Electron Identification with Deep Neural Networks

Dominique Godin
Kazuya Mochizuki
Emre Onur Kahya
Arthur Boisvert
Jean-François Arguin
Alain Tapp
Otilia Ducu
Julien Maurer
Ioan-Mihail Dinu

ATLAS EXPERIMENT

Université de Montréal

Electron Identification Motivation

- Though relatively rare, prompt electrons are very important for ATLAS experiment, Higgs' physics program and search for physics beyond Standard Model

- Example: 4 out of 5 Higgs' main signatures at LHC contain e^-

- Electron classification is therefore crucial to discriminate prompt electrons from other electron-like signatures

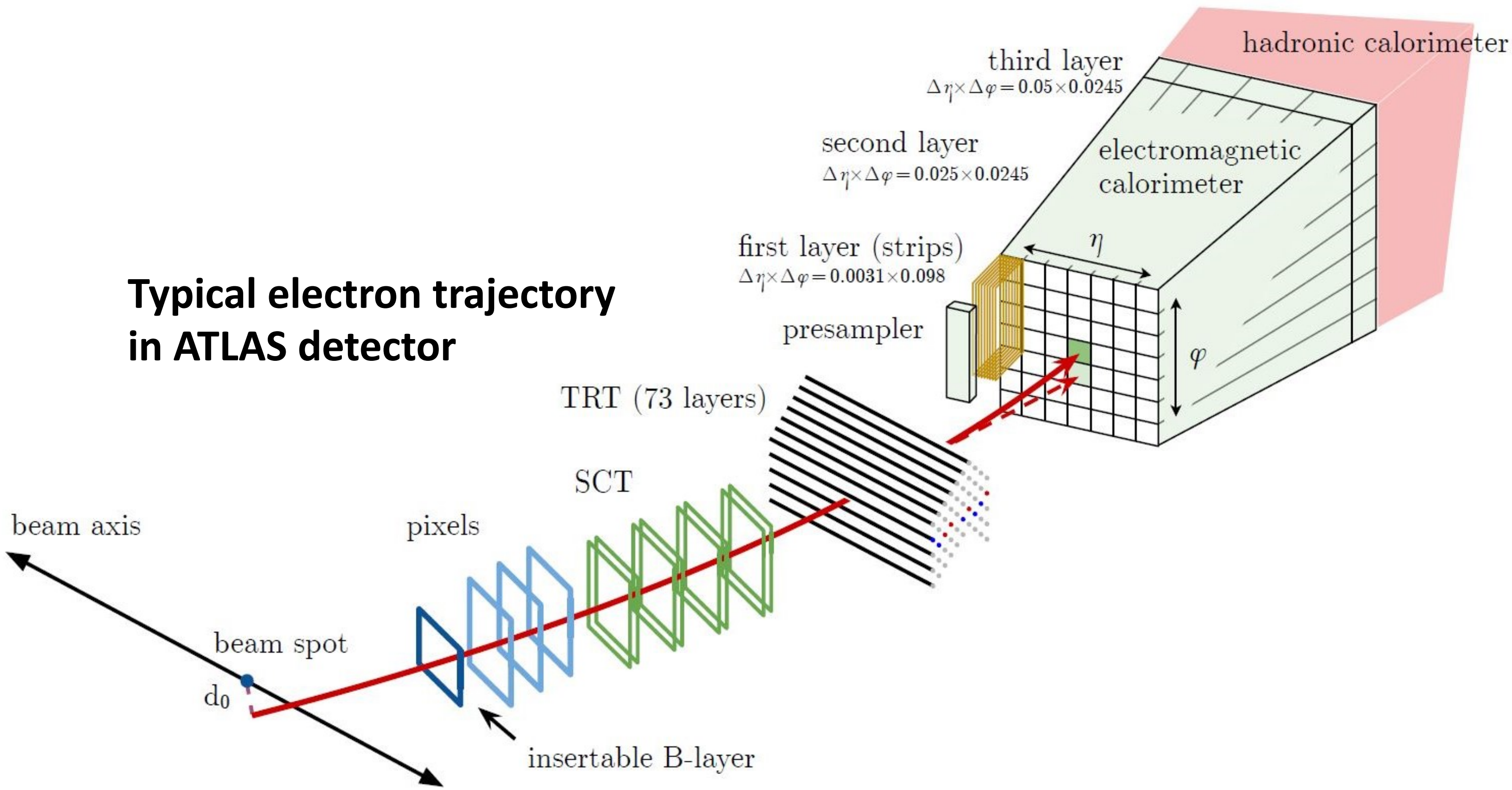
- Machine learning could replace the more simple algorithms that are presently used in ATLAS for this identification

- The goal of this project is to use the best techniques available to eventually classify e^- as of many different types as we want (i.g. signal, charge-flipped, from b-jets, γ-conversion, fake)

Electron Data from ATLAS Detector

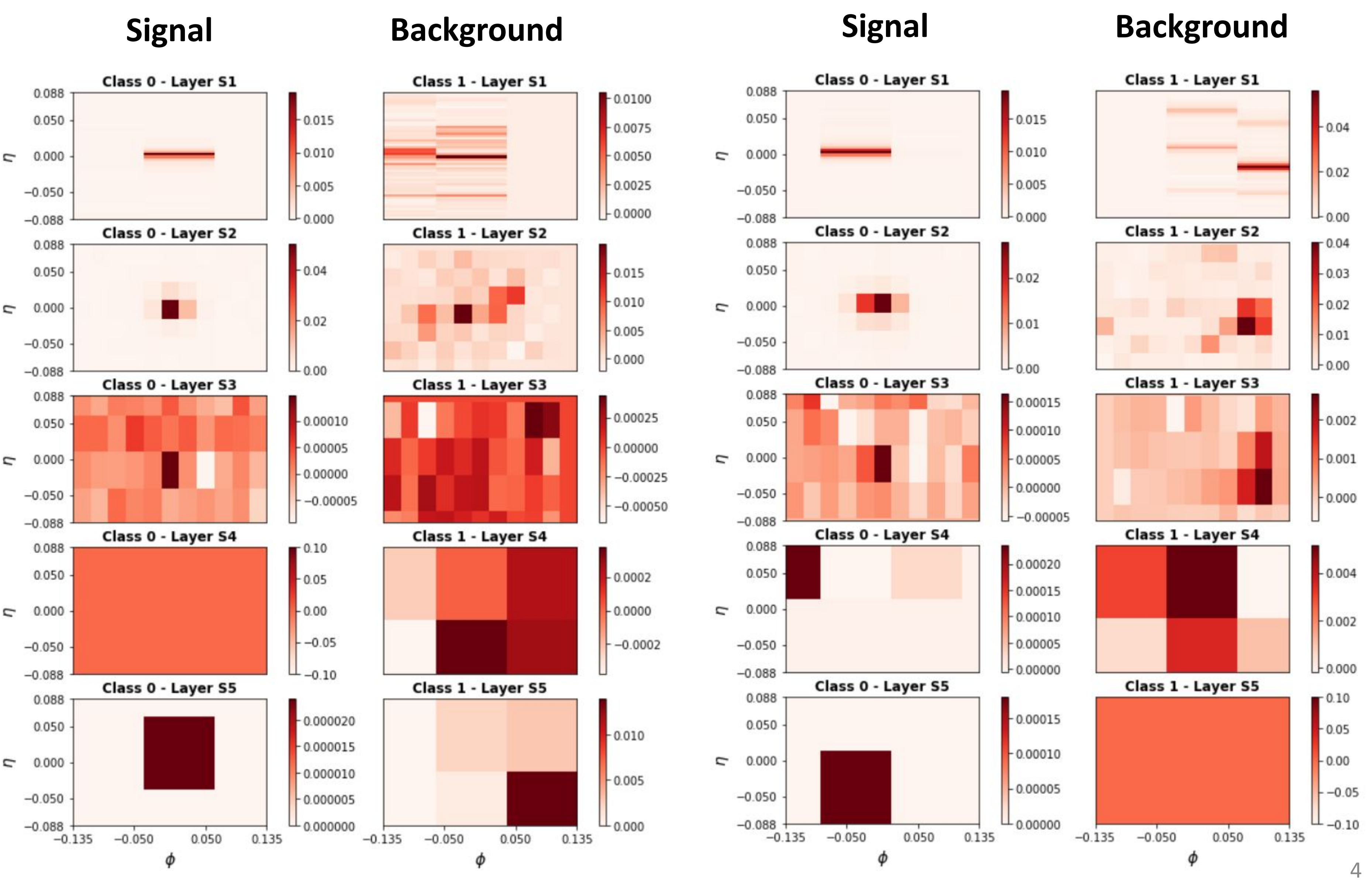
Available MC data for each electron

• 15 track candidates (e_frac, d_eta, d_phi, d_0) :	(15 x 4)
• 3 images from electromagnetic calorimeter:	3 x (56 x 11) in (η x ϕ)
• 2 images from hadronic calorimeter:	2 x (7 x 11) in (η x ϕ)



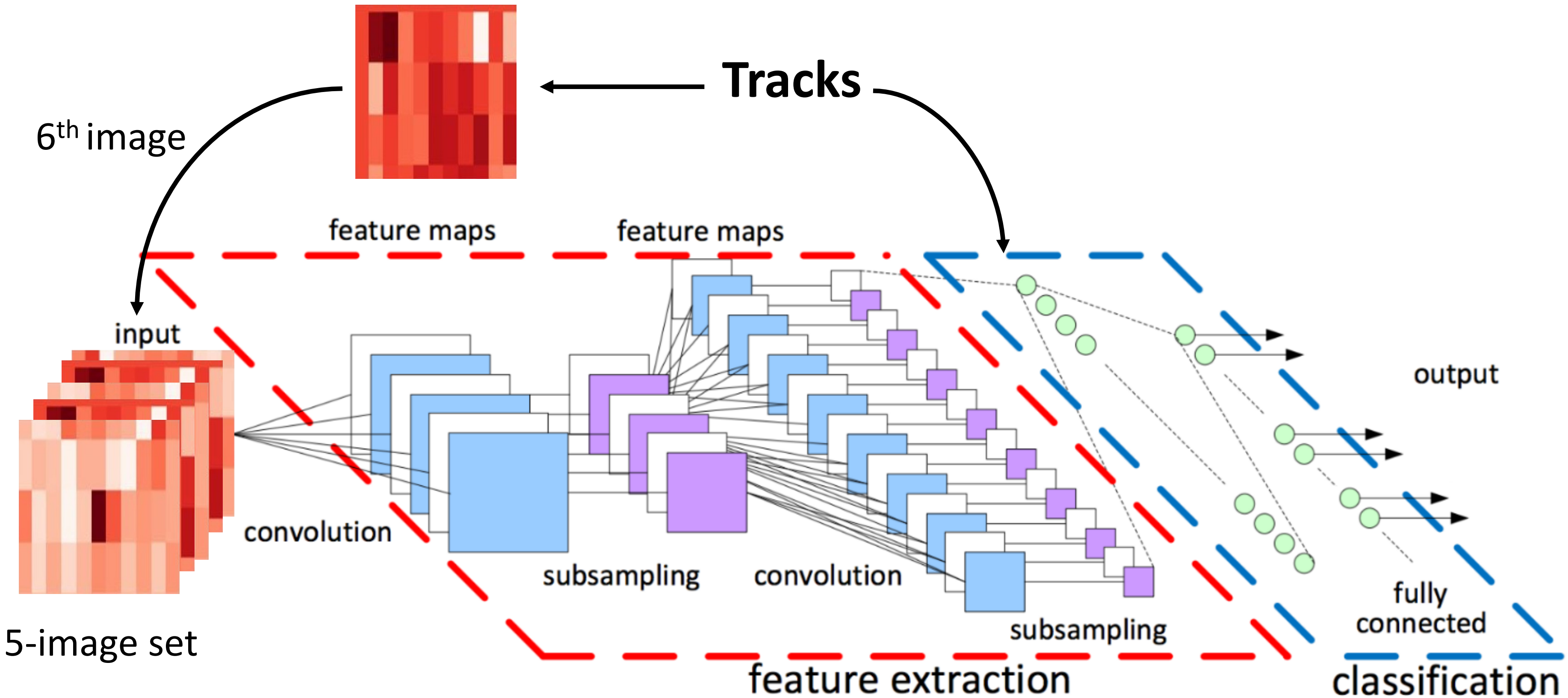
Typical electron trajectory in ATLAS detector

Calorimeter Images MC Examples



Preliminary NN Architecture

- Binary classification: 1) prompt e^- 2) fake + non prompt e^-
- Multi-channels CNN with calorimeter images + Tracks information
- The 5-image sets are concatenated in volume images (5-channel images)
- Tracks: either used as an image (6th channel) or concatenated in the FC layer
- Both methods seems to produce similar performance



Inside CNN Architecture

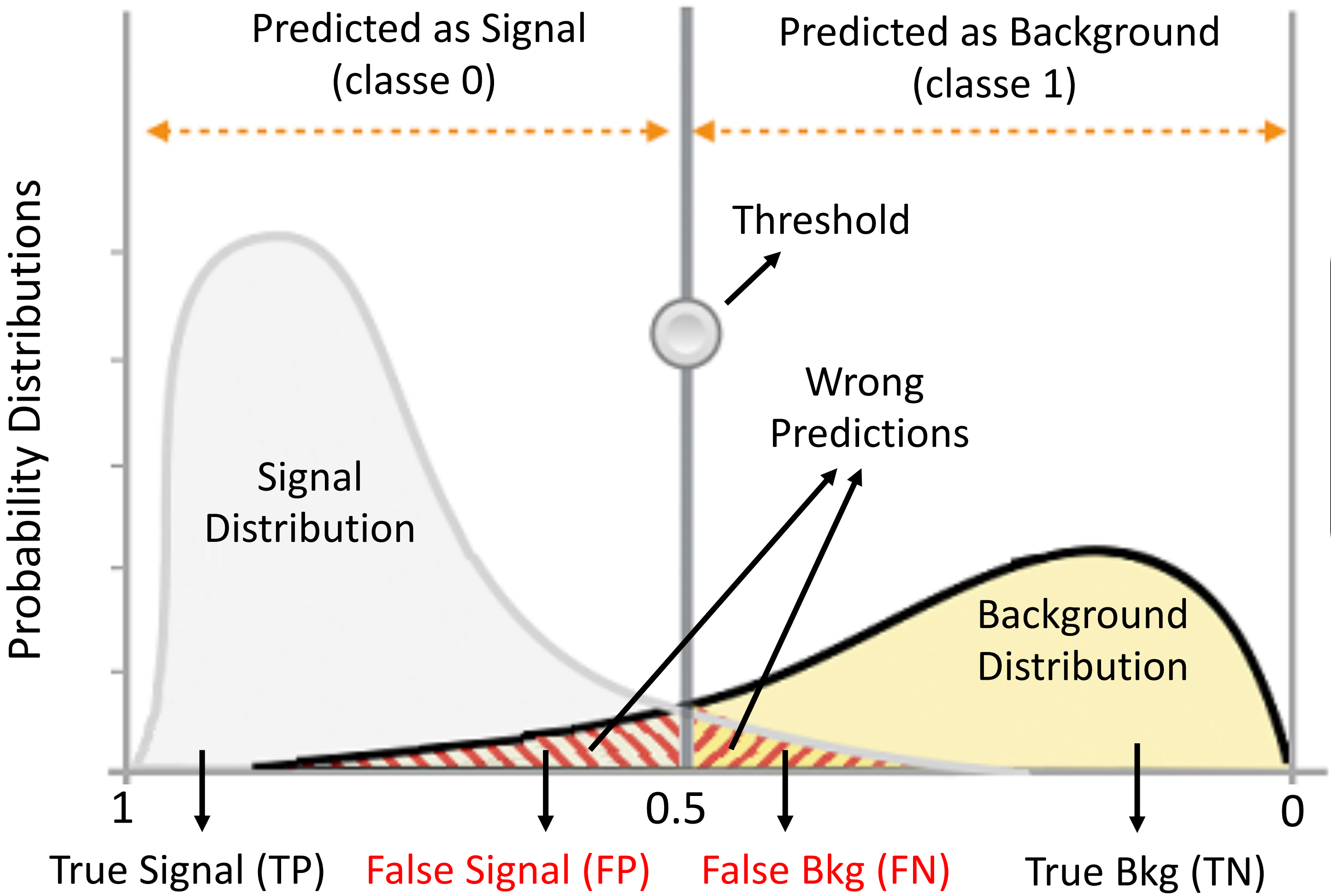
Detailed Architecture

- Layer 1: 100 convolution maps with 3x3 filter
- Layer 2: 2x2 MaxPooling
- Layer 3: 50 convolution maps with 2x2 filter
- Layer 4: 2x2 MaxPooling
- Layer 5: maps vectorization tracks information concatenation
- Layer 6: 100-neuron fully connected
- Layer 7: binary Softmax layer (output probability-vector)

Capacity: **118 652** trainable parameters (i.e. weights) altogether

Classification vs ROC Curves

- NN with Softmax layers gives class probabilities for each sample (e.g. 0.43, 0.57)
- ROC curves are used in binary classification to study the output of a classifier
- Goal: maximize the true positive rate while minimizing the false positive rate



$$TPR = \frac{TP}{TP + FN}$$

vs

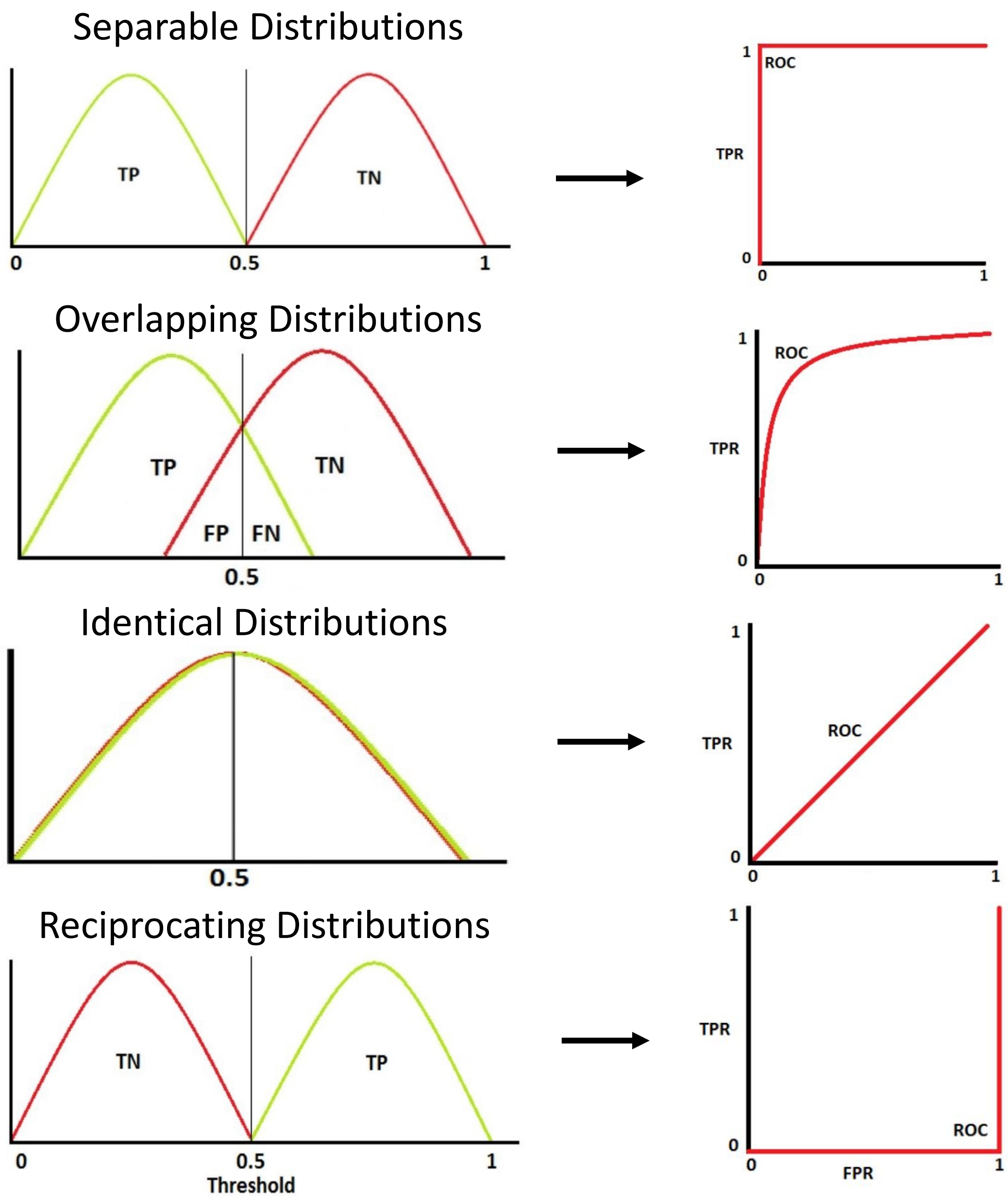
$$FPR = \frac{FP}{FP + TN}$$

More on ROC Curves

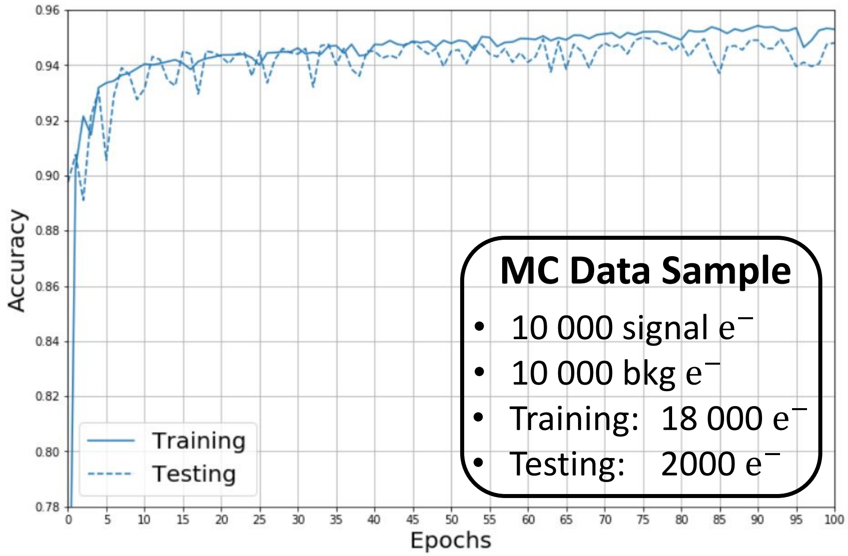
- **ROC curves**: better figure of merit than accuracy
- **AUC** (area under curve): also very useful to estimate overall performance

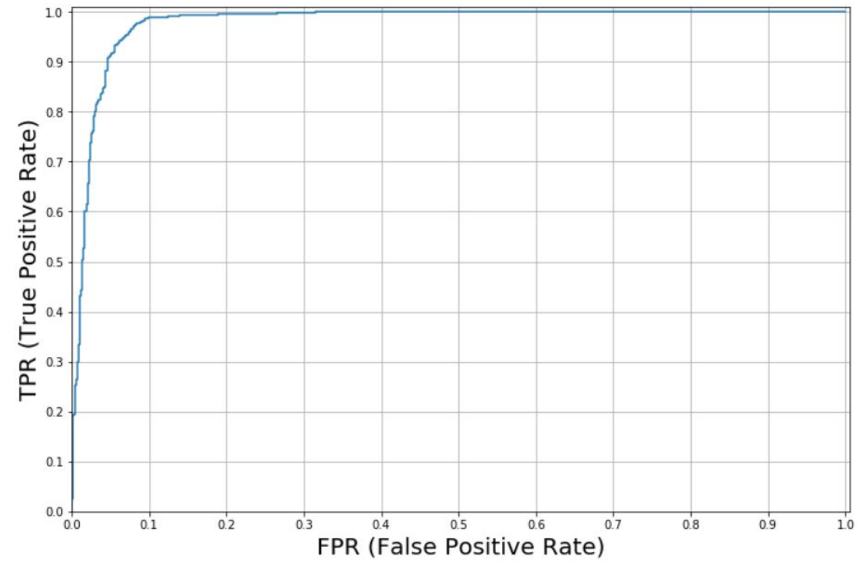
$$TPR = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{FP + TN}$$

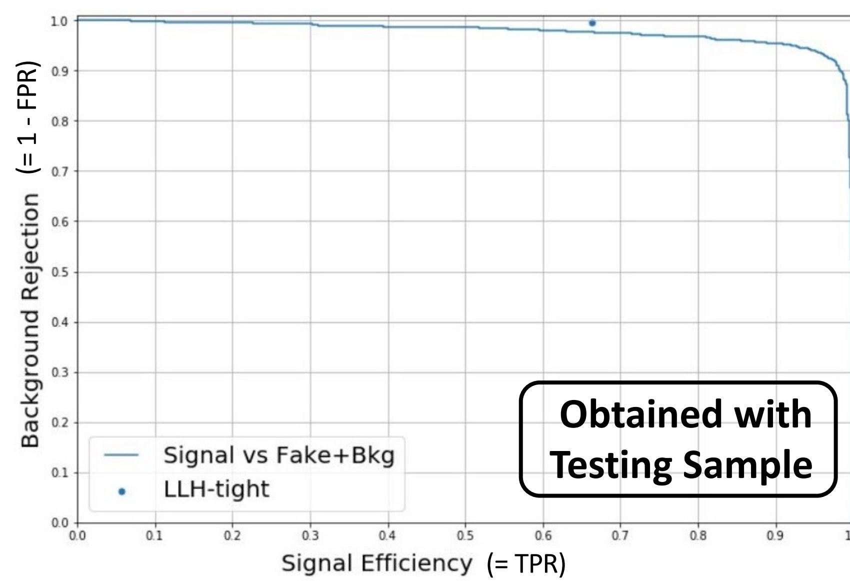


CNN Results: Accuracy and ROC Curve



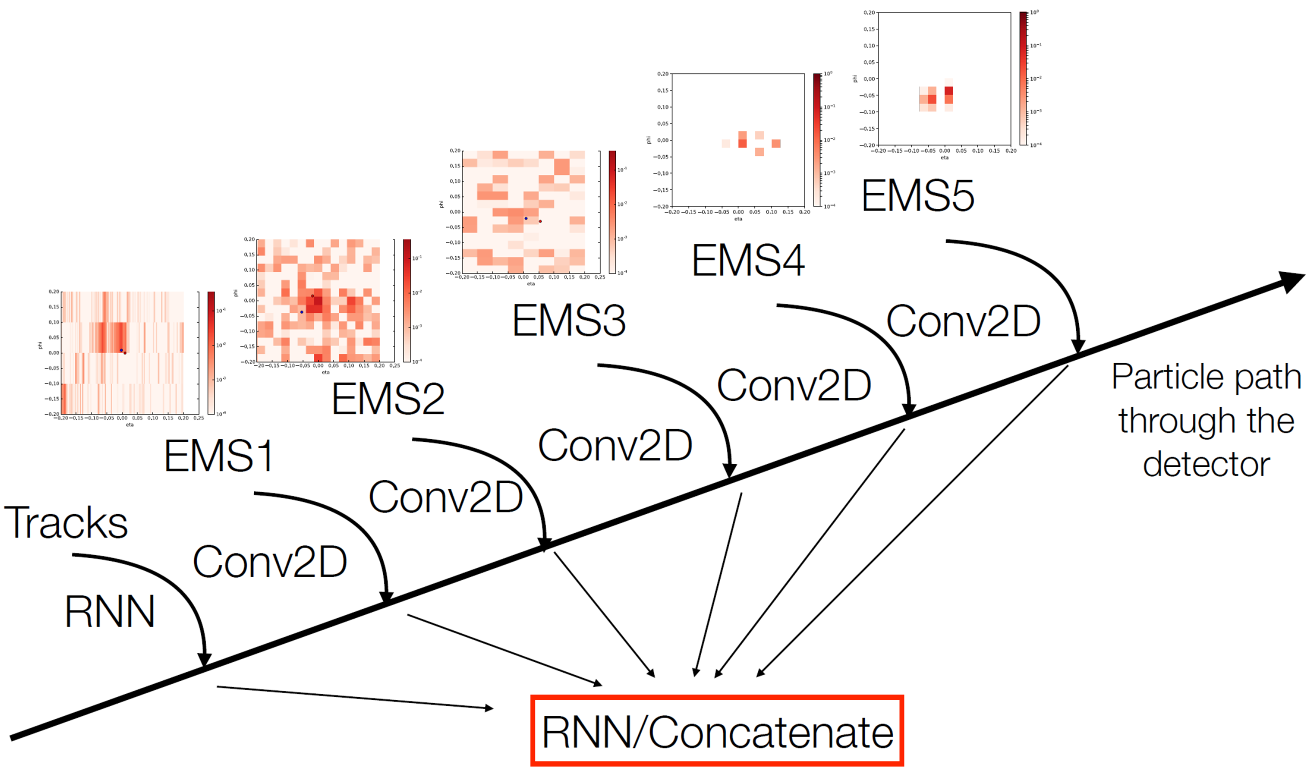


- Maximum accuracy is achieved rapidly
- 4s/epoch (generator + 4 CPU's + RTX 1080Ti)
- Batch size of 500 e^{-} was used
- No sign of serious overfitting
- Bkg vs Signal ROC curve is obtained by switching axis, expressing 1- FPR vs TPR
- LLH-tight is still performing better but NN might outperform loose and medium LLH
- Future: deeper architectures and 5-class NN



Future: RNN and 5-class Identification

- 5-class e^- identification: 1) signal 2) charge-flipped 3) from b-jets
  4) γ-conversion 5) fake
- More advanced architecture: RNN made up of 6 sub-networks (1 RNN + 5 CNN's)



- Such an architecture already showed promising results for leptons τ classification

Conclusion and Future Developments

- Electron identification is of great importance for the Atlas experiment
- The electron classification into different classes is a challenging problem
- CNN’s and RNN’s showed to be effective for τ and jets classification

Long-term possible developments

- Transition from RNN to LSTM for longer tracks or images sequences
- Training with low-level variables available from the ATLAS detector
- Eventually performing 5-class identification with real data instead of MC
- Adapt architectures to regression for estimation of variables (e.g. E ou P)
- Explore pre-processing methods for better image information extraction