Deep Learning for Inferring Cause of Data Anomalies

M.Borisyak[1,2], D.Derkach[1,2], O.Koval[2,3], F.Ratnikov[1,2], A.Ustyuzhanin[1,2]
V.Azzolini[4], G.Cerminara[5], G.Franzoni[5], F. De Guio[6], M.Pierini[5], A.Pol[7], F.Siroky[5], J.R.Vlimant[8]

[1] NRE Higher School of Economics, Moscow, Russia
[2] Yandex School of Data Analysis, Moscow, Russia
[3] Skolkovo Institute of Science and Technology, Moscow, Russia
[4] Massachusetts Institute of Technology, Cambridge, USA
[5] CERN, European Organization for Nuclear Research, Geneva, Switzerland
[6] Texas Tech University, Lubbock, USA
[7] Universtite Paris-Saclay, Paris, France
[8] California Institute of Technology, Pasadena, USA

ACAT 2017
21-25 August 2017
University of Washington, Seattle

We introduce a special 'multi-head' neural network configuration to predict anomalies in different sub-detectors of the CMS detector
This configuration **allows to learn to identify source of anomaly without explicit knowing ground truth labels** for possible sources
Being applied to CERN CMS data, this approach proves ability to decompose anomaly by separate sub-detectors.

Introduction

Daily operation of a large-scale experiment is a resource consuming task, particularly from the perspectives of routine data quality monitoring. Typically, data comes from different sub-detectors or other subsystems, and the global data quality depends on the performance of each such channel. In this work, we consider the **problem of prediction which sub-detector has caused anomalies in the detector behaviour.** This approach uses only aggregated global quality tag used for training, but allows predicting anomalies for separate sub-detectors.

Data and feature extraction

CERN CMS open data collected in 2010 is used.
LumiSections (minimal chunk of data defined in metadata) are labelled as "good" or "bad".
Collect information for each LumiSections:

- physics data streams: minimal bias, muons or photons;
- channels (reconstructed physics objects): muons, photons, Particle Flow jets or calorimiter jets;
- quantile objects by momentum: 0, 0.2, ... , 1.;
- objects kinematics and spatial location: η, φ, p_T, v_x, v_y, v_z, m;
- characteristics of distributions within lumisection: 5 percentiles, mean and variance.

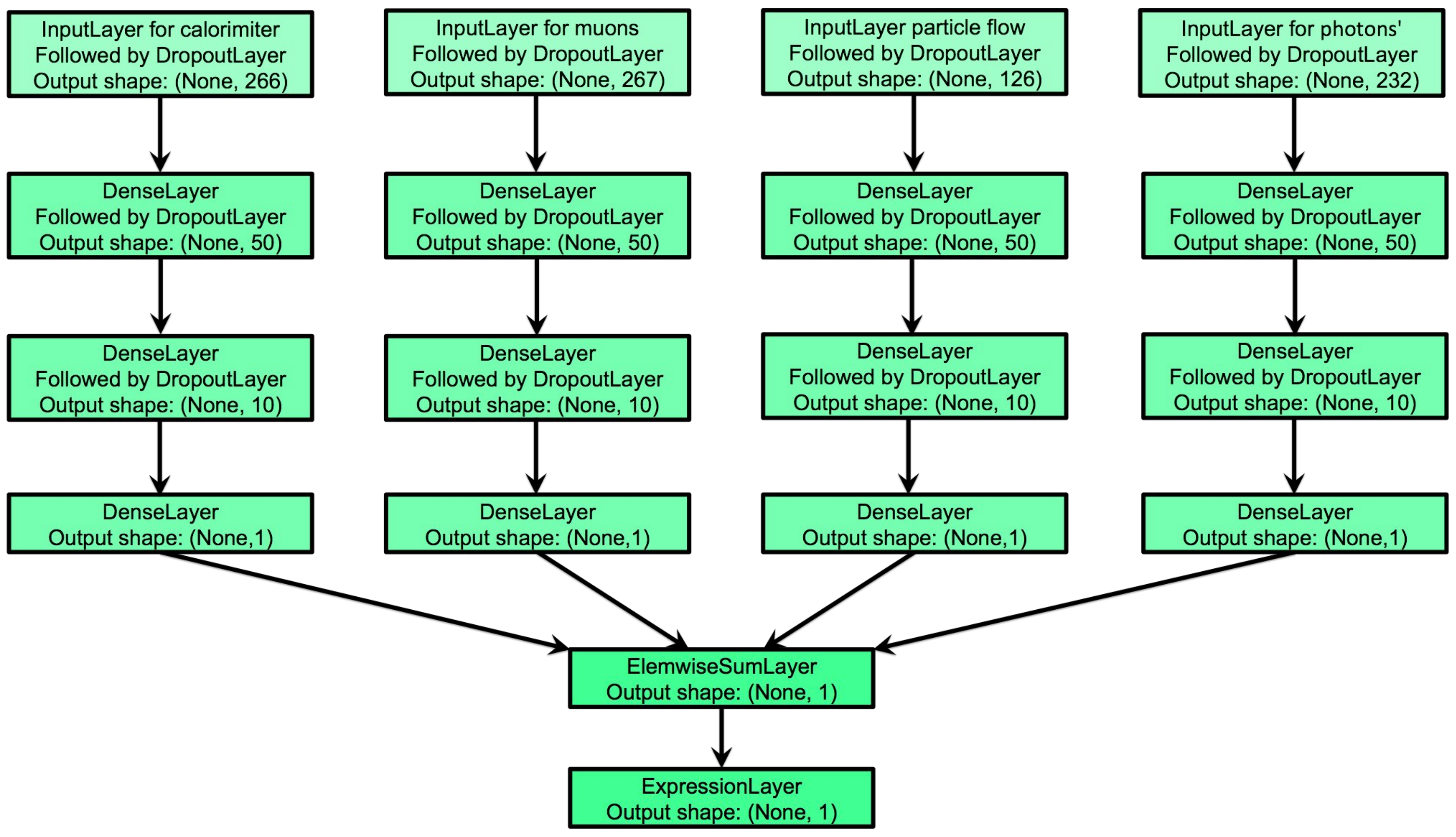
Additionally:

- scalar sum of momentum for all objects in the event;
- instant luminosity;
- number of particles in event.

Use it as features for the following analysis

Method

- build neural sub-networks on features from each channel;
  - each branch returns a score for its channel;
- connect networks by a "Fuzzy AND" operator $exp[\sum_{i=1}^{4}(f_{i\ subnetwork}-1)]$;
- train network to recover global labels;
- consider estimation of score for each channel as qualification for the channel.



InputLayer for calorimiter
Followed by DropoutLayer
Output shape: (None, 266)

DenseLayer
Followed by DropoutLayer
Output shape: (None, 50)

DenseLayer
Followed by DropoutLayer
Output shape: (None, 10)

DenseLayer
Output shape: (None,1)

InputLayer for muons
Followed by DropoutLayer
Output shape: (None, 267)

DenseLayer
Followed by DropoutLayer
Output shape: (None, 50)

DenseLayer
Followed by DropoutLayer
Output shape: (None, 10)

DenseLayer
Output shape: (None,1)

InputLayer particle flow
Followed by DropoutLayer
Output shape: (None, 126)

DenseLayer
Followed by DropoutLayer
Output shape: (None, 50)

DenseLayer
Followed by DropoutLayer
Output shape: (None, 10)

DenseLayer
Output shape: (None,1)

InputLayer for photons'
Followed by DropoutLayer
Output shape: (None, 232)

DenseLayer
Followed by DropoutLayer
Output shape: (None, 50)

DenseLayer
Followed by DropoutLayer
Output shape: (None, 10)

DenseLayer
Output shape: (None,1)

ElemwiseSumLayer
Output shape: (None, 1)

ExpressionLayer
Output shape: (None, 1)

Each subnetwork returns score

- close to 1 for good lumisections
- close to 1 for anomalies "invisible" from subnetwork's channel data
- close to 0 for anomalies "visible" from subnetwork's channel data

Thus NN decomposes anomalies by channels.

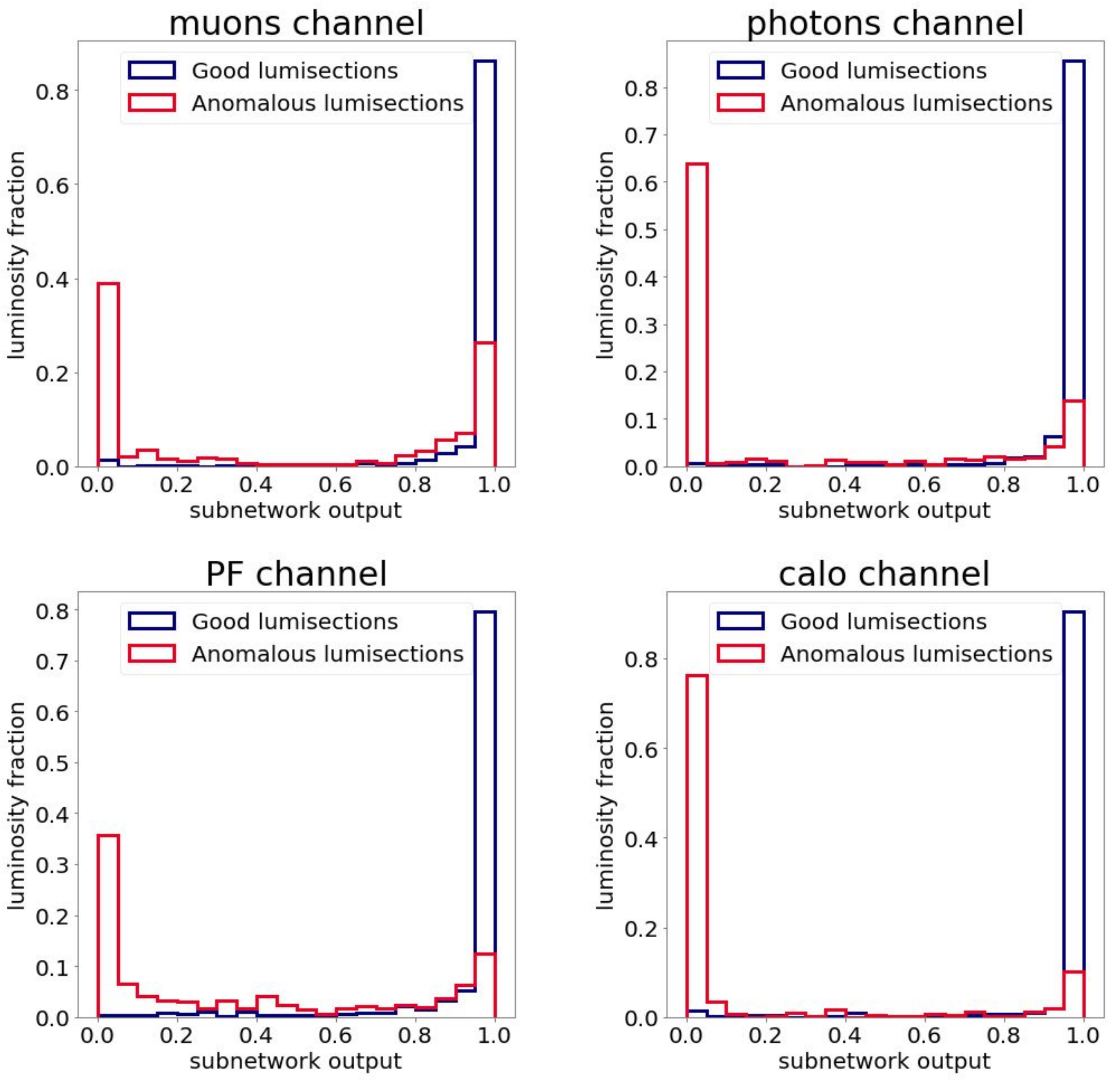
Soft pretraining

"Fuzzy AND" approach assumes that some anomalies can not be seen from all channels. This causes a problem of small gradients for close to the hyperplane samples potentially visible from particular channel, but already with negative labels from other channels. To resolve the problem and to accelerate the convergence we use a dynamic loss function:

$$L' = (1 - C) * L + C * (L_1 + L_2 + L_3 + L_4)/4 \quad ,$$

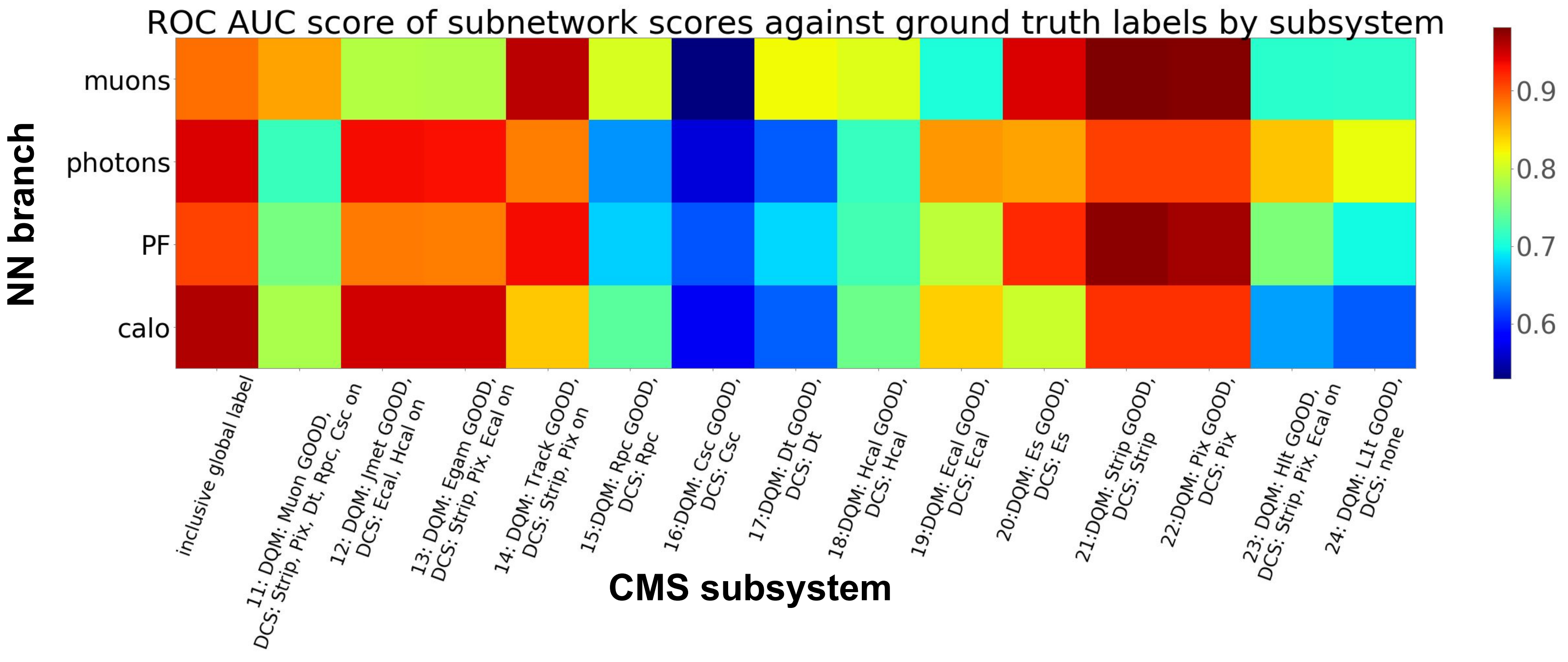
L_i - cross-entropy for "fuzzy AND" output of the network,
L - 'companion' losses, cross-entropy of corresponding subnetwork scores against global labels,
C - decreasing along iterations constant to regulate amount of "pretraining" $(C < 1, C \underset{iterations}{\longrightarrow} 0)$

Decomposition Results


muons channel
photons channel
PF channel
calo channel

Verification

ROC AUC score of subnetwork scores against ground truth labels by subsystem



There is clear correlation between subnetworks' outputs and corresponding subsystem labels, as expected.
NB: **subsystem labels are not used for NN training**