Top Quark Tagging using Convolutional Neural Networks

Aya Beshr
University of Ain Shams, Cairo, Egypt
Department of Physics

CERN Work Project Report
CERN, ATLAS, Jet Substructure Group
Supervisor: Mario Campanelli

[Dated: August 13, 2019]

Abstract

The scope of this project concentrates mainly on tagging the hadronic top quark signal versus W and QCD jets taken both as background. Jets are represented as 2d projections either in $\eta-\varphi$ palne or lund plane. The performance obtained was compared with the two approaches using one of the Machine Learning algorithms called Convolutional Neural Network (CNN).

I. Introduction

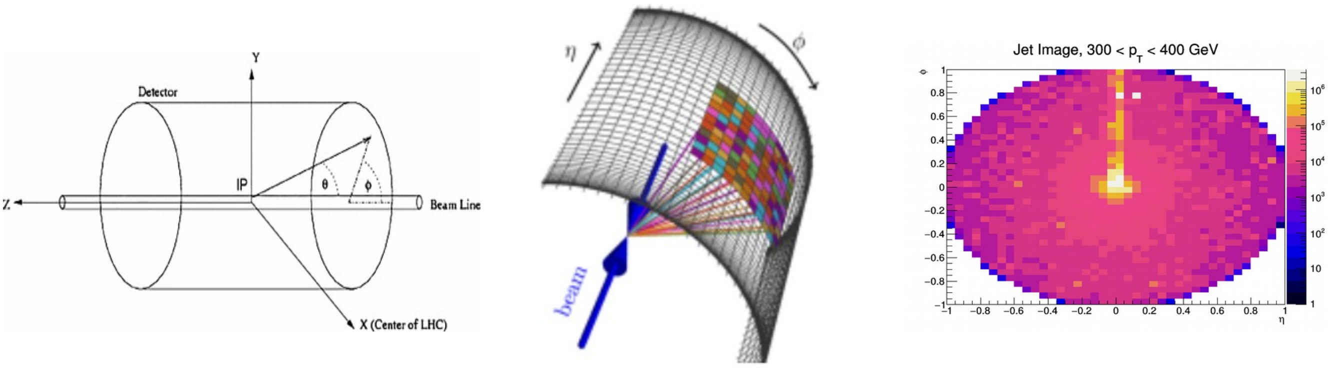
The Large Hadron Collider (LHC) is the world's largest and most powerful particle accelerator. It first started up on 10 September 2008 at CERN - Geneva. The LHC consists of a 27-kilometre ring of superconducting magnets with a number of accelerating structures to boost the energy of the particles along the way. Inside the accelerator, two high-energy particle beams travel opposite to each other at close to the speed of light before they are made to collide. Many particles are produced out of this collision. A type of them is a spray or cone of particles called jets.

Jet is a narrow cone of hadrons and other particles produced by the hadronization of a quark or gluon in a particle physics or a heavy ion experiment such as that produced at the LHC and all travel in the same direction. As quarks carry color charge they cannot be exist in free state due to color confinement that allows only for colorless bound states.

Classification of different types of jets depends on the particles initiating them. Some are known as Quantum Chromodynamics (QCD) or standard model jets which are originating from quark or gluon and others as signal jets originated from certain type of quark or vector boson as top, bottom, W or $Z^{'}$.

This project mainly concentrates on tagging the hadronic (top) quark signal versus (W) and (QCD) jets taken both as background. Jets are represented as 2d projections either in (η−φ) palne or lund plane. The performance obtained was compared with the two approaches.

Jet images are also known as (η − φ) images are two dimensional images formed in the (η − φ) plane of ATLAS detector, where (η) is called pseudorapidity ($\eta = \ln \tan(\theta/2)$), (θ) is the angle between beam line and any point in (η − φ) plane and (φ) is the angle about the beam line . The images are translated to have the jet axis aligned with the origin and rotated to have the highest pt topo-cluster aligned with the y-axis as illustrated in figure(1c).



(a) coordinate system of ATLAS detector

(b) η − φ representation

(c) jet image formed by translation then rotation at certain pt-range for top signal

Figure (1)

Lund jet planes are abstract representations for the branching of jet constituents. Each branching of jet constituent is represented by a point in 2d plane and constructed with variables ΔR, z . In this work lund planes are constructed using logarithmic scale for two variables mainly ln(1/z) on the y-axis and ln(R/ΔR) on the x-axis, where (Z) is the momentum fraction, (ΔR) is the separation between the emitted particle and the jet core and R is the jet radius. Experimentally, the points of the lund planes are reconstructed tracing back the jet reconstruction. The lund planes consists of several regions as shown in figure(2).

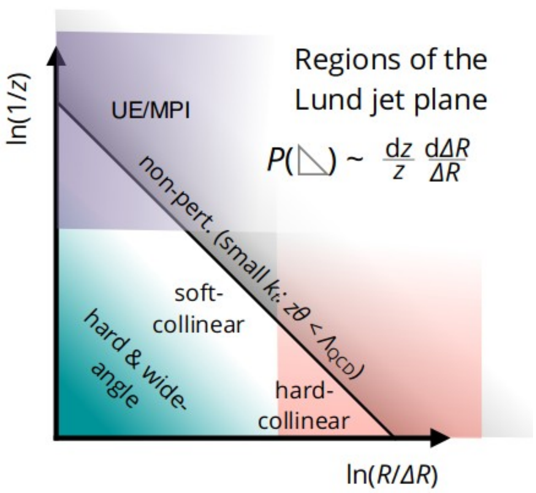


Figure (2): *different regions of lund plane*

In order to create lund planes anti-k_t algorithm was used to reconstruct the topo-clusters and Cambridge/Achen or C/A algorithm was firstly used for reclustering the tracks within jets and then used another time for tracing back the clustering and obtain a point in the lund plane for each step.

A type of Machine Learning algorithm was used to compare results for both the jet images and lund planes called Convolution Neural Network or CNN. This type is dealing with the input data in form of images . A digital image is a binary representation of visual data. It contains a series of pixels arranged in a grid-like fashion that contains pixel values to denote how bright and what color each pixel should be. It consists of several layers. The layers are arranged in such a way so that they detect simpler patterns first (lines, curves, etc.) and more complex patterns (faces, objects, etc.) further along.

According to the type and the number of the layers used a certain model is being built for training and evaluating the neural network. In this work the model is a pre-made model called LeNet5 and the architecture of the model can be described as follows:

1- (2) Convolution layers (3x3): each layer have an input which is a tensor with shape (number of images) x (image width) x (image height) x (image depth) and convolutional kernels whose width and height are hyper-parameters, and whose depth must be equal to that of the image. Convolutional layers convolve the input and pass its result to the next layer.

2- (2) PRELU layers: refers to Parametric Rectified Linear Unit which is a type of CNN layers which uses type of activation function.

3- (1) Max Pooling layer: Convolutional networks may include pooling layers which reduce the dimensions of the data by combining the outputs of neuron clusters at one layer into a single neuron in the next layer. *Max pooling* uses the maximum value from each of a cluster of neurons at the prior layer.

4- (2) Dropout layers: Dropout is a technique used to improve over-fit on neural networks. Some deep learning models use Dropout on the fully connected layers, but is also possible to use dropout after the max-pooling layers as it is the case here.

5- (1) Flatten layer: is a layer that flattens the output of convolutional neural network layers, in preparation for the fully connected layers that make a classification decision. Flattening is a key step in all Convolutional Neural Networks (CNN)

6- (2) Dense layers: also known as fully connected layer, in which the results of the convolutional layers are fed through one or more neural layers to generate a prediction.

7- (2) Activation layers: are an **activation** function that decides the final value of a neuron.

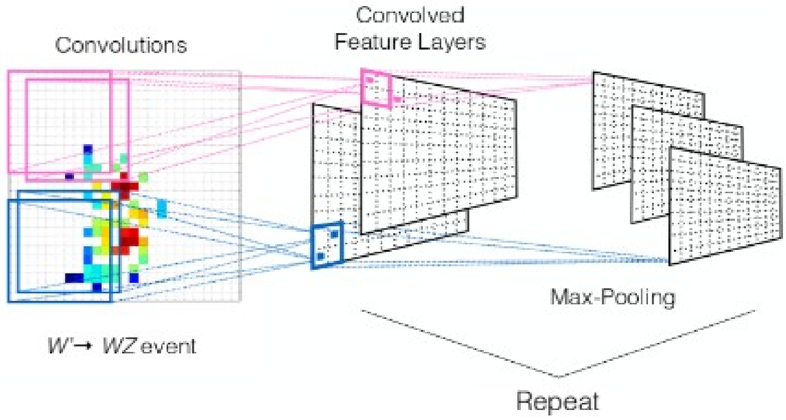
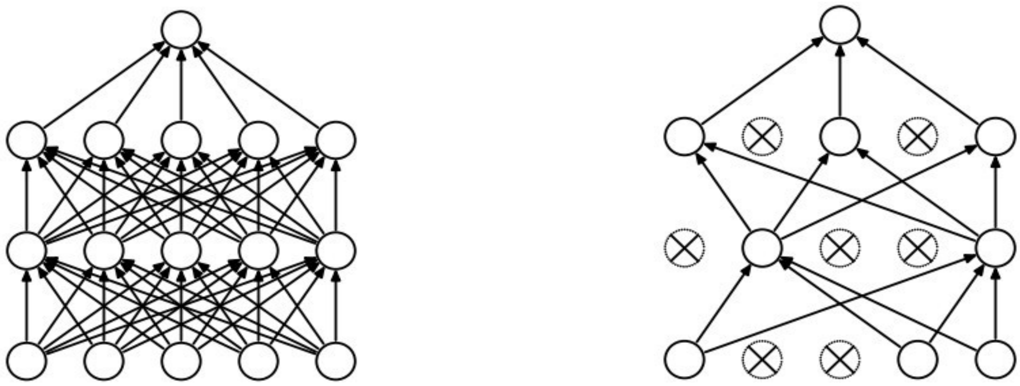


Figure (3): *convolutional and max pooling layers applied on W jet image as e.g*

A representation of dropout layers and dense layers can be shown in figure(4) which represents multilayer perceptron network with three layers of five units and an output unit. The network is shown with and without dropout.



(a) dense layer *(b) dropout layer*

Figure (4)

II. Results

The model used for training the neural network is LeNet5 for number of epochs = 9 and number of samples = 2000 where the top jet samples = 1000 and the background samples (W+QCD) = 1000. Performance of the neural network was measured using the receiver operating characteristic (ROC) curve as shown in figure(7). This plots the true positive rate against false positive rate. Table(1) sho

the lund plane and jet images. It is easily obvious that the performance increases with increasing the p_t, also it is much better for the lund plane than jet images.

Pt-ranges (GeV)	**300-400**	**500-600**	**700-850**	**900-1500**	**1900-2400**	**2400-2800**	**2800-3000**
Lund Planes	0.780	0.830	0.87	0.90	0.92	0.91	0.92
Jet Images	0.76	0.83 +/- 0.01	0.860 +/- 0.01	0.89	0.89	0.87	0.88

Table(1)

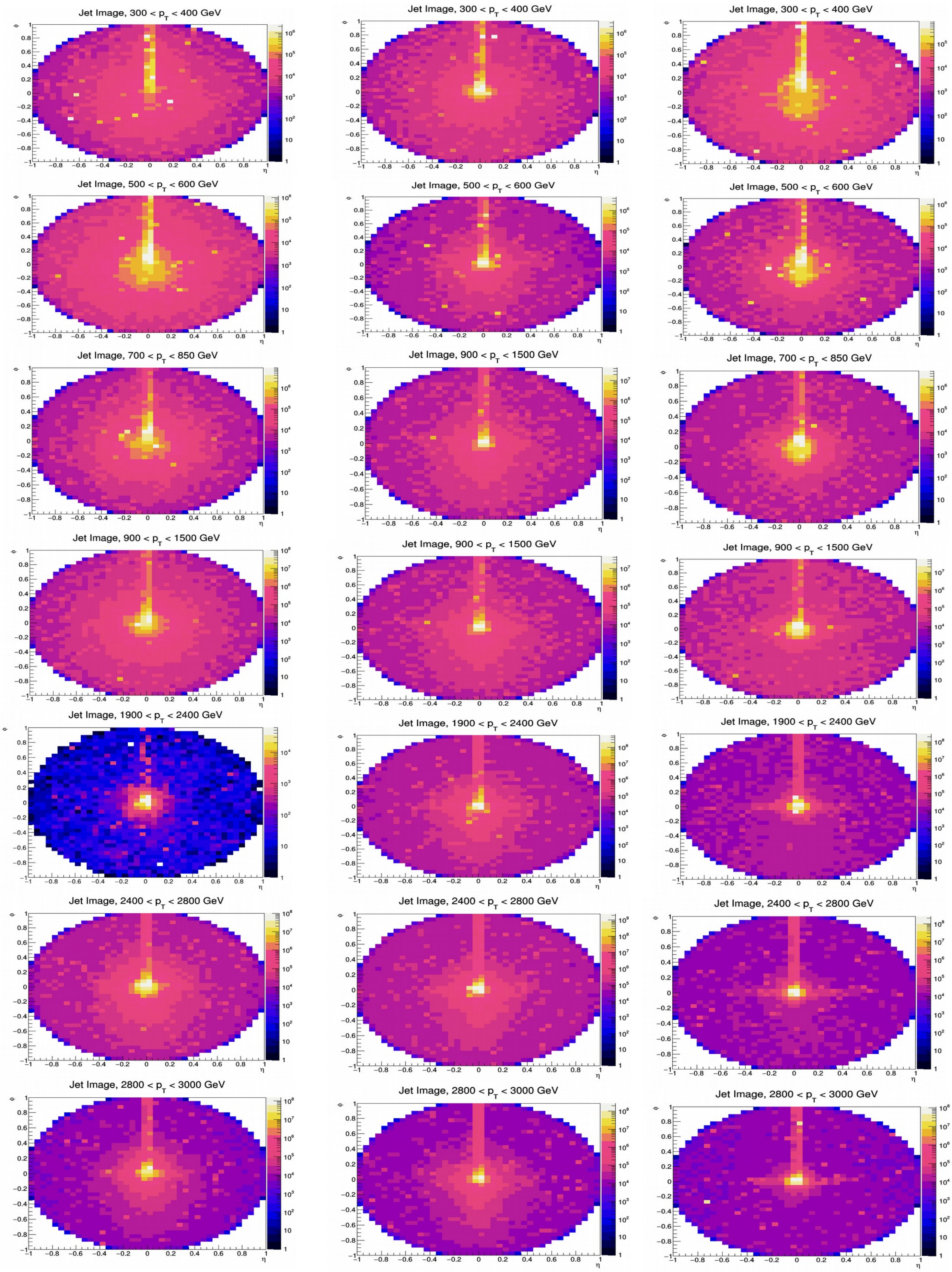
Another study was done in which background rejection was applied at various signal efficiencies within pt range (2400-2800 GeV) and it is obviously that the background rejection for W background is higher than that for the QCD background for both lund plane and jet images as shown in tables(2) and (3).

Signal Efficency	**50%**	**80%**	**90%**
50% W + 50%QCD	35.32	11.04	5.79
W	99.89	27.18	12.89
QCD	25.68	9.36	5.08

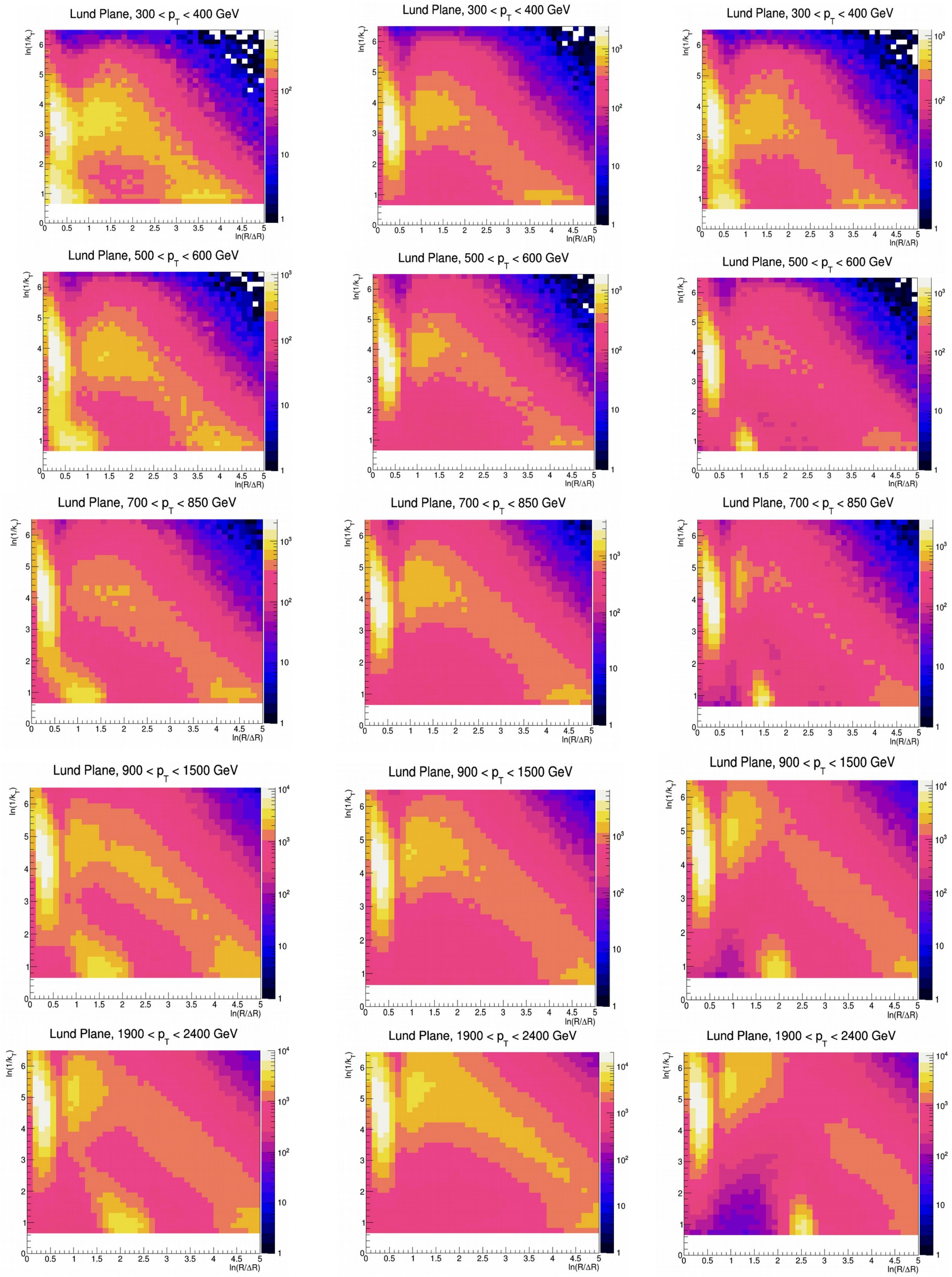
Table(2)

Signal Efficency	**50%**	**80%**	**90%**
50%W + 50%QCD	15.64	5.91	3.94
W	57.71	16.85	9.14
QCD	11.43	3.91	2.62

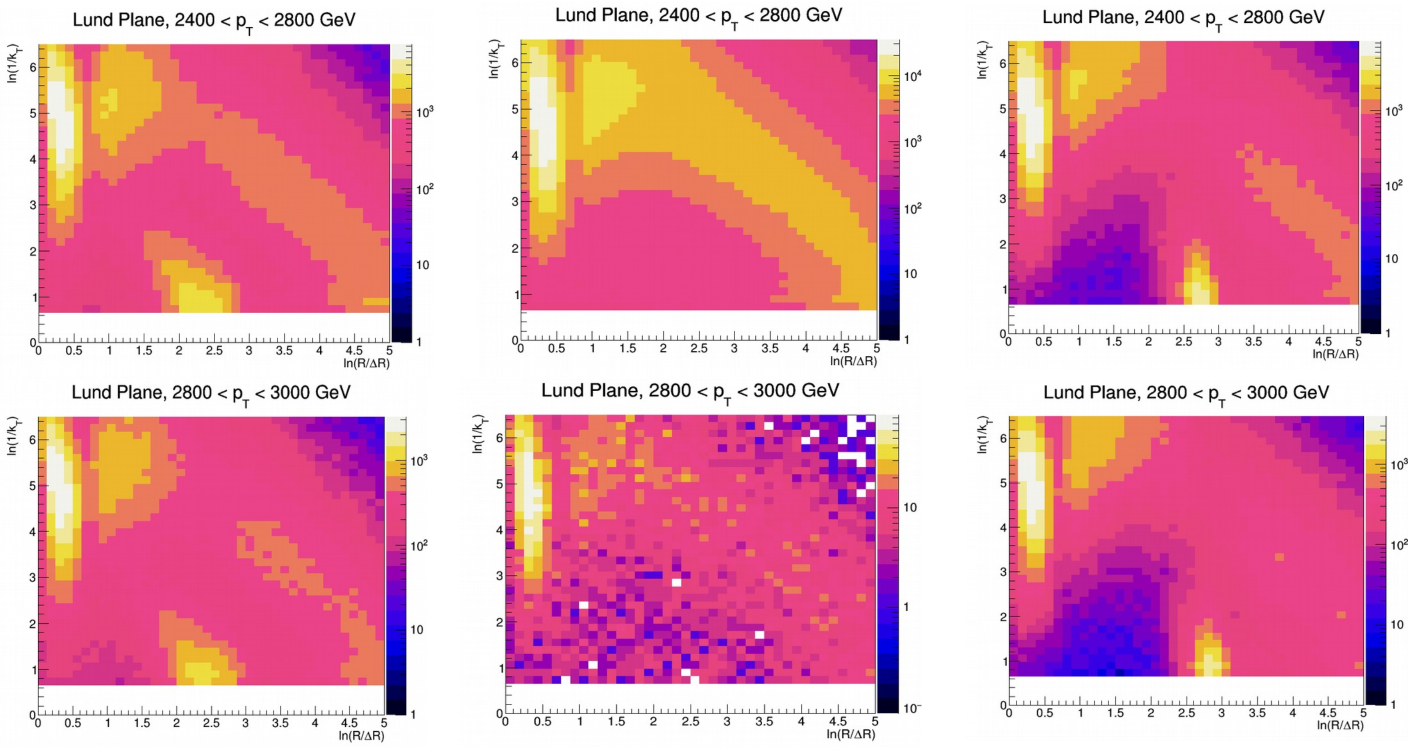
Table(3)



Figure(5) *represents the jet images created by translation and rotation for top (signal), QCD and W (background) jets respectively at various pt ranges*



Figure(6) *lund planes created for top (signal), QCD and W (background) jets at various pt ranges respectively*



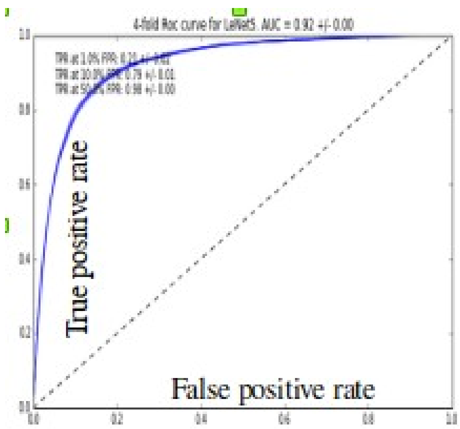


Figure (7) *ROC curve for CNN for certain pt range*

III. Conclusion

Performance of CNN for identification of top jets from the standard model (QCD + W) jets is better for the lund planes than for the jet images at higher pt-ranges. Rejection is higher for W background than QCD background for both lund planes and jet images.

IV. Outlook

Convolutional Neural Networks have shown a very good performance for the top jets tagging. As a kind of a future work completing this study LSTM model can be used for tagging jets.

V. Acknowledgments

Firstly, Special thanks to my supervisor Mario Campanelli for his invaluable guidance, patience, expertise and explanation. I would like also to thank Alexander Sopio for helping me a lot during the project and spend a lot of time assisting me despite his PhD work.

VI. References

[1] Nitish Srivastava et al. “Dropout: a simple way to prevent neural networks from overfitting”. In: The Journal of Machine Learning Research 15.1 (2014), pp. 1929–1958.

[2] Sachiyo Daely. “Analysis of lund jet plane using image recognition”.

[3] https://indico.cern.ch/event/764444/attachments/1731005/2797581/talk_milan18.pdf

[4]https://missinglink.ai/guides/keras/using-keras-flatten-operation-cnn-models-code-examples/