# Deep Neural Network for ALICE Prompt Photon Discrimination

Yue Shi Lai

LBNL RNC

February 14, 2017

# Introduction

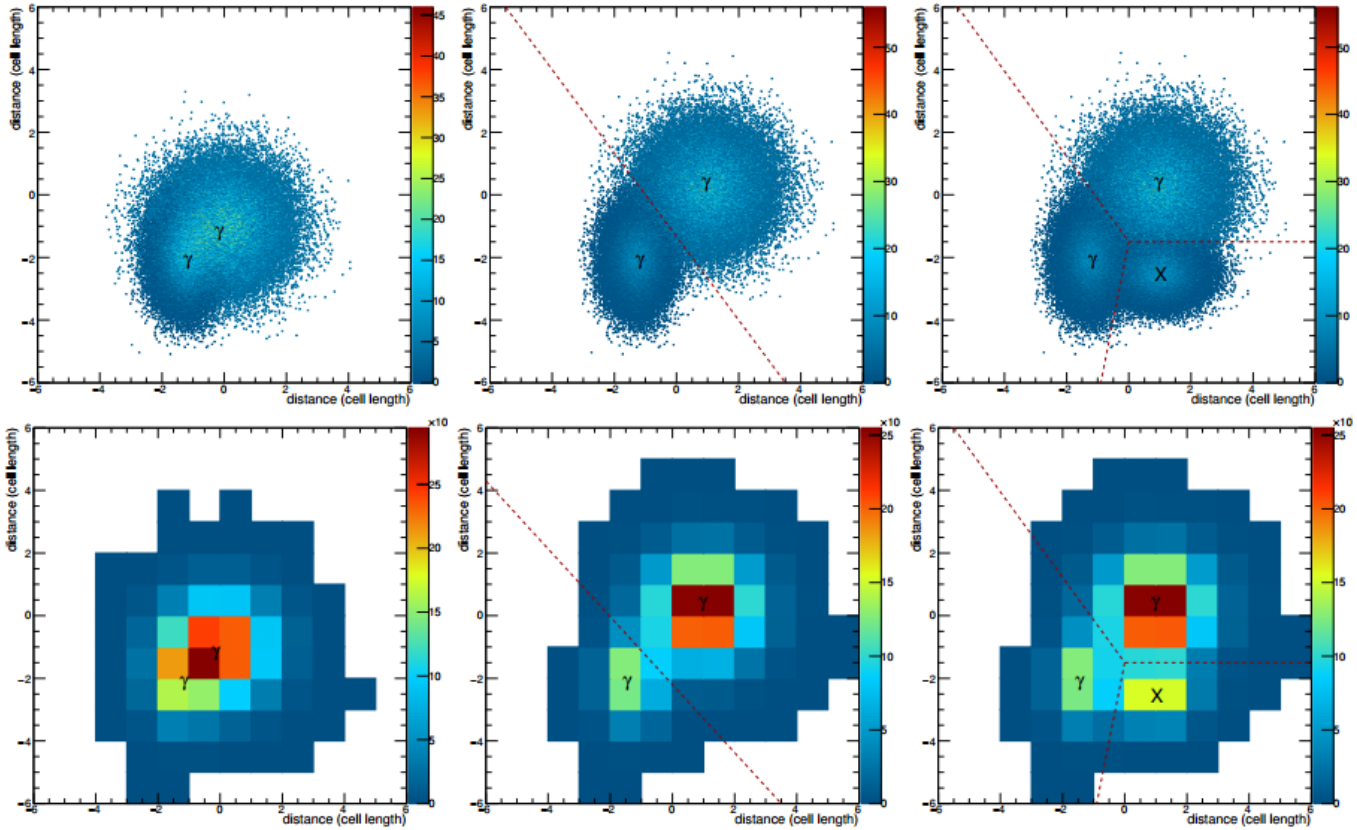
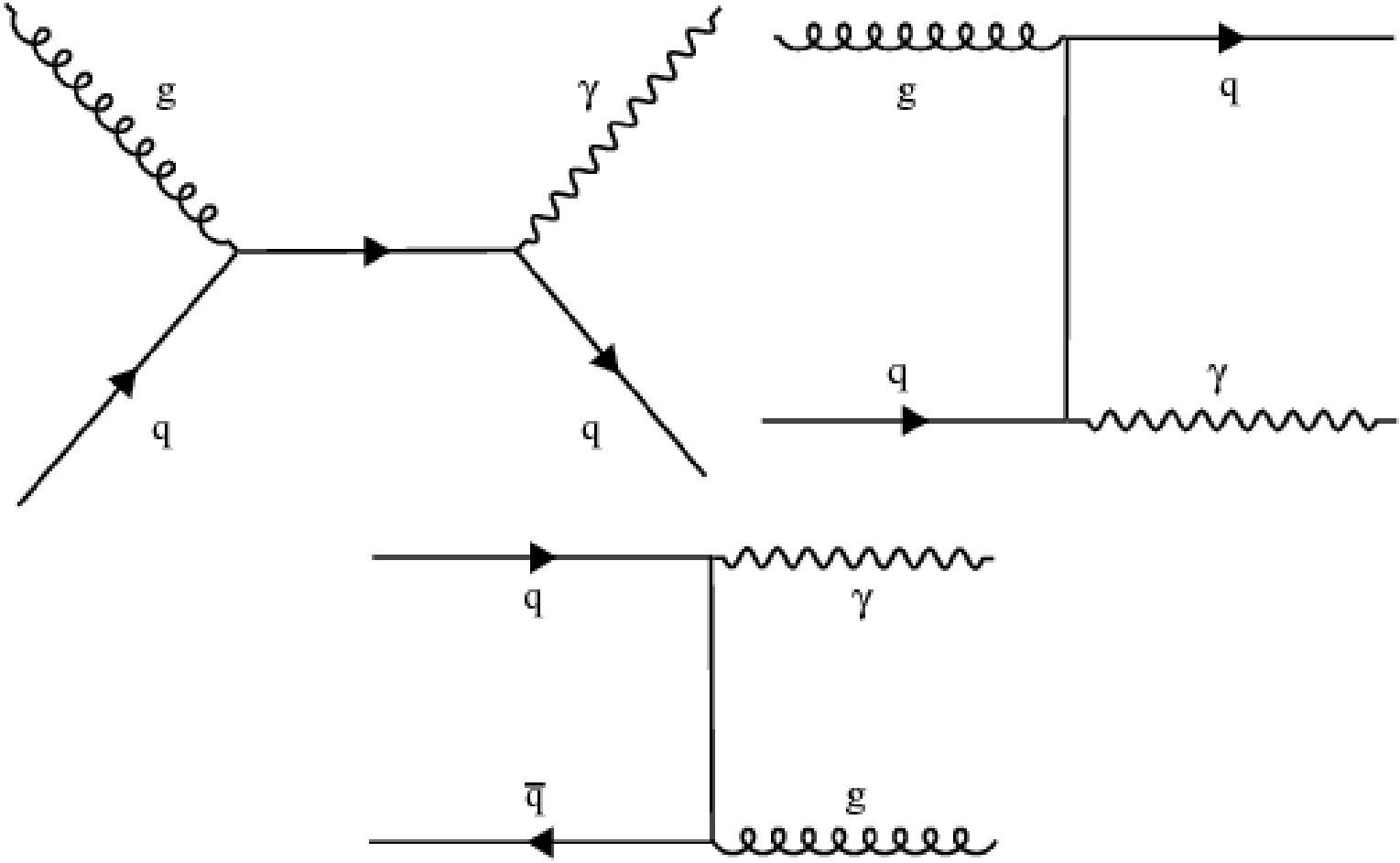



**Figure 8:** Energy deposit of neutral pions and their decay products in the calorimeter with different cluster shapes with fine granularity (upper row) and how they would look with the approximate resolution of the EMCal (lower row). The red dashed lines indicate the approximate splitting done by the $v2$-clusterizer, while the $v1$ clusterizer would see them as one single cluster with different number of local maxima.

- Electromagnetic calorimeters (ECAL/EMCAL) operate really like imaging devices
- The problem here is to discern objects that are at pixel/subpixel level

F. Bock, ALICE-AN-3067 (unfortunately internal)

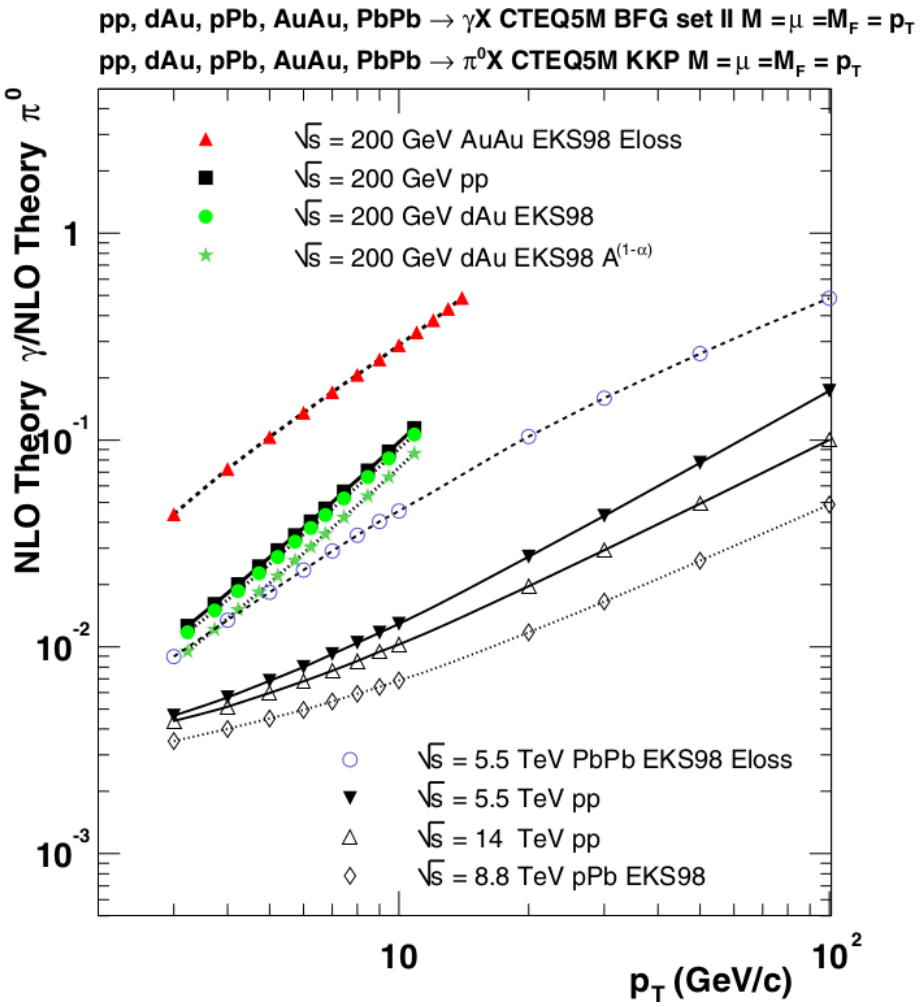
# Introduction




H. Hesari *et al.*, Phys. Rev. D91 (2015), 057502, https://doi.org/10.1103/PhysRevD.91.057502

- One of the most important measurement with EMCAL is the prompt/direct photon
- To leading order the kinematic tag for to the quark/gluon production

# Introduction




F. Arleo *et al.*, https://arxiv.org/abs/hep-ph/0311131

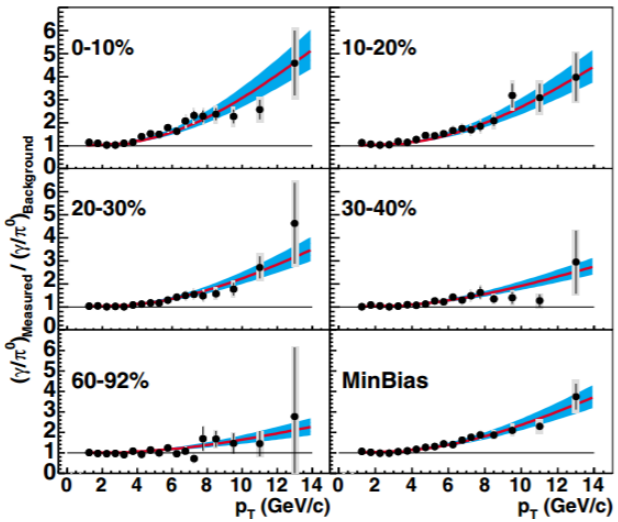
- But there is always a large background from $\pi^0 \rightarrow \gamma\gamma$
- Signal/background for 10-20 GeV/$c$ is $\approx$ 0.05–0.1

# Goal

$$R_\gamma = \frac{(\gamma/\pi^0)_{\text{Measured}}}{(\gamma/\pi^0)_{\text{Background}}} \approx \frac{\gamma_{\text{Measured}}}{\gamma_{\text{Background}}}, \qquad (1)$$

and any significant deviation of the double ratio above unity indicates a direct photon excess. In Fig. 1 an excess

- $\gamma_{\text{inclusive}} = \gamma_{\text{prompt}} + \gamma_{\text{decay}}$
- $R_\gamma = \dfrac{\text{inclusive}}{\text{decay}}$ can be measured
- But you have to distinguish $\gamma$ vs. $\pi^0$ to high momentum to make the measurement




https://doi.org/10.1103/PhysRevLett.94.232301

# Previous Work

$$\lambda_0^2 = 0.5(\delta_{\phi\phi} + \delta_{\eta\eta}) + \sqrt{0.25(\delta_{\phi\phi} - \delta_{\eta\eta})^2 + \delta_{\eta\phi}^2}, \quad \text{(B.1)}$$

$$\lambda_1^2 = 0.5(\delta_{\phi\phi} + \delta_{\eta\eta}) - \sqrt{0.25(\delta_{\phi\phi} - \delta_{\eta\eta})^2 + \delta_{\eta\phi}^2}, \quad \text{(B.2)}$$

where $\delta_{\phi\phi}$, $\delta_{\eta\eta}$ and $\delta_{\phi\eta}$ are weighted coefficients by the cell energy:

$$\delta_{\alpha\beta} = \sum_i \frac{w_i \alpha_i \beta_i}{w_{tot}} - \sum_i \frac{w_i \alpha_i}{w_{tot}} \sum_i \frac{w_i \beta_i}{w_{tot}}, \quad \text{(B.3)}$$

$$w_i = TMath :: Max(0, w_0 + \ln(\frac{E_i}{E_{cluster}})), \quad \text{(B.4)}$$

$$w_{tot} = \sum_i w_i, \quad \text{(B.5)}$$

where $w_0 = 4.5$ [9] [5], $i$ indicates a cell that belongs to the cluster and $\eta_i$ and $\phi_i$ are the indexes of the cell inside a EMCal super-module in longitudinal and azimuthal direction, respectively, with $0 \leq \eta_i < 48$ and $0 \leq \phi_i < 24$.

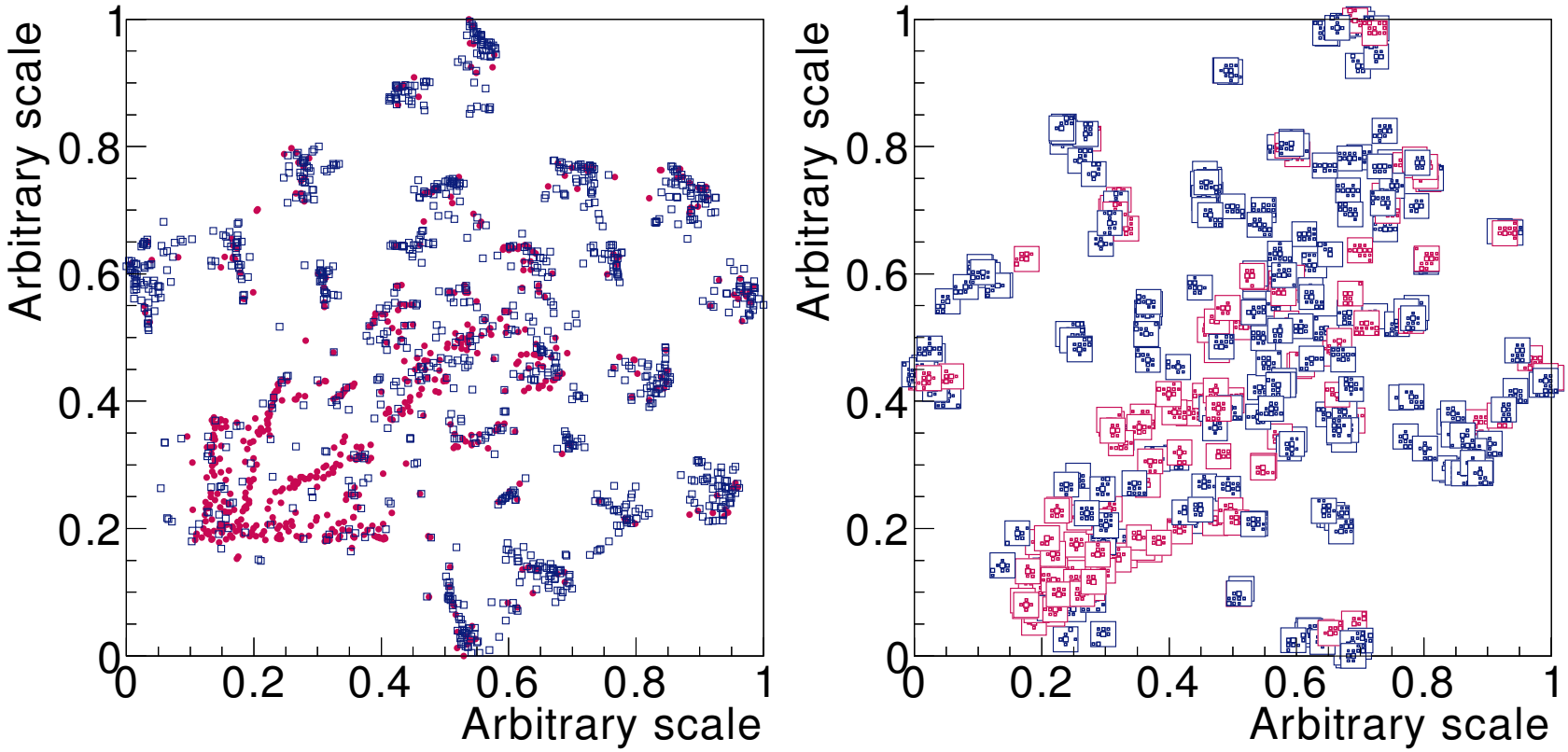
- The traditional discriminant the ALICE experiment is using
- Basically an ellipse fit, where the major axis is the discriminant

H.-t. Zhang, ALICE-AN-2326 (unfortunately internal)

# Ingredients

- ALICE LHC15a3b (EM-enriched QCD MC, generator-level vs. GEANT), 4 $\hat{p}_{\perp,\mathrm{min}}$ bins, 80054 training and 80054 test clusters
- Keras 1.1.2, TensorFlow 0.11.0
- 8–16 GeV clusters
- Extract $5 \times 5$ cells $c_i = E_i^{\mathrm{cell}} / E^{\mathrm{cluster}}$, $i = 1, \ldots, 25$ around each cluster's maximum cell, add cluster's $\eta$
  - $(c_1, \ldots, c_{25}, \eta)$
- Rectified Linear Unit (ReLU) activation (not possible with TMVA)
- 26 parameters $\rightarrow$ 512 neurons $\rightarrow$ 512 neurons $\rightarrow$ 512 neurons $\rightarrow$ 512 neurons $\rightarrow$ (softmax to) 2 neurons (not possible with TMVA)
- Tap 1 of the 2 output neurons as prompt $\gamma$ discriminant
- Dropout at probability 0.1 for regularization (not possible with TMVA)
- 128 batches, 12 epochs, $\approx$ 5 minutes training on a GPU (not possible with TMVA)

# Data in t-SNE

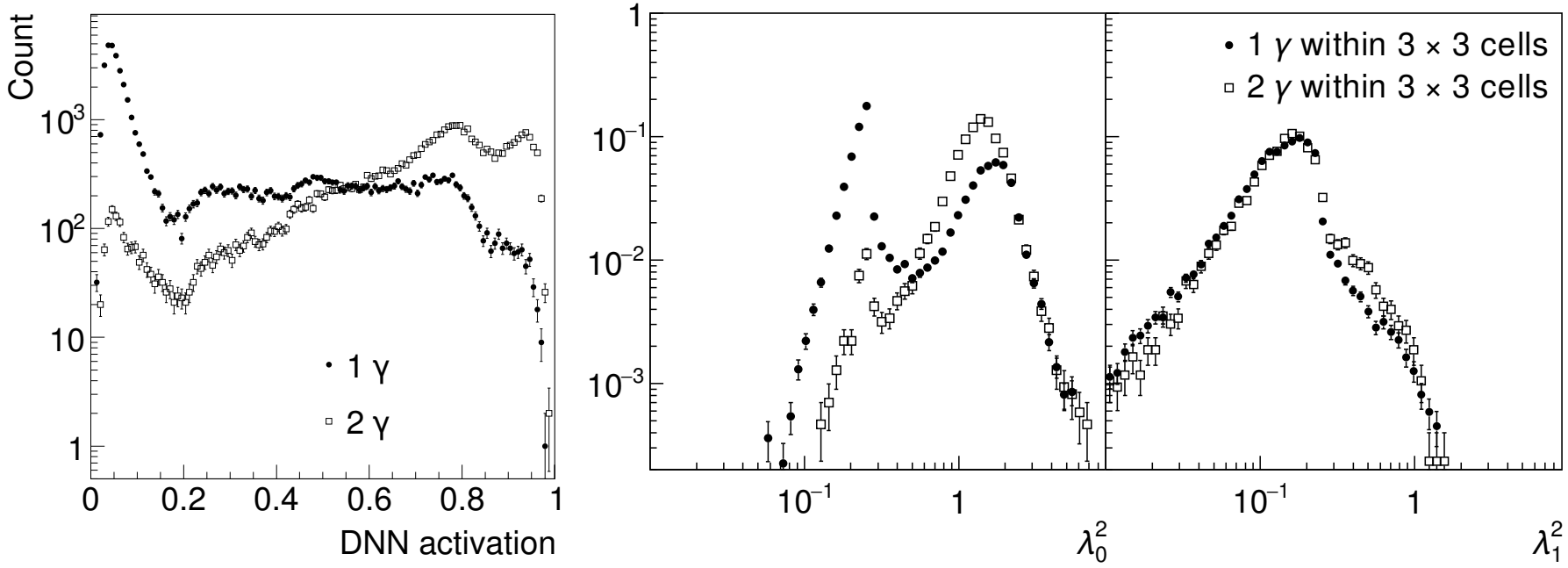



- You get a region with well isolated photon, radiating to a diffuse region, then 20 regions of clearly detectable two photons
- 20 regions = 5 $\times$ 5 minus the middle cross (5)
- Problem is clearly not rotationally symmetric
- Some of the ongoing work is to trace the “contamination” by problems in the detector simulation (alignment)

# How it looks in Keras

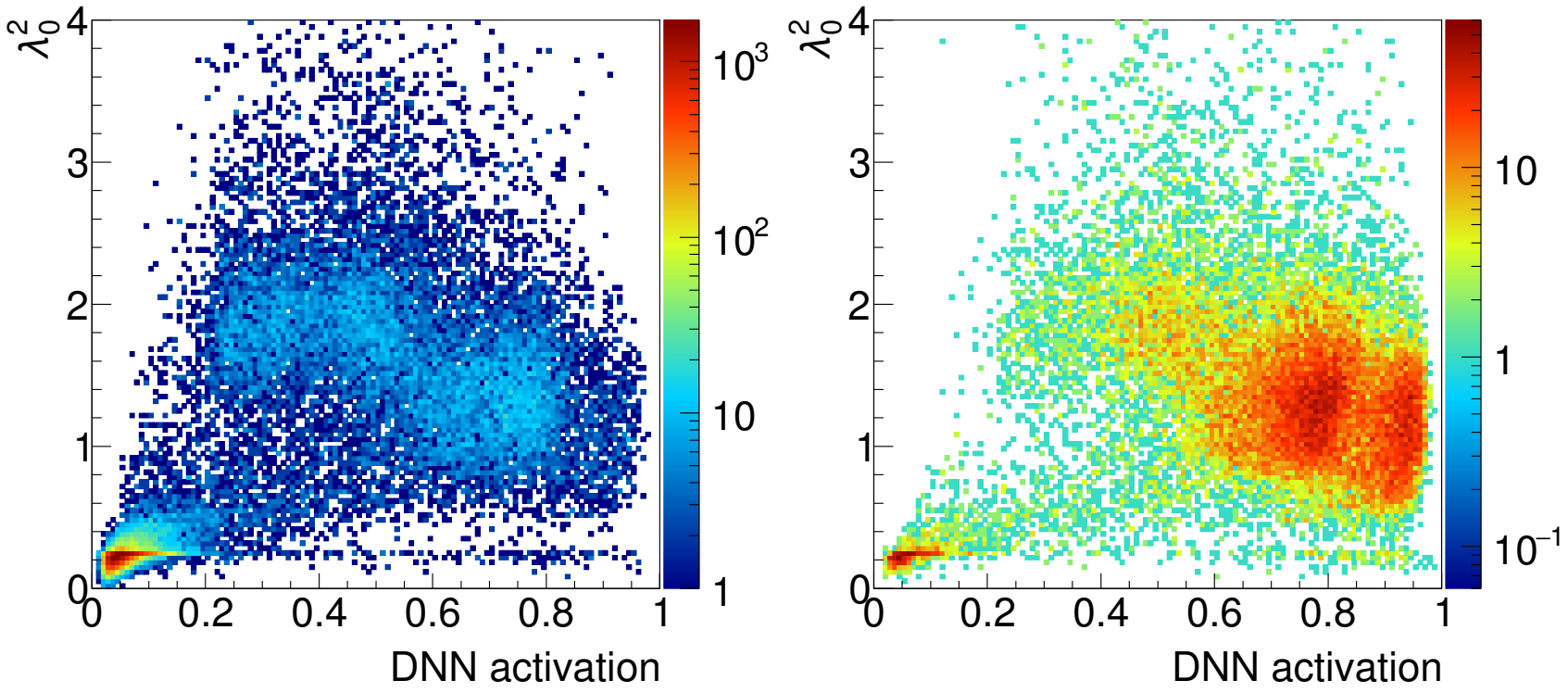
```
model = Sequential()
ndense = 512
dropout = 0.1

model.add(Dense(ndense, input_shape=(nfeature,)))
model.add(Activation('relu'))
model.add(Dropout(dropout))
for i in range(4):
    model.add(Dense(ndense))
    model.add(Activation('relu'))
    model.add(Dropout(dropout))
model.add(Dense(2))
model.add(Activation('softmax'))

model.compile(loss='categorical_crossentropy',
              optimizer='adadelta',
              metrics=['accuracy'])
```

# DNN activation




- A much higher discrimination power for photons, where signal/background gain $\approx$ 20–30
- Has also a high purity $\pi^0$ region

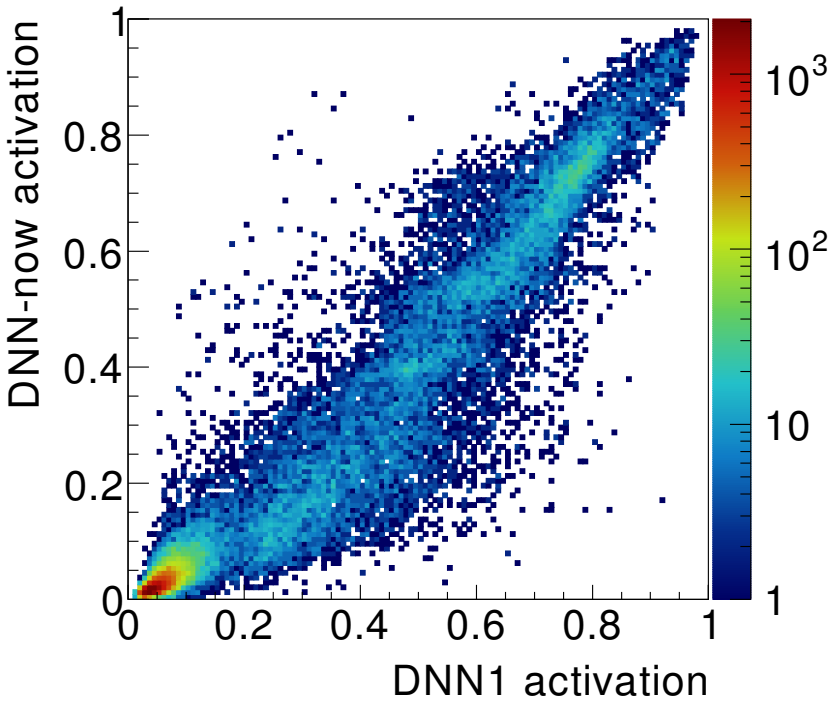
# Scatter plot DNN vs. $\lambda_0^2$



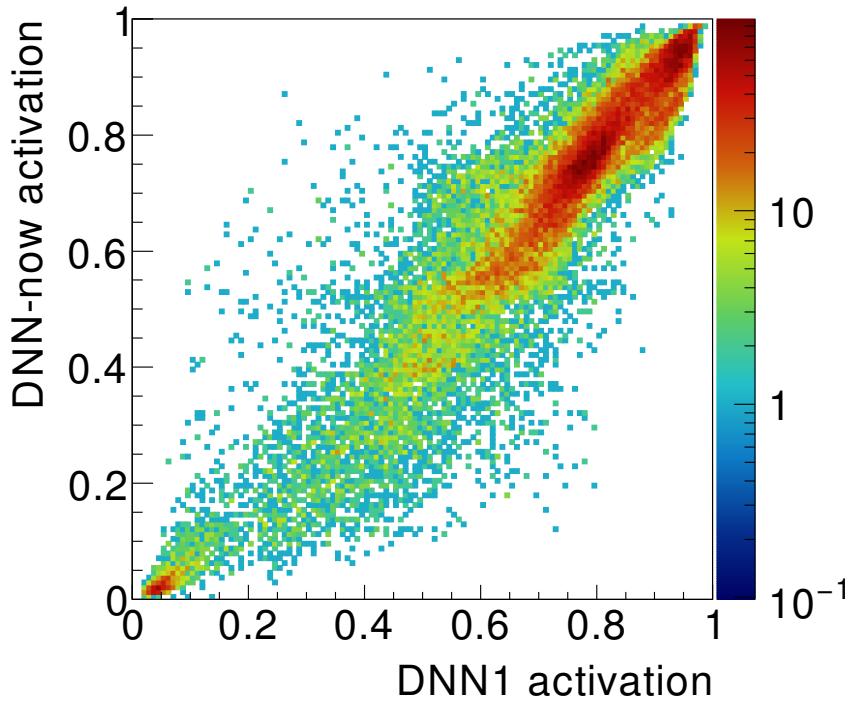

1 γ 2 γ

- DNN recovers γ at $\lambda_0^2 \approx 0.4$ and out to 1.5
- There is also a 2nd, low purity region the DNN detects, around $\lambda_0^2 \approx 1.5$–2 (possibly clusters not properly split)
- DNN rejects a "stripe" of γ at $\lambda_0^2 \approx 0.2$ and poor SNR

# Scatter plot among two generations of DNNs



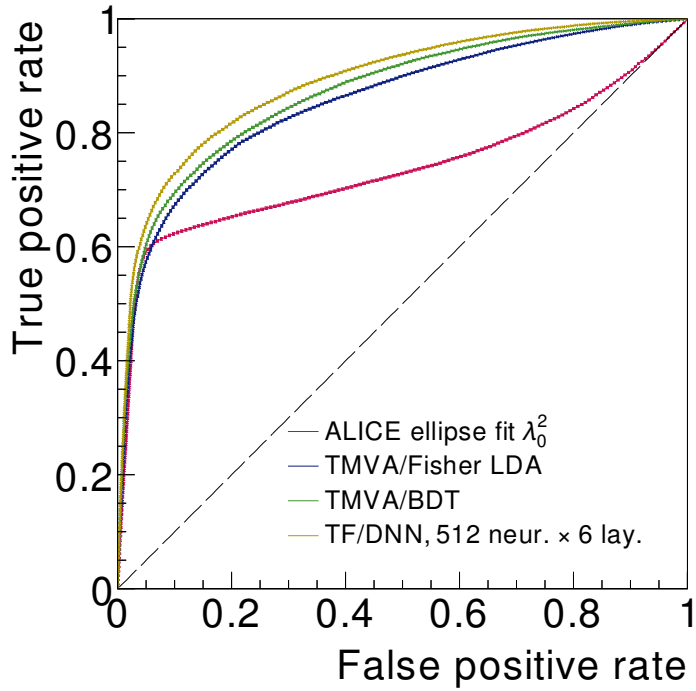

1 γ



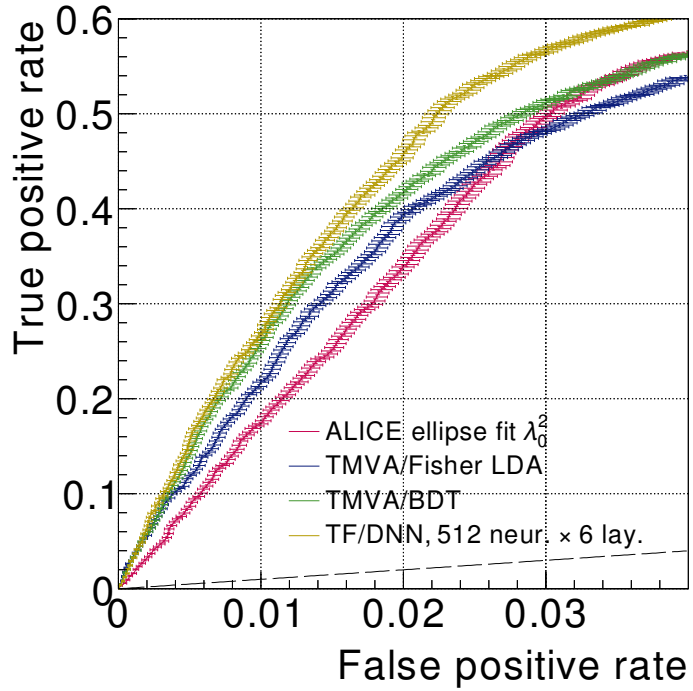

2 γ

- Change is to center/scale the individual values by mean and $\sigma$ to -1...1
- Mostly a non-linear function

# ROC curve comparison



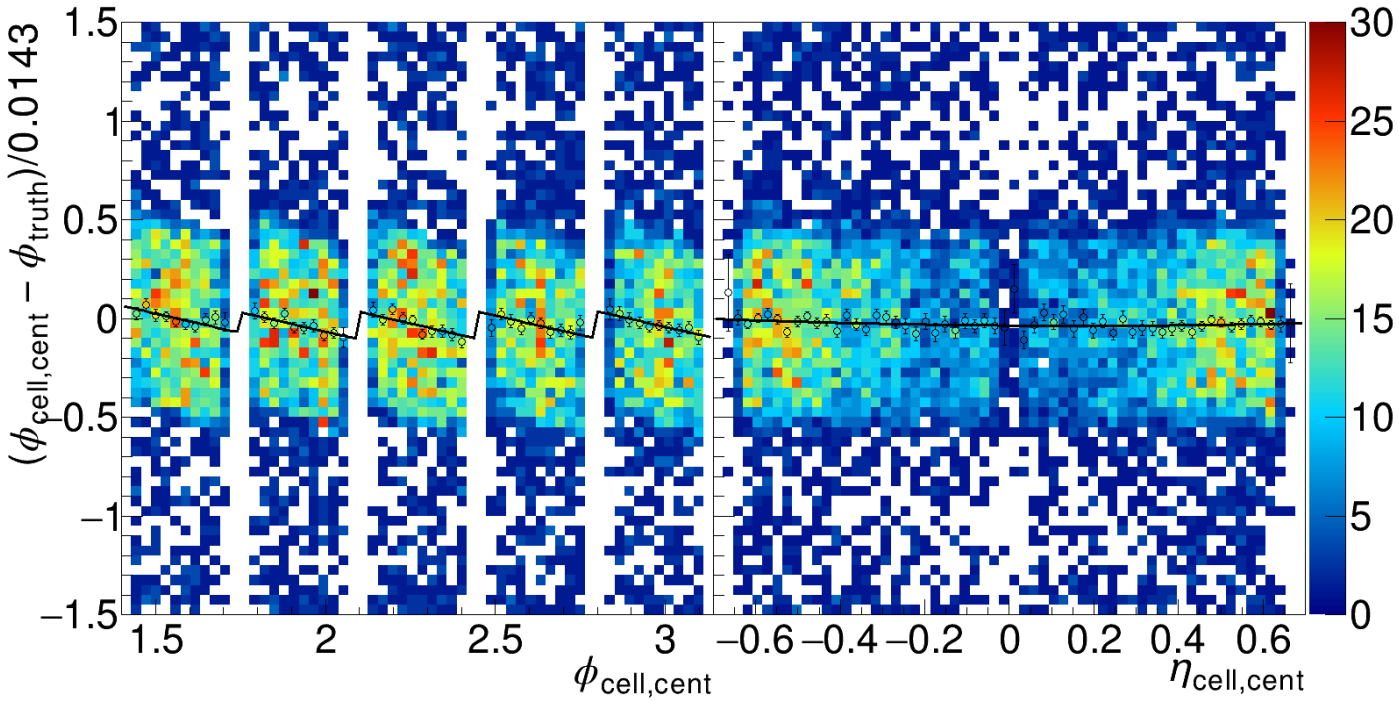

"Normal" (e.g. engineering)




HEP/TMVA orientation

- Error bars are 68% Jeffreys $B(1,1)$ ("Bayesian") confidence interval
- DNN consistently outperforms $\lambda_0^2$ by $\approx$ 30–40%

# Ongoing Work




- Detector simulation checks
- Embedding into data background
- Scaling study on Cori phase I/II