detected particles. In practice, experimental artifacts such as pileup and electronic noise make the inference of energy from current a difficult process. The beam intensity of the Large Hadron Collider will be significantly increased during the Phase-II long shut-down of 2025-2027. Signal processing techniques that are used to extract the energy of detected particles in the ATLAS detector will suffer a significant loss in performance under these conditions. This paper compares the presently used optimal filter technique to convolutional neural networks for energy reconstruction in the ATLAS liquid argon hadronic end cap calorimeter. In particular, it is shown that convolutional neural networks trained with an appropriately tuned and novel loss function are able to outperform the optimal filter technique.

KEYWORDS: signal processing, convolutional neural networks, liquid argon calorimeter, ATLAS detector

Contents

1	Introduction	1
2	Simulation of Current in a HEC Calorimeter Cell	1
3	Energy Reconstruction Using the Optimal Filter Technique	2
4	Energy Reconstruction Using a CNN	3
4.1	Using an RMSE Loss Function	4
4.2	Using an Alternative Loss Function	4
5	Conclusion	7

1 Introduction

The inference of energy from measured current in ionization detectors is often a non-trivial process, particularly in the regime of high energy particle physics. The conditions in the liquid argon (LAr) calorimeter of the ATLAS detector [1–3] expected after the high luminosity (HL) upgrade to the Large Hadron Collider (LHC) [3, 4] are such that the precision of energy measurements will suffer a loss in performance. This paper seeks to simulate the conditions of the HL-LHC and the electronic response of a particular detector cell in the hadronic end cap (HEC) subsystem of the ATLAS LAr Calorimeter. The presently used algorithm for energy reconstruction from measured signal, known as the optimal filter (OF) [5], will be compared to convolutional neural network (CNN) architectures. It will be shown that a model architecture trained with an appropriate loss function on simulated data outperforms the optimal filter in relevant metrics.

2 Simulation of Current in a HEC Calorimeter Cell

The total energy deposited in a HEC detector cell E is given by the sum of two different components: pileup p , which includes deposited energy from multiple low-energy collisions and is proportional to the luminosity of the LHC, and signal S , which is representative of particular and rare processes. The energy can be expressed as a function of discrete time t , with 25 ns or 1 bunch crossing (BC) spacing:

$$E(t) = S(t) + p(t) . \quad (2.1)$$

In this study, $p(t)$ is simulated according to [6], where the total pileup energy at a given BC is the sum of energies from many minimum bias events. The probability density function for the energy of a single minimum bias event is given by

$$P(E) = q\delta(E) + (1 - q) \left(F_1 e^{-\sqrt{E/E_1}} + F_2 e^{-\sqrt{E/E_2}} \right) , \quad (2.2)$$

where $\delta(E)$ is the Dirac delta function and the expression is written such that q is the probability of zero energy deposition. The parameters used are $F_1 = 1.721 \text{ GeV}^{-1}$, $E_1 = 1.012 \text{ GeV}$, $F_2 = 0.0226 \text{ GeV}^{-1}$, $E_2 = 1.012 \text{ GeV}$,

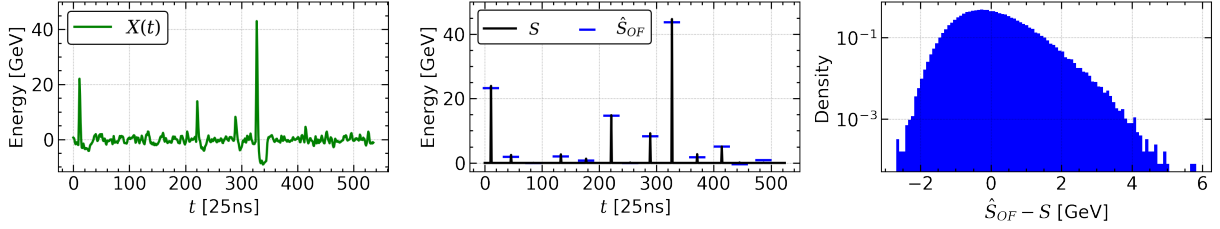


Figure 1. Simulated current $X(t)$ representative of the ATLAS experiment [left], signal and optimal filter estimation of the signal [middle] and histogram of residuals $S - \hat{S}_{OF}$ for all 30 million BCs [right]. For display purposes, $X(t)$, S , and \hat{S}_{OF} are aligned so that the peaks coincide.

and $q = 0.8314$; they correspond to the expected pileup conditions at $\eta = 2.25$ in the first layer of the HEC as obtained through simulations in [6] and were provided by one of the authors of this note¹. The number of minimum bias events per BC used for simulation is a Poisson random variable with mean $\langle \mu \rangle = 200$ – the expected value at the HL-LHC [4]. These energies are summed together at each BC time t to obtain the total pileup $p(t)$.

$S(t)$ is simulated such that consecutive ionization pulses are non-overlapping²; specifically, the spacing between non-zero values is a random variable $\Gamma \sim \mathcal{U}(30 \text{ BC}, 50 \text{ BC})$ where $\mathcal{U}(a, b)$ is the uniform distribution between values a and b . The distribution of energies during signal injection is given by

$$f(s) = \frac{1}{3} \left[\mathcal{U}(0 \text{ GeV}, 0.3 \text{ GeV}) + \mathcal{U}(0.3 \text{ GeV}, 5 \text{ GeV}) + \mathcal{U}(5 \text{ GeV}, 50 \text{ GeV}) \right], \quad (2.3)$$

designed to consider relevant signal energies for the HL-LHC.

The total energy E deposited in a HEC detector cell at a given BC produces a triangular-shaped ionization drift current pulse which, upon convolution with the electronics chain response, results in the measurable current $X(t)$. The electronics chain is described in [7, 8], and features a bipolar shaper. Sample time series of $X(t)$ and $S(t)$ are shown in the left and middle subplot of Figure 1, respectively.

3 Energy Reconstruction Using the Optimal Filter Technique

The signal energy is obtained by inferring $S(t)$ from $X(t)$ and is presently obtained using the OF [5], which produces an estimator $\hat{S}_{OF}(t)$ given by a weighted sum on $X(t)$ with N coefficients a_i :

$$\hat{S}_{OF}(t + N) = \sum_{i=1}^N a_i X(t + i). \quad (3.1)$$

All time series used have a length of 30 million BCs. $N = 25$ coefficients are used in the OF; previous studies have shown there is no significant performance gain with more coefficients [9]. Shown in Figure 1 is a sample of the measurable current $X(t)$ [left], the signal $S(t)$ and optimal filter predictions \hat{S}_{OF} when signal is injected [middle], and a histogram of the residuals between the true and predicted signals for the entire time series [right].

¹The statistical uncertainty of these parameters is small and sufficient to ensure an uncertainty on $P(E)$ that is smaller than the uncertainty of the electronics chain modeling, which is at the level of 2%.

²Consecutive ionization pulses induced by $S(t)$. In general, ionization pulses from pileup $p(t)$ will be overlapping.

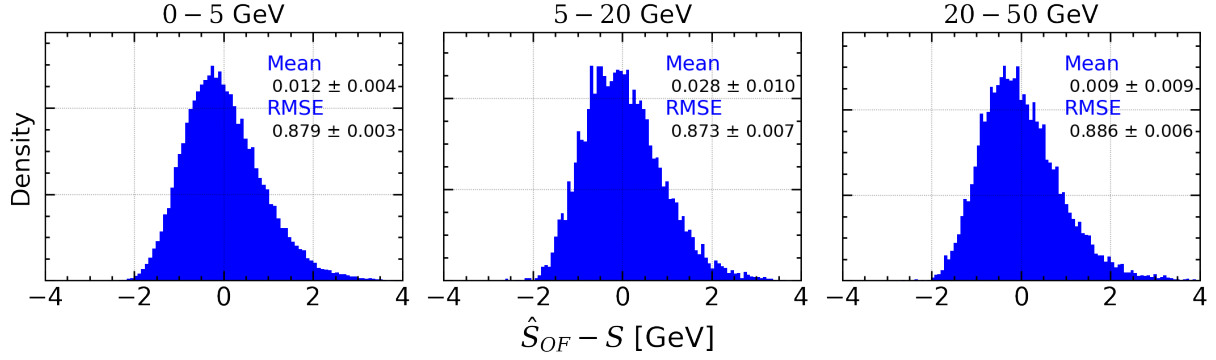


Figure 2. Histograms of $\hat{S}_{OF} - S$ for different signal magnitudes S , specified by the title of each subplot. The mean and RMSE are shown on each subplot.

Subset	P_1	P_2	P_3	P_4	P_5	P_6	P_7
Values [GeV]	0-0.3	0.3-1	1-2	2-5	5-10	10-20	20-50

Table 1. Signal partitions.

A key feature of the OF is its ability to estimate $S(t)$ with little bias in separate energy regions: this is apparent in Figure 2, where residuals are separated based on signal magnitude. Such an estimator is said to have no energy range bias; the extent to which an estimator \hat{S} makes locally biased predictions can be quantified by the following metric

$$B_L(S, \hat{S}) = \sum_{j=1}^n \left| \frac{1}{N_j} \sum_{S \in P_j} (\hat{S}(\theta) - S) \right|, \quad (3.2)$$

where θ represents the parameters of the estimator, P_j is an interval of energies and N_j is the number of events that lie in the interval. Desirable features of any estimator \hat{S} are small B_L and RMSE values. The P_j 's used are given by Table 1.

4 Energy Reconstruction Using a CNN

This section provides evidence that a CNN [10] architecture may outperform the OF for energy reconstruction in a LAr detector cell. A CNN architecture is chosen because it has been shown to be competitive with other types of Machine Learning architectures [11], it is relatively fast to train on systems with a graphical processing unit, and it could likely be implemented on field programmable gate arrays (FPGAs) [12] that will be deployed in ATLAS in the Phase-II upgrade [13]. All models considered have less than 100 parameters in view of FPGA implementation, though the feasibility of implementation is not guaranteed by this constraint.

Models used were based on a WaveNet architecture [11] with relatively few layers and filters, and small kernel sizes. A single convolutional layer with a filter size of 3 was used as the final layer for each model. The filter depth, which is defined as the number of input data points used to make a prediction for a single output data point, is dependent on the model hyperparameters. Similar to the OF, the models used were not causal and were able to use 11 future BCs when making predictions. All models were trained for 2000 epochs. Regularization and cross validation were not considered necessary; models did not overfit as the size of the training data was

Hyperparameters				Results		
Dilations	Filters	Kernel size	Filter Depth	Rank	RMSE [GeV]	B_L [GeV]
(1,1)	3	7	13	1	0.7294 ± 0.0002	1.117 ± 0.015
(1, 1, 2, 2)	2	5	25	2	0.7300 ± 0.0004	1.118 ± 0.007
(1, 1, 1, 2)	2	5	21	3	0.7308 ± 0.0005	1.120 ± 0.009
(1, 2)	3	7	19	4	0.7318 ± 0.0006	1.080 ± 0.020
(1, 1, 2)	2	7	25	5	0.7321 ± 0.0003	1.131 ± 0.005
(1, 3)	3	7	25	6	0.7323 ± 0.0007	1.098 ± 0.017
(1, 2)	3	5	13	7	0.7324 ± 0.0008	1.067 ± 0.014
(1, 3)	3	5	17	8	0.7341 ± 0.0004	1.081 ± 0.006
(1, 1, 3)	2	5	21	9	0.7357 ± 0.0002	1.075 ± 0.006
(1, 3)	4	3	9	11	1.0678 ± 0.0014	2.022 ± 0.017
Optimal Filter				10	0.8810 ± 0.0016	0.108 ± 0.021

Table 2. Hyperparameter search: the ranking is based on the RMSE metric. The OF metrics are shown in the bottom row.

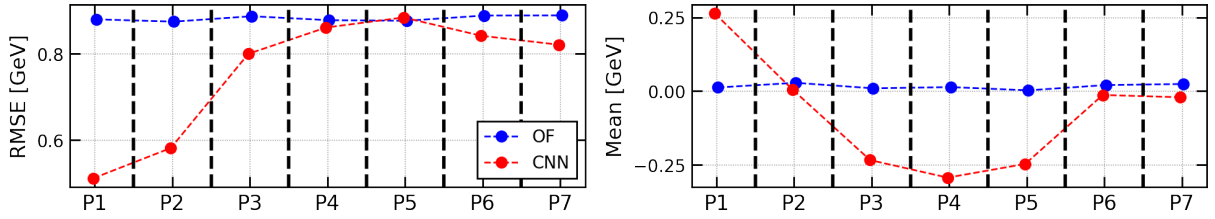


Figure 3. RMSE and Mean of residuals $\hat{S} - S$ for the top model architecture of Table 2 and for each of the subsets P_j in Table 1.

large and the models contained few parameters. The models were trained on 80% of the data and evaluated on the other 20%.

4.1 Using an RMSE Loss Function

Table 2 shows the results of a hyperparameter search where models are trained using an RMSE loss function and a Nadam variation of gradient descent [14]. All models have less than 100 parameters. While the CNNs tend to make predictions with lower RMSE than those of the OF, the large B_L values are indicative of large local biases in CNN predictions. This is visible, for example, in the predictions of the CNN model architecture with the lowest RMSE score, shown in Figure 3. The CNN is able to make predictions with a lower RMSE than the OF, especially at lower energies, but it also tends to overpredict small energies and underpredict intermediate energies. Corresponding histograms of residuals $\hat{S} - S$ in each region P_j are shown in Figure 4. Compared to the OF, over all reconstructed signals \hat{S} , the RMSE obtained with the CNN is $(17.2 \pm 0.2)\%$ lower.

4.2 Using an Alternative Loss Function

To prevent CNNs from being optimized in such a way that they make locally biased predictions, the following loss function is proposed

$$L(\theta) = \text{RMSE}(\hat{S}(\theta), S) + \alpha \sum_{j=1}^n \left(\frac{1}{N_j} \sum_{S \in P_j} (\hat{S}(\theta) - S) \right)^2, \quad (4.1)$$

where RMSE is the standard RMSE loss function and α is an adjustable constant. In Figure 5, it is shown that a proper adjustment of α results in a CNN predictor with smaller RMSE and B_L than the OF. For small α , the

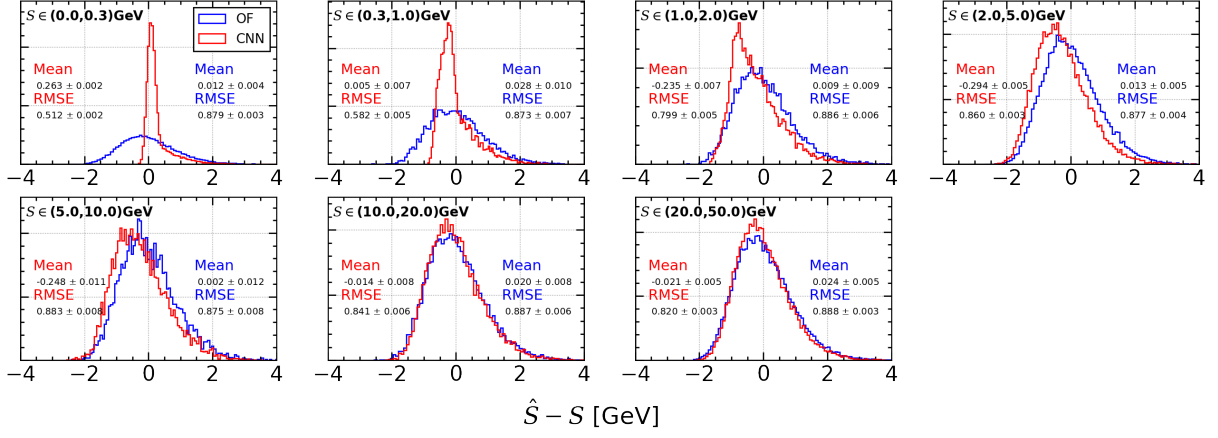


Figure 4. Histograms of $\hat{S} - S$ for models trained with a standard RMSE loss function, for each of the subsets P_j in Table 1.

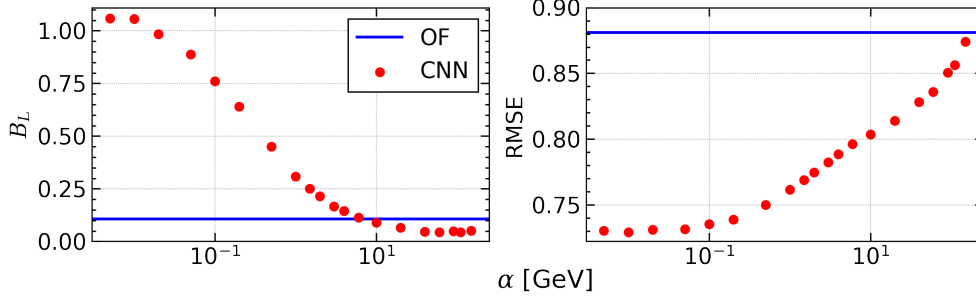


Figure 5. RMSE and B_L of S and \hat{S} for different values of α . The model hyperparameter configuration is the top configuration from Table 2. The blue lines represent the optimal filter metrics.

loss function is dominated by the RMSE and the model's RMSE and B_L are consistent with that seen in Table 2. As α increases, locally biased predictions become increasingly penalized, and the model correspondingly has a lower B_L value after training. The trade-off, however, is a slight reduction in the RMSE performance. A tunable α permits an adjustable trade-off between RMSE and B_L metrics; this could potentially be used to minimize net statistical uncertainty in the ATLAS experiment. Hyperparameters were not re-tuned for each value of α ; it thus may be the case that a different optimal hyperparameter configuration exists for each value of α .

The mean and RMSE of $\hat{S} - S$ for the top model architecture trained with the value $\alpha = 40$ are shown in Figure 6; contrasted with Figure 3, it is apparent that the new loss function successfully prevents the CNN model from making locally biased predictions. Corresponding histograms of residuals $\hat{S} - S$ in each region P_j are shown in Figure 7. Compared to the OF, over all reconstructed signals \hat{S} , the RMSE obtained with the CNN is $(6.0 \pm 0.2)\%$ lower.

The optimal hyperparameter configuration and loss function can additionally be used to train a model on a dataset where the signal S from Figure 1 is spaced sufficiently close together that overlapping ionization pulses are obtained from consecutive signal events. This is simulated by using $\Gamma \sim \mathcal{U}(3 \text{ BC}, 50 \text{ BC})$; a sample of $X(t)$ obtained using this Γ is shown in Figure 8. The performance of the CNN is significantly greater than the OF in this regime; this is shown in Figure 9. Corresponding histograms of residuals $\hat{S} - S$ in each region P_j are

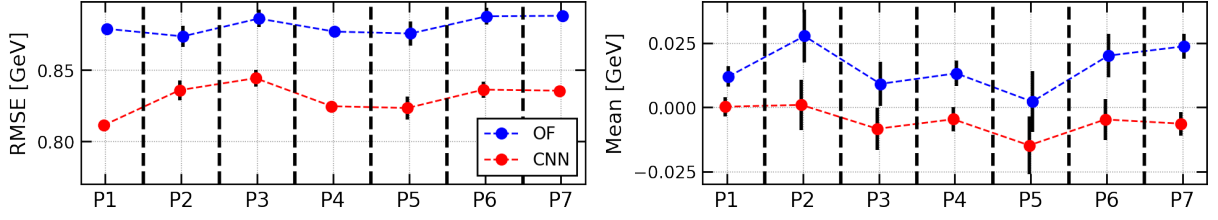


Figure 6. RMSE and Mean of residuals $\hat{S} - S$ for the optimal filter and top model architecture in Table 2 trained with $\alpha = 40$ in Equation 4.1 for each of the subsets P_j in Table 1.

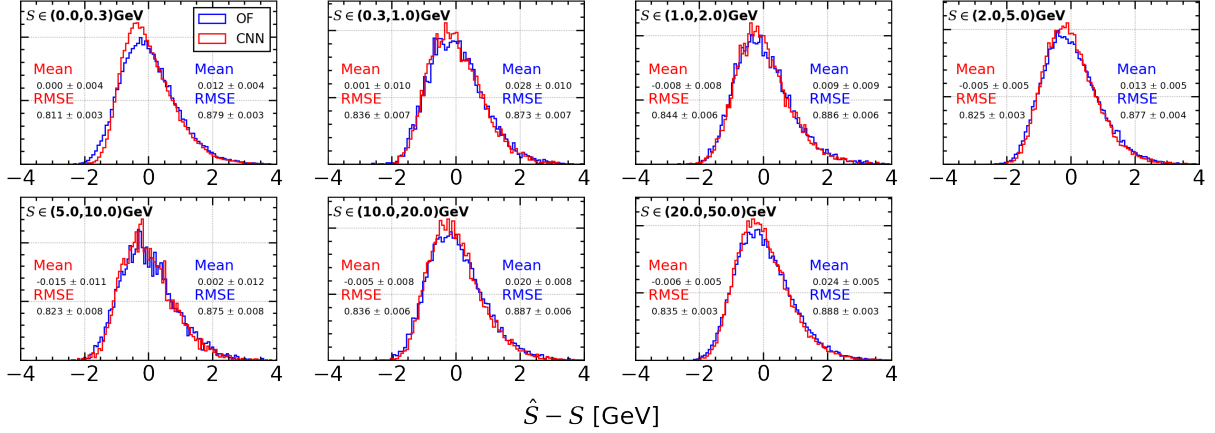


Figure 7. Histograms of $\hat{S} - S$ for models trained with the loss function given by Equation 4.1 with $\alpha = 40$, for each of the subsets P_j in Table 1.

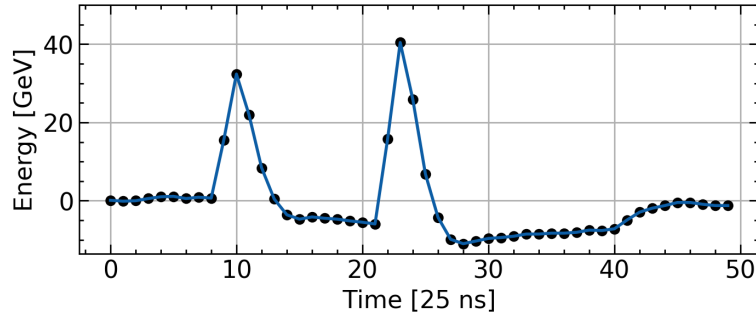


Figure 8. Two overlapping ionization pulses from signal events. Black dots represent data and blue lines represent the linear interpolation between data points.

shown in Figure 10. Compared to the OF, over all reconstructed signals \hat{S} , the RMSE obtained with the CNN is $(26.2 \pm 0.2)\%$ lower. In this regime, the OF makes locally biased predictions in all intervals but the CNN does not. In addition, the relative difference between the RMSE of the CNN and OF increases.

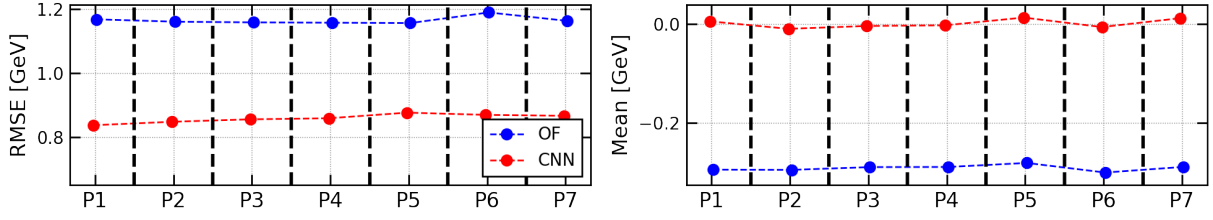


Figure 9. RMSE and Mean of $\hat{S} - S$ for the optimal filter and the top model architecture in Table 2 trained with $\alpha = 40$ in Equation 4.1 for each of the subsets P_j in Table 1. The data set consisted of signal hits close-by in time.

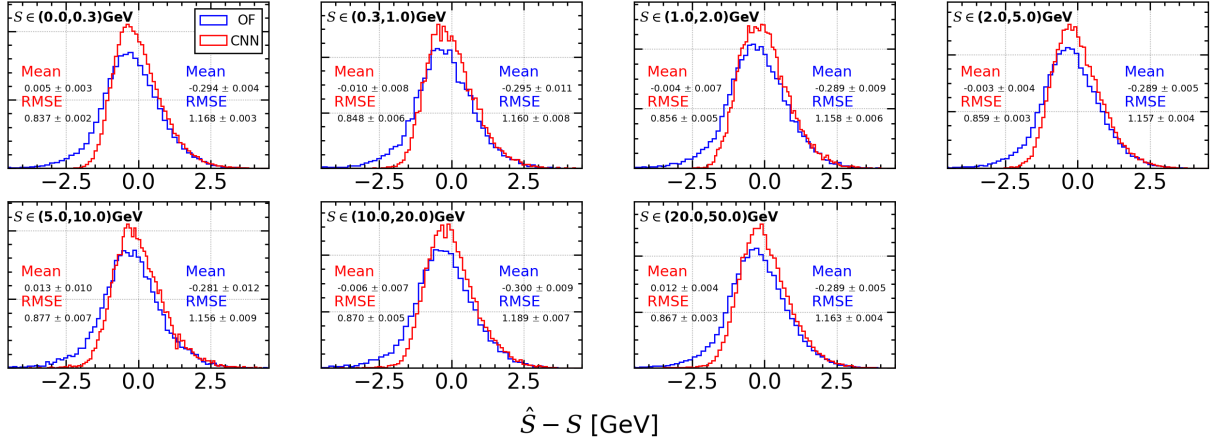


Figure 10. Histograms of $\hat{S} - S$ for models trained with the loss function given by Equation 4.1 with $\alpha = 40$, for each of the subsets P_j in Table 1. Signal is simulated here such that consecutive ionization pulses occasionally overlap.

5 Conclusion

A convolutional neural network (CNN) architecture was explored as an alternative to the optimal filter (OF) energy reconstruction algorithm in a HEC ATLAS liquid argon calorimeter cell. Signal was simulated according to expected pileup conditions in ATLAS at the HL-LHC without overlapping signal ionization pulses. A standard RMSE loss function was explored as a first option for training CNNs; while the CNN saw a reduction in RMSE of $(17.2 \pm 0.2)\%$ compared to the OF, the CNN also developed locally biased predictions in different energy ranges. A novel loss function for training was developed to address this bias. This loss function eliminated energy range biases and the CNN still outperformed the OF in RMSE by $(6.0 \pm 0.2)\%$. Signal was then simulated such that consecutive events could lead to overlapping ionization pulses; in this case, the CNN had the greatest performance improvement over the OF, with a $(26.2 \pm 0.2)\%$ decrease in RMSE and no energy range bias. With the new loss function, a CNN architecture outperforms the OF in RMSE while eliminating energy range biases.

Acknowledgments

The authors would like to thank Dr Steffen Stärz and Alessandro Ambler for fruitful discussions. Support from the Natural Sciences and Engineering Research Council of Canada, Compute Canada, and CMC Microsystems

is gratefully acknowledged.

References

- [1] The ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, 2008 *JINST* **3** 08.
- [2] The ATLAS Collaboration, *ATLAS liquid argon calorimeter: Technical design report*, CERN-LHCC-96-41.
- [3] L. Evans and P. Bryant, *LHC Machine*, 2008, *JINST*, **3** 08.
- [4] G. Apollinari et al., *High-Luminosity Large Hadron Collider*, CERN-2015-005.
- [5] W.E. Cleland and E.G. Stern, *Signal processing considerations for liquid ionization calorimeters in a high rate environment*, 1994 *Nucl. Instrum. Methods Phys.* **338** 467.
- [6] A. Kiryunin, L. Kurchaninov, and P. Stavina, *Reconstruction of Signal Parameters in the HEC Readout Driver*, (2001) ATLAS HEC-Note-110.
- [7] J. Ban et al., *Cold electronics for the liquid argon hadronic end-cap calorimeter of ATLAS*, 2006 *Nucl. Instrum. Methods Phys.* **556** 01.
- [8] L. Kurchaninov, *Modeling of the HEC Electronics Chain*, (2001) ATLAS HEC-Note-109.
- [9] Polson, L., *Application of machine learning for energy reconstruction in the ATLAS liquid argon calorimeter*, (2021) M.Sc. thesis [<http://hdl.handle.net/1828/13097>].
- [10] Y. LeCun et al., *Deep Learning*, 2015 *Nature* **521** 436.
- [11] A. Oord, *WaveNet: A Generative Model for Raw Audio*, (2016) [<https://arxiv.org/abs/1609.03499>].
- [12] Aad, G., Berthold, A.S., Calvet, T. et al. *Artificial Neural Networks on FPGAs for Real-Time Energy Reconstruction of the ATLAS LAr Calorimeters*, 2021 *Computing and Software for Big Science* **5** 19.
- [13] The ATLAS Collaboration, *ATLAS Phase-II Upgrade Scoping Document*, CERN-LHCC-2015-020.
- [14] S. Ruder, *An overview of gradient descent optimization algorithms*, (2016) [<http://arxiv.org/abs/1609.04747>].