Merged Di-photon Identification for the ATLAS Experiment at the Large Hadron Collider

Sanha Cheong, Jannicke Pearkes, Aviv Cukierman†
{ sanha, jpearkes, acukierm } @ stanford.edu

Department of Physics, Stanford University
SLAC National Accelerator Laboratory

Search for New Particles

The **Large Hadron Collider (LHC)** collides protons at extremely high energies to study the fundamental building blocks of the universe. **The ATLAS detector**, shaped like a cylinder with two proton beams coming in along its axis, detects the particles produced from proton collisions at its center and measures their energies and trajectories.

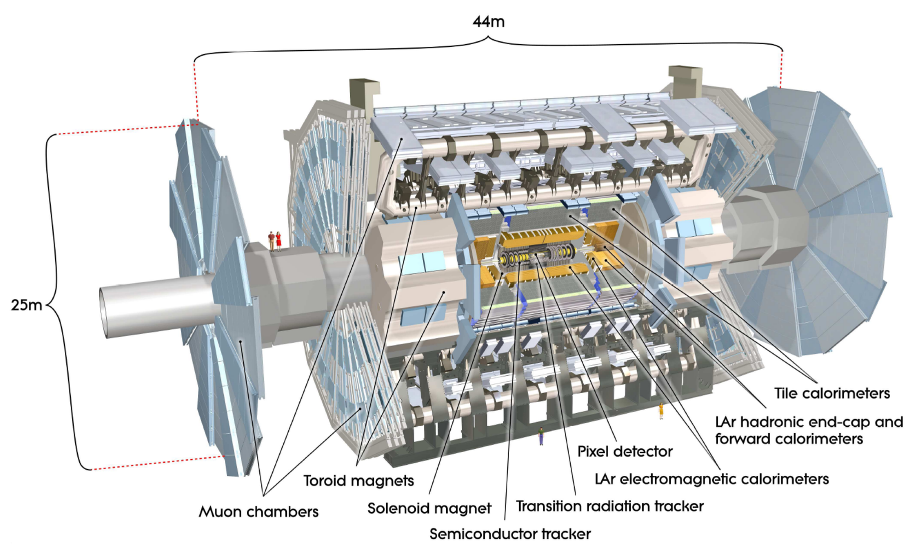


Figure 1: An overview of the ATLAS detector [4]

Some of the most exciting physics the ATLAS Collaboration is looking for is **new particles**. One particular candidate of interest is a new (pseudo-)scalar particle (a). It can be produced by a Higgs boson (h) decay as seen in Fig. 2. This new particle a can decay into a pair of photons ($\gamma\gamma$) or a pair of gluon jets (gg). A recent analysis [1] has set limits on the existence of this new particle by exploring the $\gamma\gamma gg$ final state. This analysis is sensitive to our ability to reconstruct and resolve final-state photons.

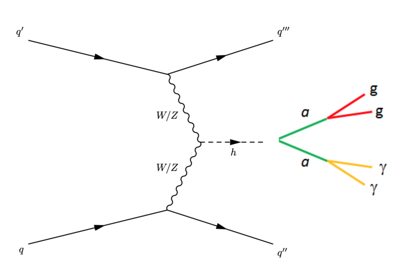


Figure 2: Hypothetical Higgs decay $h \to aa \to \gamma\gamma gg$

If the mass of a is very small, the two final-state photons tend to be highly collimated. Because collimated photons are difficult to identify correctly, current searches are limited to a relatively high mass region. In order to extend the search space to lower mass regions, we utilize neural networks (NN's) to identify signatures in the ATLAS detector created by **a pair of collimated photons**.

Objectives

- Use **deep neural networks** to identify collimated di-photon events at the ATLAS detector.
- A process creating a new particle like an a is extremely rare, and there are many backgrounds. It is important to achieve extremely **high background rejection** (1/False Positive Rate).
- Out-perform the **baseline** existing ATLAS photon reconstruction algorithms (not ML-based)
- For physics analyses, it is important for us to retrieve signal efficiency across different energies. Hence, we want our **performance to be independent of energy**.

Topo-Clusters & Jets

We extract our signal and background data from Monte Carlo simulations produced by the ATLAS Collaboration. These are detailed simulations of how particles are produced from proton collisions and how they interact with the detector.

Interacting with the calorimeters, particles like quarks (q), gluons (g), and photons (γ) create showers of particles, leaving wide-spread energy deposits called **jets** (Fig. 3); while most quark and gluon jets are contained within the hadronic calorimeter (HCAL), photon jets are mostly contained within the electromagnetic calorimeter (ECAL). A small group of hits in the calorimeter cells are clustered into **topo-clusters**. Hence, each jet is a collection of constituent topo-clusters. Jets can consist of different numbers of topo-clusters.

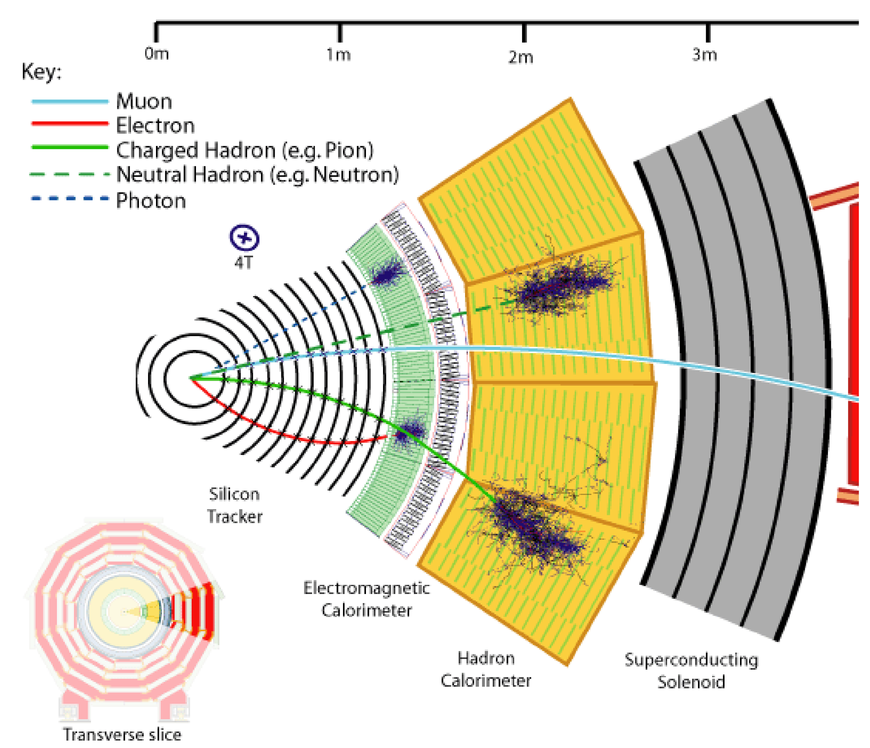


Figure 3: Different signatures in a particle detector [2]

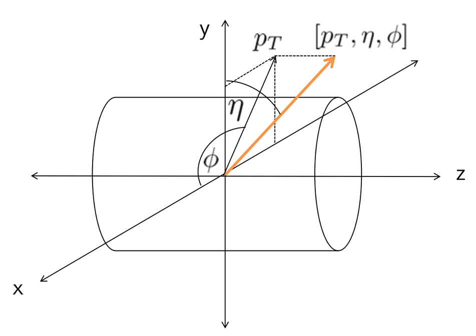


Figure 4: Coordinate system

Each topo-cluster has **4 features**: $(p_T, \eta, \phi, p_{EM})$:

- Transverse momentum p_T: energy intensity
- Pseudo-rapidity η: direction of particle
- Azimuthal angle ϕ: direction of particle
- EM Fraction p_{EM}: ratio of energy deposited in ECAL (green region in Fig. 3)

Each example in our dataset is a jet; signal jets are di-photon jets, while background jets are unbiased di-jet process jets (mostly q/g jets). The signal jets have two high p_T, high p_{EM} topo-clusters close to each other, characteristic of the collimated photons. Figure 5 show an example signal jet image [3] in the p_T and p_{EM} channels.

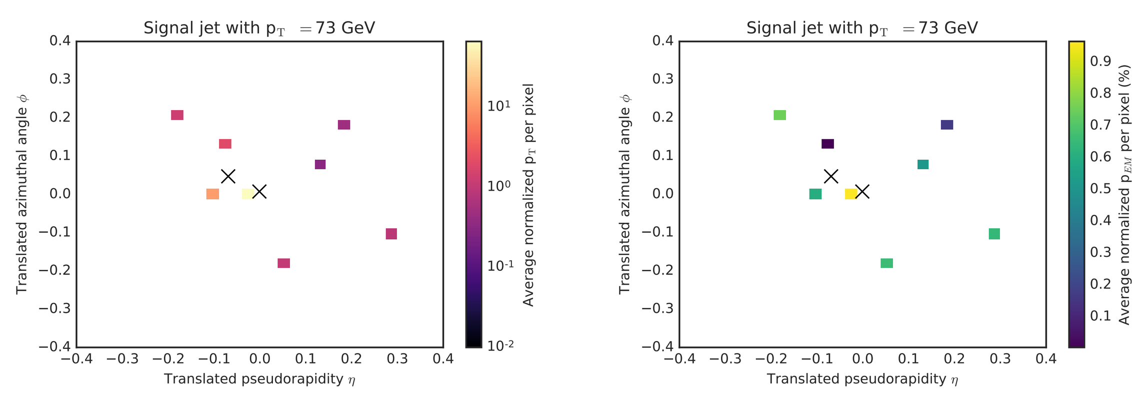


Figure 5: An example signal jet image in p_T and p_{EM} channels

Methods

We have experimented with **3 different neural network architectures**: fully connected network (FCN), long short-term memory recurrent neural network (LSTM), and 2D convolutional neural network (CNN). Each network was trained with the Adam optimizer with early-stopping based on the history of validation loss. The network architectures were tuned with cross-validation. The input to each architecture was:

- **FCN**: a flattened feature vector of all topo-clusters, zero-padded where necessary
- **LSTM**: a sequence of topo-clusters, each with up to 4 features
- **CNN**: pixelated image of each jet with 2 spatial coordinates (η, ϕ) and 2 channels (p_T, p_{EM})

In order to achieve performance independent of jet p_T, we **flattened the p_T-distributions** of signal and background jets. This was done by smoothly approximating the jet p_T distributions using kernel density estimation and assigning each sample a weight of $w_i = 1/p_{y_i}(p_{T,i})$.

We used 128,000 jets for training, 25,600 for validation, and 61,982 for test. The signal ratio is about 20% in each dataset.

Understanding What Matters

We performed a number of experiments to probe what features were most useful for classification.

Number of Constituent Topo-clusters

We varied the number of input topo-clusters for each jet from 5 to 50. This did not result in large changes to the validation set accuracy achieved by each architecture. This indicates that the 5 leading topo-clusters contain most of the crucial information for classification.

Ablation Studies

We also trained the NN's with only partial information to see which variables are the most important for classification. The results are reported in Table 1.

Input Used	FCN	LSTM	CNN
p_T	90.1	89.0	88.8
p_{EM}	92.8	93.4	90.7
η	83.2	N/A	N/A
ϕ	85.8	N/A	N/A
All	96.6	96.6	95.9

Table 1: Results from ablation studies. Reported values are highest validation accuracies achieved.

This result shows that **p_{EM} is the most important feature**, with p_T being the second. It is also rather surprising that we can achieve good accuracy just with η or ϕ. One interpretation of this result is that the NN's learns to distinguish signal and background based on the number of constituents. The results of the ablation studies can be interpreted by studying the average jet images in Figs. 6. We can see that the p_T spectrum of the background jets is much more diffuse than the signal jets. In a similar fashion, Fig. 6 shows that the signal jets typically have a center with a high p_{EM} (due to the photons) that rapidly drops off, whereas the background jets typically have a smaller overall p_{EM} as and a more diffuse spectrum.

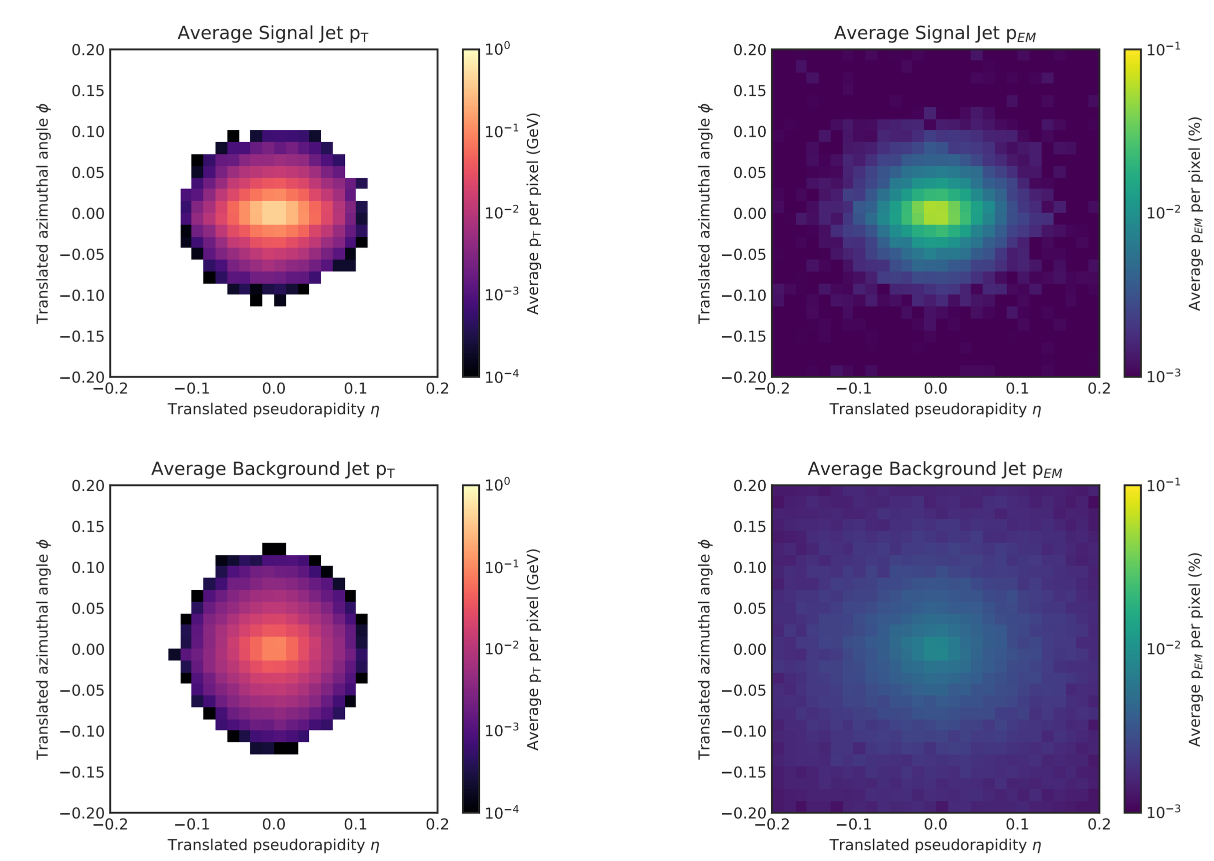


Figure 6: Average jet images

References

[1] Aaboud, M. et al. (2018). Search for Higgs boson decays into pairs of light (pseudo)scalar particles in the $\gamma\gamma jj$ final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Submitted to: Phys. Lett.*

[2] Barney, D. (2016). CMS Detector Slice. CMS Collection.

[3] de Oliveira, L., Kagan, M., Mackey, L., Nachman, B., and Schwartzman, A. (2016). Jet-images — deep learning edition. *JHEP*, 07:069.

[4] Pequenao, J. (2008). Computer generated image of the whole ATLAS detector.

Final Results & Analysis

The best-performing networks of FCN, LSTM, and CNN were able to achieve **96.40%, 96.31%, and 95.40% test accuracies** respectively. All three architectures greatly outperform the baseline. The ROC curves and AUC's, along with the baseline, are shown in Fig. 7.

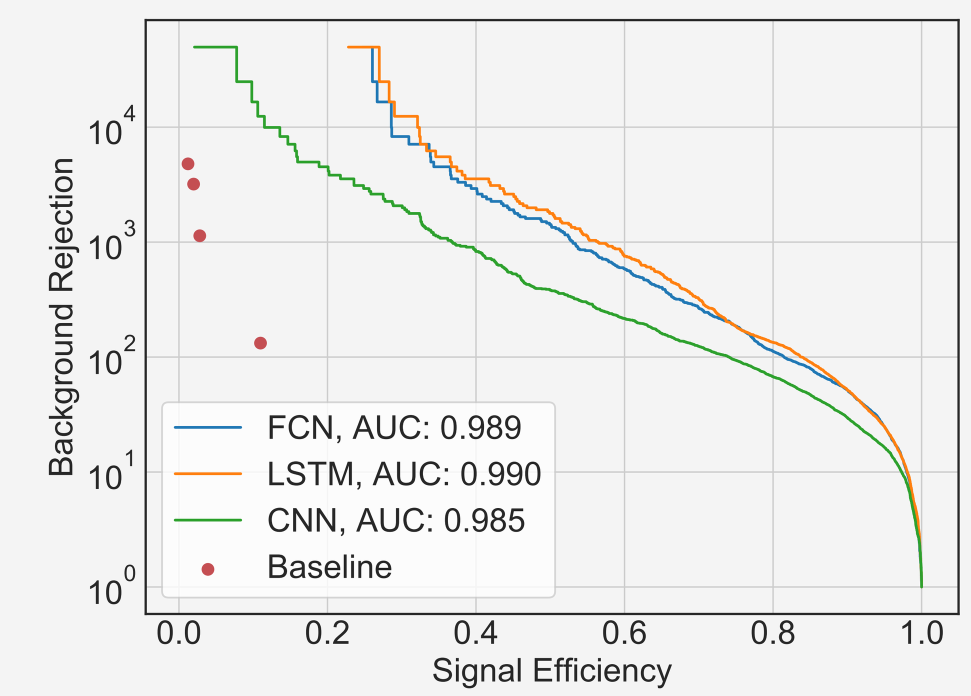


Figure 7: ROC curves of best-performing NN's

Fig. 8 shows the weighted test accuracy for different bins of jet p_T. The test accuracies are consistently over 95% for all architectures for all jets with $p_T > 40\,\mathrm{GeV}$.

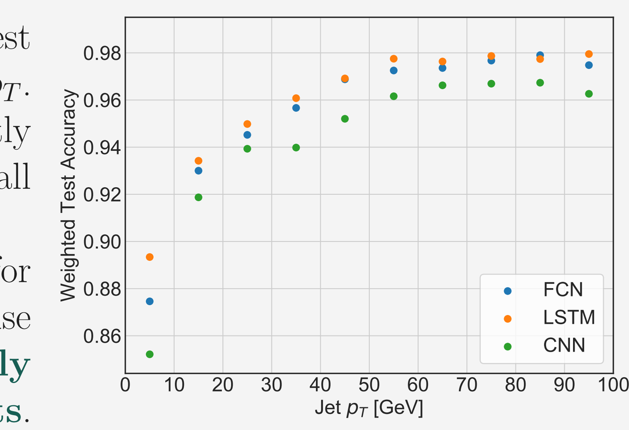
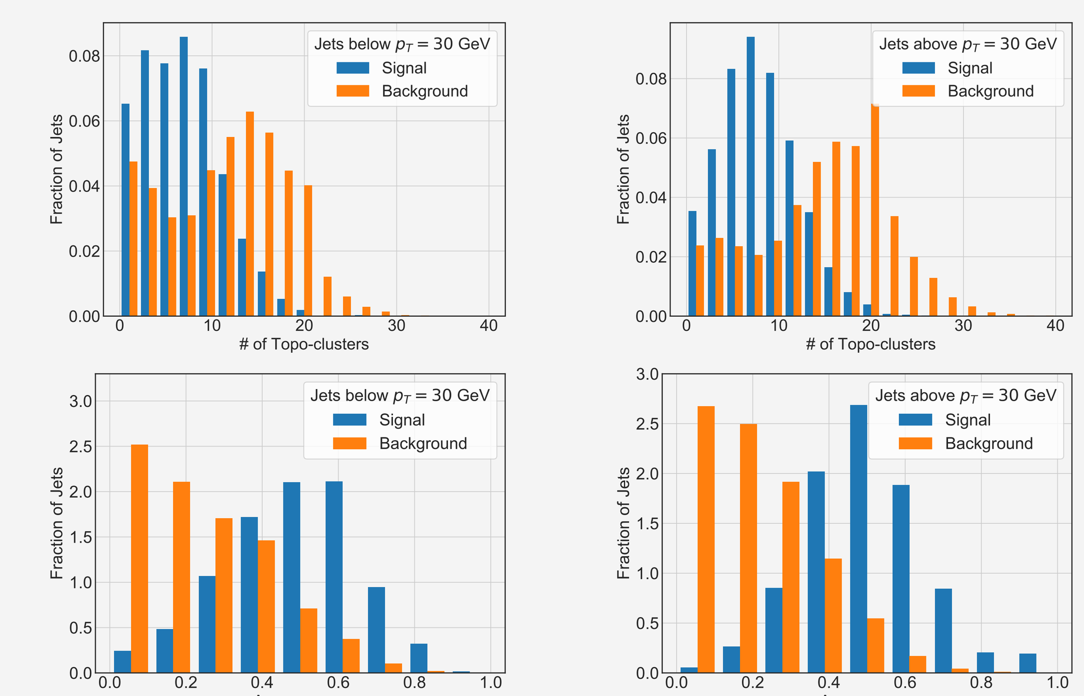


Figure 8: Test accuracies for different p_T regions

The test accuracy degrades for lower energy jets, potentially because the **classification is genuinely more difficult for low-p_T jets.** As support for this claim, we see that the feature distributions of signal and background jets overlap much more, for jets with $p_T < 30\,\mathrm{GeV}$ as shown below.



Acknowledgements

We thank Aviv Cukierman† for his help in obtaining the original data samples and useful discussions throughout this study. We also thank Prof. Ariel Schwartzman from the SLAC ATLAS group for helpful discussions and feedbacks.