



The Compact Muon Solenoid Experiment

Conference Report

Mailing address: CMS CERN, CH-1211 GENEVA 23, Switzerland



21 November 2018 (v2, 30 November 2018)

Exploring End-to-end Deep Learning Applications for Event Classification at CMS

Michael Andrews, Manfred Paulini, Sergei Gleyzer and Barnabas Poczos for the CMS Collaboration

Abstract

An essential part of new physics searches at the Large Hadron Collider (LHC) at CERN involves event classification, or distinguishing potential signal events from those coming from background processes. Current machine learning techniques accomplish this using traditional hand-engineered features like particle 4-momenta, motivated by our understanding of particle decay phenomenology. While such techniques have proven useful for simple decays, they are highly dependent on our ability to model all aspects of the phenomenology and detector response. Meanwhile, powerful deep learning algorithms are capable of not only training on high-level features, but of performing feature extraction. In computer vision, convolutional neural networks have become the state-of-the-art for many applications. Motivated by their success, we apply deep learning algorithms to low-level detector data from the 2012 CMS Simulated Open Data to directly learn useful features, in what we call, end-to-end event classification. We demonstrate the power of this approach in the context of a physics search and offer solutions to some of the inherent challenges, such as image construction, image sparsity, combining multiple sub-detectors, and de-correlating the classifier from the search observable, among others.

Presented at *CHEP 2018 Computing in High-Energy Physics 2018*

Exploring End-to-end Deep Learning Applications for Event Classification at CMS

Michael Andrews¹, Manfred Paulini¹, Sergei Gleyzer², and Barnabas Poczos³
on behalf of the CMS Collaboration

¹Department of Physics, Carnegie Mellon University, Pittsburgh, USA

²Department of Physics, University of Florida, Gainesville, USA

³Machine Learning Department, Carnegie Mellon University, Pittsburgh, USA

Abstract. An essential part of new physics searches at the Large Hadron Collider (LHC) at CERN involves event classification, or distinguishing potential signal events from those coming from background processes. Current machine learning techniques accomplish this using traditional hand-engineered features like particle 4-momenta, motivated by our understanding of particle decay phenomenology. While such techniques have proven useful for simple decays, they are highly dependent on our ability to model all aspects of the phenomenology and detector response. Meanwhile, powerful deep learning algorithms are capable of not only training on high-level features, but of performing feature extraction. In computer vision, convolutional neural networks have become the state-of-the-art for many applications. Motivated by their success, we apply deep learning algorithms to low-level detector data from the 2012 CMS Simulated Open Data to directly learn useful features, in what we call, end-to-end event classification. We demonstrate the power of this approach in the context of a physics search and offer solutions to some of the inherent challenges, such as image construction, image sparsity, combining multiple sub-detectors, and de-correlating the classifier from the search observable, among others.

1 Introduction

An important part of new physics searches at the Large Hadron Collider (LHC) involves the classification of collision events to distinguish between potential signal events, and events from background processes. For physics searches by the Compact Muon Solenoid (CMS) [1] collaboration, this is currently accomplished by first reconstructing the raw data collected by the detectors into progressively more physically-motivated quantities [2] until arriving at tabular-like particle-level data. The traditional analysis approach [3, 4] then uses these condensed inputs to construct an event classifier that capitalizes on the decay structure or topology of the processes involved. While such approaches have been widely successful in understanding the Standard Model of Particle Physics (SM), they potentially lose information in the process that may hinder more exhaustive searches for physics Beyond the Standard Model (BSM).

In this paper, we propose a class of event classifiers that directly use low-level detector data as inputs, or an *end-to-end* (E2E) event classifier. These are made possible by recent advances in Deep Learning and convolutional neural networks (CNNs) in particular, that

have allowed for breakthroughs in computer vision and pattern recognition with image-like data. At the same time, such E2E classifiers are also general event classifiers in that their construction is topology-independent, making them well-suited to merged and variable decay structures.

While exploring the full potential of E2E classifiers lies outside the scope of this work, we choose a simple but illustrative process to better understand what such classifiers are able to learn, and to address some of the challenges involved in their use. We therefore study the decay of the Standard Model Higgs boson to two photons using the 2012 CMS Simulated Open Data, building on earlier work presented in [5].

2 Open Data Simulated Samples

The CMS Open Data provides a number of high-quality, simulated 2012 CMS data events using Geant4 [6] to model the interaction of particles with the detector material and the most detailed geometry model of CMS.

For our signal sample, we choose the gluon fusion Higgs to diphoton dataset [7], $gg \rightarrow H \rightarrow \gamma\gamma$. For the background samples, we choose quark fusion to prompt diphoton [8], $q\bar{q} \rightarrow \gamma\gamma$, and $\gamma + \text{jet}$ production [9]. These two are representative of the most challenging background types: kinematically-differentiated decays or irreducible backgrounds ($\gamma\gamma$) and particle shower-differentiated decays due to unresolved objects ($\gamma + \text{jet}$). The samples account for the multi-parton interactions from the underlying event as well as pile-up (PU). The PU distributions are run era dependent, ranging from a peak average PU of $\langle \text{PU} \rangle = 18 - 21$ [10].

We categorize the samples based on pseudorapidity η . The central sample is restricted to $|\eta| < 1.44$ and the central+forward sample to $|\eta| < 2.3$, with the region around the electromagnetic calorimeter barrel-endcap boundary, $1.44 < |\eta| < 1.54$, excluded. For both categories, we require exactly two reconstructed photons, each with transverse momentum $p_T > 20$ GeV. The reconstructed mass of the diphoton system is required to be $m_{\gamma\gamma} > 90$ GeV. With these cuts, we obtain 63502 and 135602 events in the $\gamma\gamma$ dataset for the central and central+forward categories, respectively. For the remaining datasets, we sample the same number of events with similar PU distributions to minimize learning based on differences in PU.

3 CMS Detector & Images

The Compact Muon Solenoid (CMS) detector is arranged as a series of concentric cylindrical sections split into a barrel section and two circular endcap sections. The innermost part includes the tracking system for measuring charged particle tracks. This is then enclosed by the electromagnetic calorimeter (ECAL) which measures photon and electron energy depositions, followed by the hadronic calorimeter (HCAL) which measures the energy depositions of hadronic particles.

The CMS Open Data makes available the reconstructed hits for the ECAL and HCAL. This makes it possible to construct calorimeter images whose pixels correspond exactly to physical crystals (towers). The images are constructed using the ECAL granularity, with the HCAL hits upsampled to match. However, the key challenge in constructing lossless multi-channel detector images is differences in detector segmentation not granularity. For instance, the ECAL endcaps (EE) are segmented in (iX, iY) while the HCAL endcaps in $(i\eta, i\phi)$. We thus devise two image geometry strategies: one where the ECAL endcap segmentation is preserved and the HCAL endcap hits are projected onto an (iX, iY) grid (ECAL-centric), and another where the HCAL endcap segmentation is preserved and the ECAL endcap hits are projected onto an $(i\eta, i\phi)$ grid (HCAL-centric). For the moment, the full tracker information

is represented only using reconstructed tracks. Specifically, each track is approximated as a point in an image layer of ECAL-like resolution, with intensity equal to the track’s p_T . The position of this point corresponds to the track’s $(i\eta, i\phi)$ coordinates evaluated at the track’s point of closest approach to the beamline.

For the central category, we use only the subdetector images which overlap with the ECAL barrel (EB) (Figure 1a) giving image inputs of resolution 170×360 . For the central+forward category, we use images which overlap with both the EB+EE (ECAL-centric: Figures 1a+1b or HCAL-centric: Figures 1a+1c). These give image inputs of resolution 170×360 and 100×100 for the ECAL-centric geometry, and 280×360 for the HCAL-centric one. Each image, in turn, contains three channels or layers corresponding to the track p_T , ECAL energy, and HCAL energy.

4 Network & Training

For all image-based or end-to-end (E2E) event classifiers, Residual Net-type (ResNet-15) convolutional neural networks are used due to their simplicity and scalability with image size and network depth [11]. The various E2E classifier models are summarized in Table 1.

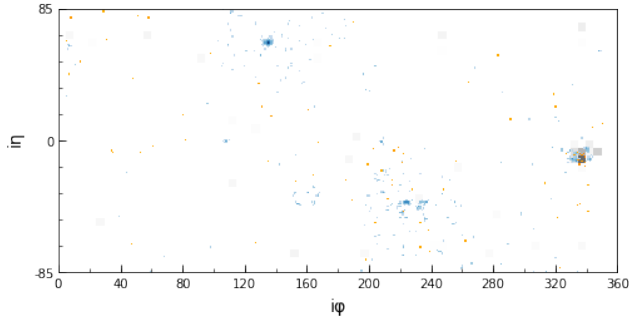
Model	Category	Architecture	Inputs
EB	Central	ResNet-15	ECAL*
CMS-B	Central	ResNet-15	Tracker, ECAL, HCAL*
ECAL	Central+Fwd	3 x ResNet-15, FCN	ECAL
CMS-I	Central+Fwd	3 x ResNet-15, FCN	Tracker, ECAL, HCAL
CMS-II	Central+Fwd	ResNet-15	Tracker, ECAL, HCAL

Table 1: Summary of end-to-end models used in this paper. *NOTE: Models from the central category only use the barrel portion of the subdetector images (c.f. Figure 1a).

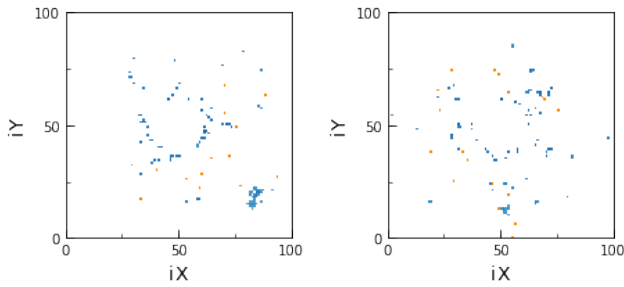
To serve as a reference for conventional event classifiers, we train a separate dense, fully-connected neural network (FCN) on the reconstructed 4-momenta of the two candidate photons in each event, which we denote as the 4-momentum classifier. The photon p_T s are divided by the reconstructed diphoton mass $m_{\gamma\gamma}$ to de-correlate the classifier from the mass of the Higgs boson [3].

To achieve mass de-correlation in the end-to-end case, we divide each image by the reconstructed diphoton mass for that event (see Figure 3). However, this only delays the onset of mass-sculpting and does not completely eliminate it—we suspect the shower profile provides an alternate avenue for learning the p_T of the shower. An updated version of this work includes a more robust solution to this problem using a CVM-based loss penalty [12]. For the present work, it suffices to implement early stopping to intercept the training before the mass is learned.

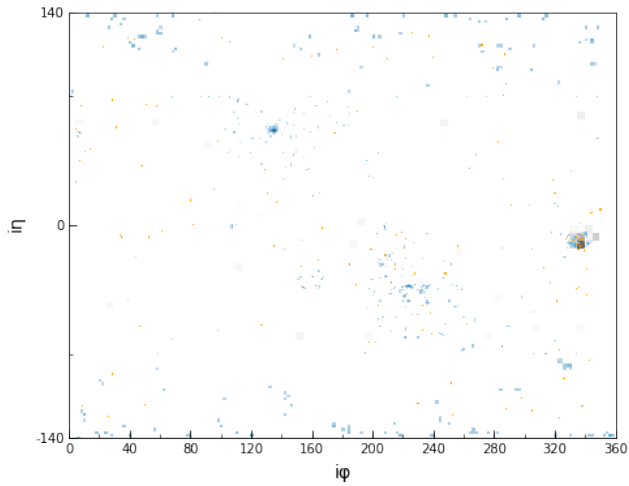
The breakdown of training and test set—which doubles as the validation set—is shown in Table 2. These limited statistics provide a slight advantage to the 4-momentum classifier which has less weights to train. Both training and validation sets contain the same number of events for each of the three event classes. All training was done using the PyTorch [13] software library running on a single NVIDIA Titan X GPU for which total training time ranges from several hours to a day. Note that individual classifier optimization was kept to a minimum to provide a broad and consistent survey of E2E classifiers.



(a) Barrel section of composite image in ECAL-centric geometry. Image resolution: 170×360 .



(b) Endcap sections of composite image in ECAL-centric geometry. Image resolution: 100×100 .



(c) Composite image in HCAL-centric geometry. Extent of EB indicated by minor ticks on y -axis. Image resolution: 280×360 .

Figure 1: Composite images of a single γ + jet event in different geometry strategies: separate barrel (1a) and endcaps (1b) for the ECAL-centric geometry, and stitched together (1c) for the HCAL-centric. Each image contains three channels: track p_T (orange), ECAL energy (blue), and HCAL energy (gray). Note the photon at $(i\eta = 70, i\phi = 130)$ vs. jet at $(i\eta = -10, i\phi = 340)$.

Category	Training Events per class	Test Events per class
Central	51200	11800
Central+forward	120000	15600

Table 2: Number of events in training and test sets *for each class*. Test set doubles as validation set due to limited statistics. The total training and test sets contain a balanced number of class samples.

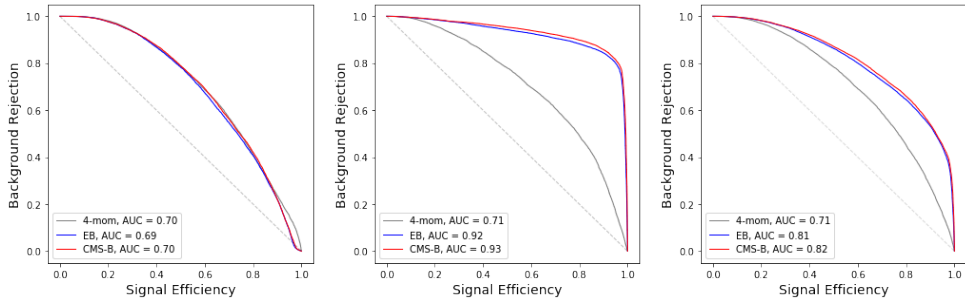


Figure 2: Multi-class Event Classification ROC curves, central $|\eta| < 1.44$ region. Left: $H \rightarrow \gamma\gamma$ vs. $\gamma\gamma$ component. Center: $H \rightarrow \gamma\gamma$ vs. $\gamma + \text{jet}$ component. Right: $H \rightarrow \gamma\gamma$ vs. Rest

5 Event Classification

In any real physics decay, energy and momentum conservation impose physical constraints on the allowed kinematics of the decaying particles. In this section, we therefore attempt to classify realistic $H \rightarrow \gamma\gamma$ vs. $\gamma\gamma$ vs. $\gamma + \text{jet}$ decays. The end-to-end (E2E) event classification results are divided by pseudorapidity (see Section 2), with the results for the central (central+forward) category shown in Figure 2 (Figure 4). The ECAL-only classifier is labeled **EB (ECAL)** and the Tracks+ECAL+HCAL classifier in the ECAL-centric geometry is labeled **CMS-B (CMS-I)**. For the central+forward region, we also include the results of the HCAL-centric classifier (**CMS-II**). In each category, we plot the signal vs. combined background ROC (1-vs-Rest), as well as the signal vs. single background ROC component (1-vs-1). For context, we also include the results of the (mass de-correlated) 4-momentum-only classifier (**4-mom**).

5.1 Central η region

We first focus on the central category where we only use detector images from the barrel section of CMS. From the 1-vs-Rest plot (Figure 2, rightmost), we see that, overall, image-based classifiers perform better than purely kinematical classifiers. This is, of course, expected in the presence of a shower-differentiated background but serves to confirm that the E2E classifier is delivering as expected. We also see that the **EB** and **CMS-B** classifiers perform comparably, with only negligible advantage to including additional subdetectors, which is expected from the signature of the decays. Note that the other subdetector images contain quite a bit of pile-up and underlying event (see Figure 1). That no degradation in performance was seen by including these additional subdetectors highlights the E2E classifier’s ability to effectively screen out features which are not relevant to the hard-scattering process.

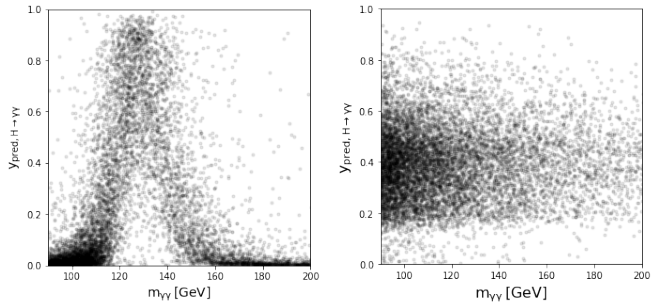


Figure 3: Central **EB** classifier score in signal label vs. reconstructed diphoton mass for true $\gamma\gamma$ events, without (left) and with (right) mass de-correlation. The impact of de-correlation is more severe if the background lacks shower differentiation.

Looking at the $H \rightarrow \gamma\gamma$ vs. $\gamma\gamma$ component (Figure 2, leftmost), we see that for kinematically-differentiated backgrounds, the E2E classifiers perform comparably to the 4-momentum-only classifier. This demonstrates that, at least in this context, we have paid no penalty in using a general classifier trained on low-level data over a specialized kinematical classifier that relied on our ability to reconstruct the event. Note that while mass de-correlation was applied, classifier performance vs. the irreducible background was not completely eliminated. This suggests the kinematical information is manifested in the detector image in two ways: the angular distribution of the photon showers and the energy scale of the shower hits. While mass de-correlation removes the latter, it preserves the former, allowing for residual performance.

Turning now to the $H \rightarrow \gamma\gamma$ vs. $\gamma + \text{jet}$ component (Figure 2, center), we see that this is primarily responsible for the E2E advantage over kinematics-only. This is expected because the jet manifests itself in the ECAL image as a differentiated shower, which, on occasion, is discernible by eye (see Figure 1). As studied in [5], E2E classifiers are highly sensitive to differences in shower shapes even when no distinguishing kinematical information is present. Moreover, the $\gamma + \text{jet}$ decay exhibits similar non-resonant kinematics to $\gamma\gamma$ and so, to the 4-momentum classifier, the two should look alike. This is confirmed by their similar 4-momentum results (c.f. Figure 2, leftmost and center). Lastly, owing to strong shower differentiation, the $\gamma + \text{jet}$ background shows strong performance relative to $\gamma\gamma$ despite being mass de-correlated. This suggests that the impact of mass de-correlation depends strongly on the importance of kinematics over shower differentiation.

5.2 Central+Forward η region

In this category, we have included the endcap images either in ECAL-centric (**ECAL**, **CMS-I**) or HCAL-centric (**CMS-II**) fashion (see Section 3). In general, we find the main conclusions from the Central category to still be relevant with minimal differences in absolute performance. This alone informs us about the scalability of E2E network architectures and their ability to deal with the increased pile-up of the forward detector regions. Despite drastic differences in network structures, we find classifier performance to not be greatly sensitive to the choice of endcap projection.

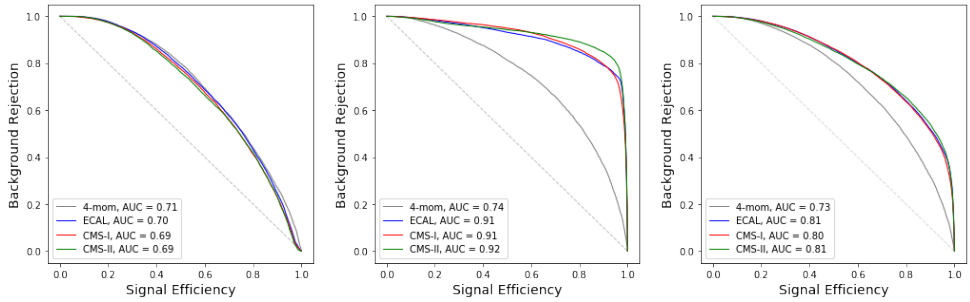


Figure 4: Multi-class Event Classification ROC curves, central+forward $|\eta| < 2.3$ region. Left: $H \rightarrow \gamma\gamma$ vs. $\gamma\gamma$ component. Center: $H \rightarrow \gamma\gamma$ vs. $\gamma + \text{jet}$ component. Right: $H \rightarrow \gamma\gamma$ vs. Rest

6 Conclusions

In this paper, we described the construction of a class of general, end-to-end, image-based event classifiers, using high-fidelity, simulated, low-level detector data as inputs. While these classifiers are best suited to challenging decays, we have applied them in a simplified search for the Standard Model $H \rightarrow \gamma\gamma$ decay to highlight their key features and challenges. Through the irreducible $\gamma\gamma$ background, we were able to infer that such classifiers are able to learn about the angular distribution of the photon showers as well as the energy scale of their constituent hits. By removing the latter through preprocessing, we showed that we were able to de-correlate the event classifier from the reconstructed diphoton mass while still preserving the former. Through the reducible $\gamma + \text{jet}$ background, we additionally showed that such classifiers can learn about the photon shower shape giving them a strong advantage over purely kinematical classifiers while being less reliant on the energy scale information. Finally, we demonstrated the scalability and flexibility of these classifiers when dealing with multiple detector images and networks where we found them to be robust versus the choice of geometry projection and the presence of underlying event and pile-up.

Acknowledgments

We would like to thank the CERN Open Data group for the release of simulated data under an open access policy. We strongly support initiatives to provide such high-quality simulated datasets that can encourage the development of novel but also realistic algorithms, especially in the area of machine learning. We believe their continued availability will be of great benefit to the high energy physics community in the long run.

References

- [1] The CMS Collaboration, JINST **3**, S08004 (2008)
- [2] The CMS Collaboration, JINST **12**, P10003 (2017)
- [3] The CMS Collaboration, Eur. Phys. J. C **74**, 3076 (2014)
- [4] P. Baldi, P. Sadowski and D. Whiteson, Nature Communications **5**, 4308 (2014)
- [5] M. Andrews, M. Paulini, S. Gleyzer and B. Poczos, 18th International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT2017) **1085**, 042022 (2018)

- [6] GEANT4 Collaboration, Nucl. Instrum. Meth. **A506**, 250-303 (2003)
- [7] The CMS Collaboration, *Simulated dataset GluGluHToGG_M-125_8TeV_pythia6 in AODSIM format for 2012 collision data*. CERN Open Data Portal. doi:10.7483/OPENDATA.CMS.WQ7P.BZP3 (2017)
- [8] The CMS Collaboration, *Simulated dataset DiPhotonBorn_Pt-25To250_8TeV_ext-pythia6 in AODSIM format for 2012 collision data*. CERN Open Data Portal. doi:10.7483/OPENDATA.CMS.WV7J.8GN0 (2017)
- [9] The CMS Collaboration, *Simulated dataset GJet_Pt40_doubleEMEnriched_TuneZ2star_8TeV_ext-pythia6 in AODSIM format for 2012 collision data*. CERN Open Data Portal. doi:10.7483/OPENDATA.CMS.2W51.W8AT (2017)
- [10] CERN, *PdMVPileUpDescription2012, PU_RD1*. https://twiki.cern.ch/twiki/bin/viewauth/CMS/PdmVPileUpDescription#PU_RD1 (2017)
- [11] K. He, X. Zhang, S. Ren, J. Sun, 2016 IEEE Conference on Computer Vision and Pattern Recognition, 770 (2016)
- [12] A. Rogozhnikova, A. Bukvac, V. Gligorovd, A. Ustyuzhaninb, M. Williams. JINST **10**, T03002 (2015)
- [13] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, A. Lerer. NIPS-W (2017)