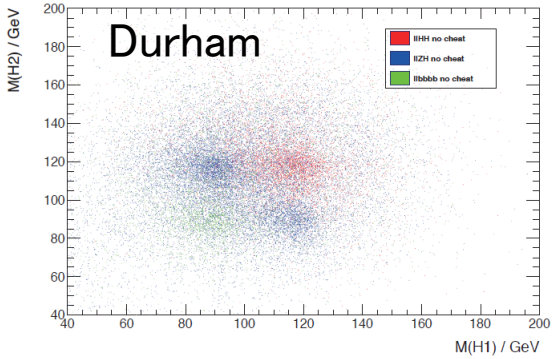
Jet clustering using DeepLearning 2 – Transfer Learning and Deep Neural Network

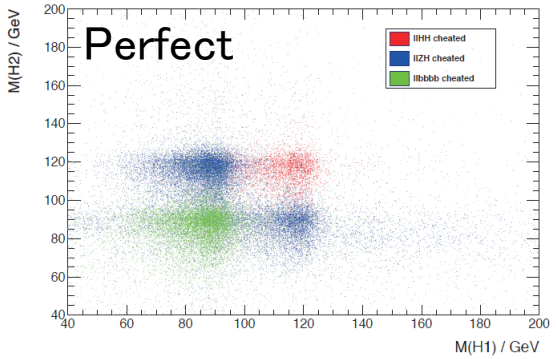
Masakazu Kurata

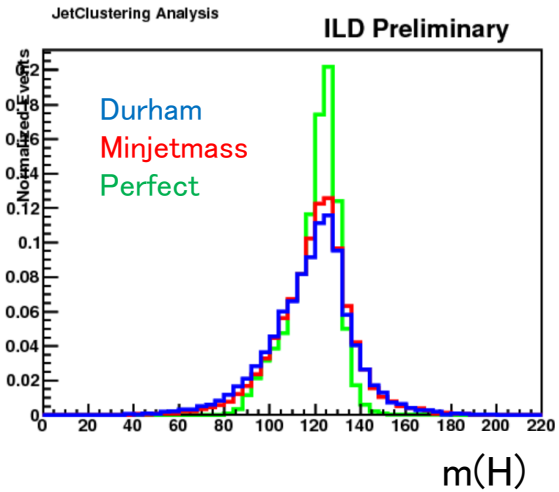
04/15/2017

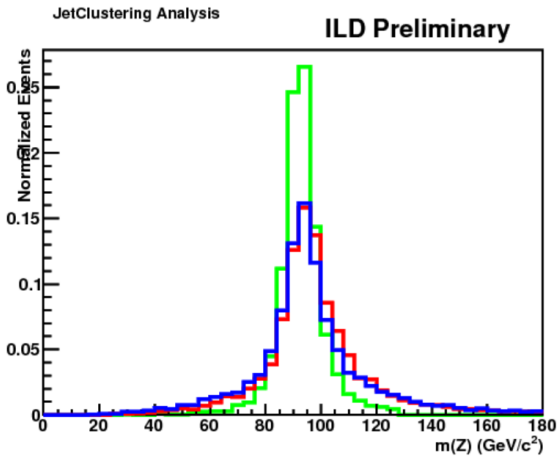
Introduction

- Jet clustering is one of the main key to obtain better physics results
  - To obtain correct jets leads to improve the mass resolution of the resonances
- Present jet clustering is far from good tool for reconstructing jets

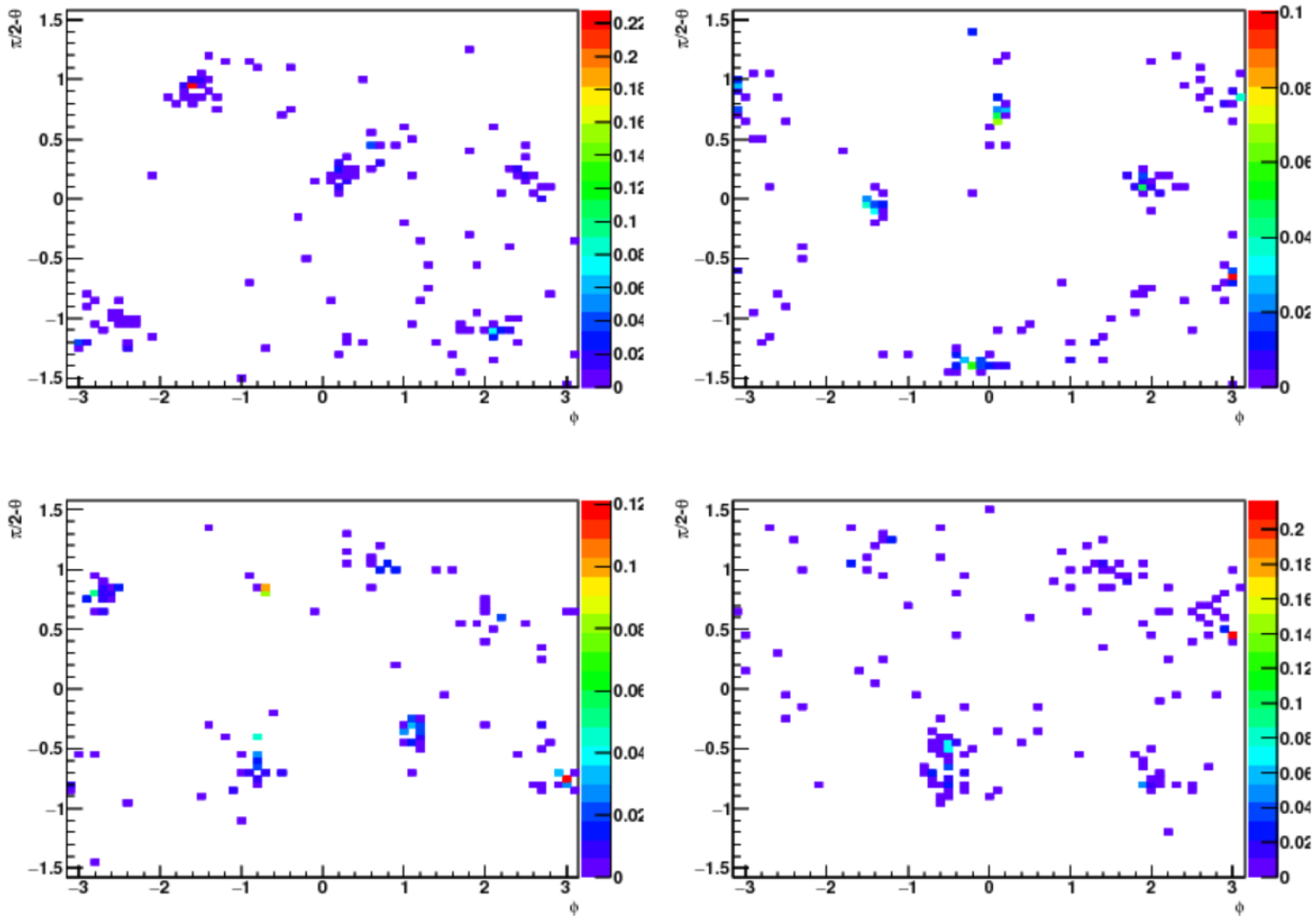








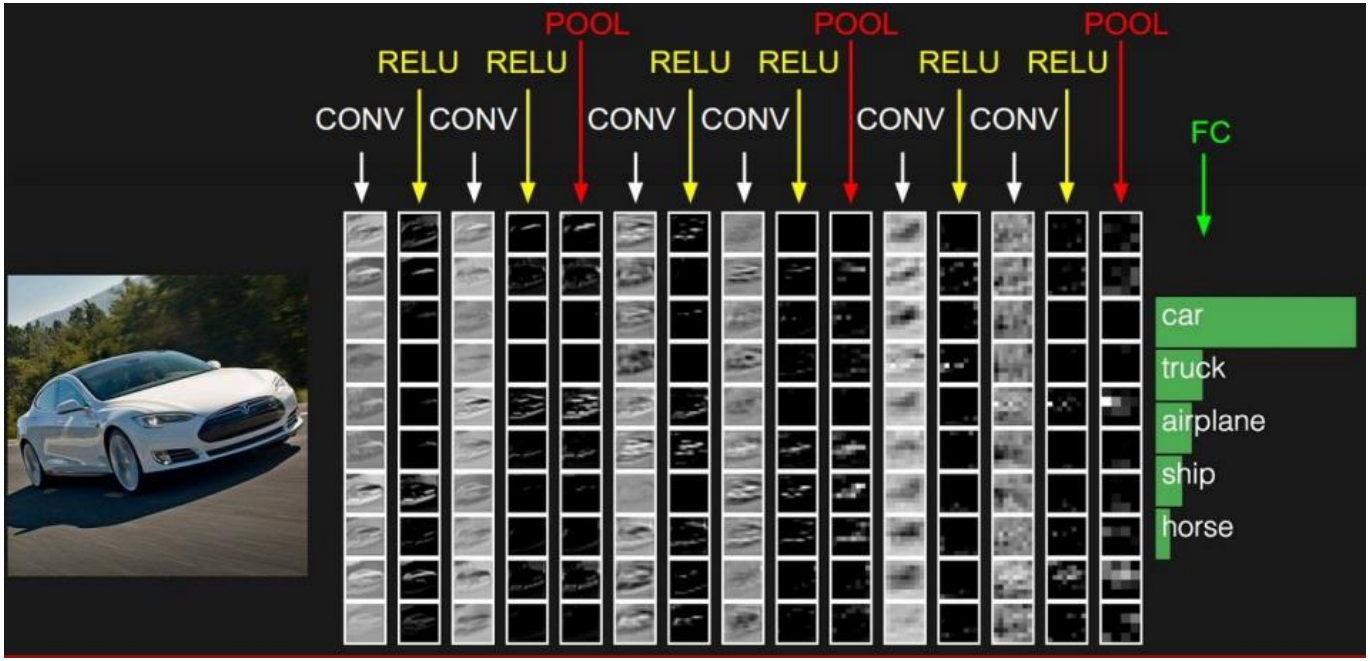
Convolutional neural network?



- First, we would like to get features of jets
- Even by eye, we can see some clusters to some degree
- Can we obtain some hint from those images?
  - Shape? Color flow? Substructure?

How to apply CNN?

- cnn is used for image processing
- cnn has very excellent efficiency for classifying images
- cnn can recognize which object is there in an image
  - Example:



- So, follow this way: How many jets are there in an event?
  - Jets = Objects
  - If cnn can determine how many jets in an event correctly, cnn will be able to recognize jet shapes in an event

Preliminary results

- Tagging efficiency using 1000 events

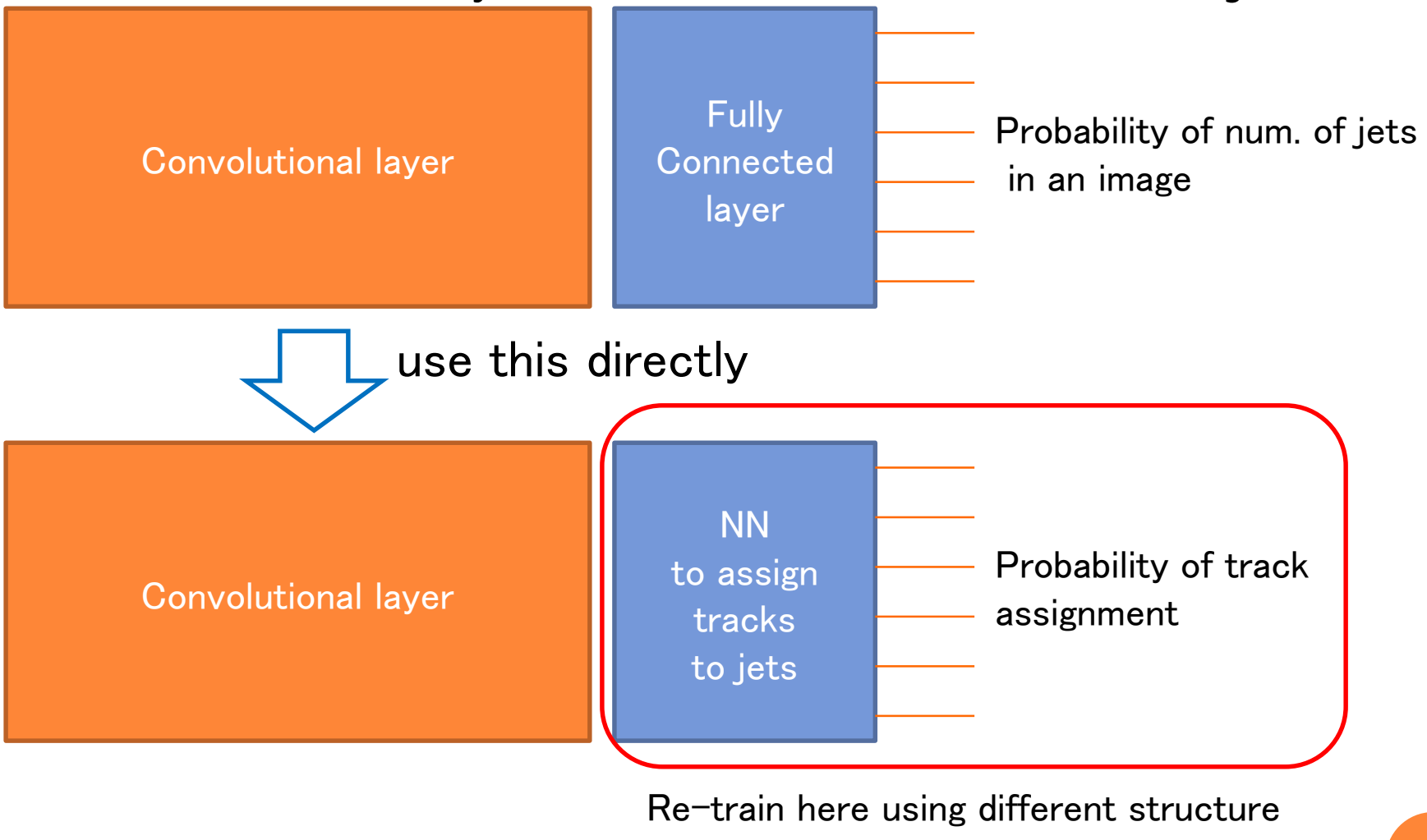
Input/output	1 jet	2 jets	3 jets	4 jets	5 jets	6 jets
1 jet	93.9 93.4	6.10 6.60	0.00 0.00	0.00 0.00	0.00 0.00	0.00 0.00
2 jets	8.80 8.40	79.3 80.1	11.8 11.3	0.10 0.20	0.00 0.00	0.00 0.00
3 jets	0.20 0.30	18.2 16.4	59.6 57.8	21.6 24.9	0.30 0.60	0.10 0.00
4 jets	0.00 0.00	0.70 0.40	16.6 16.3	62.0 61.8	18.5 19.7	2.20 1.80
5 jets	0.00 0.00	0.00 0.00	0.80 0.80	20.4 21.6	47.8 48.6	31.0 29.0
6 jets	0.00 0.00	0.00 0.00	0.00 0.00	1.40 0.80	13.7 14.6	84.9 84.6

Training data
Testing data

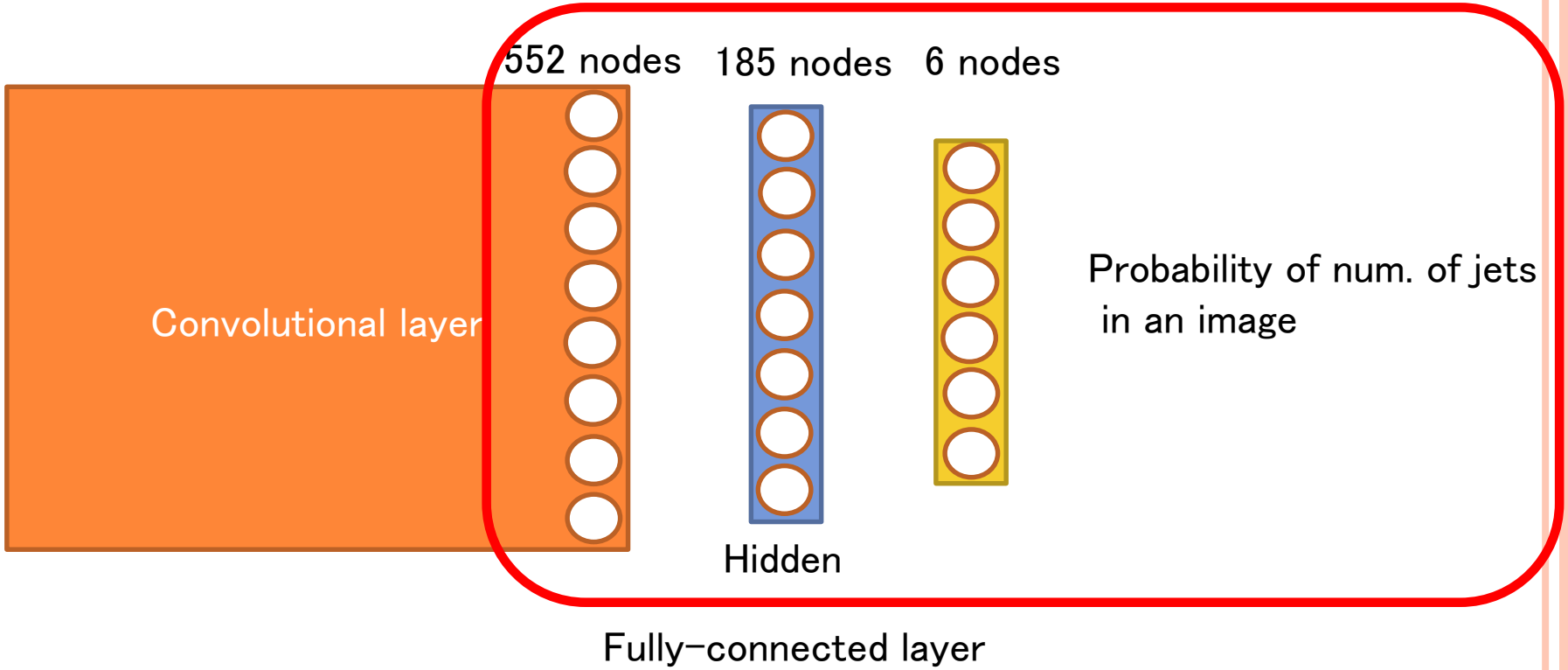
- Looks no over-fitting
- Better diagonal elements will be of course better
- Need to suppress 5jets→6jets mis-ID
- This result cannot be used for jet clustering of course

Transfer learning

- Convolutional layer will learn some features about jets

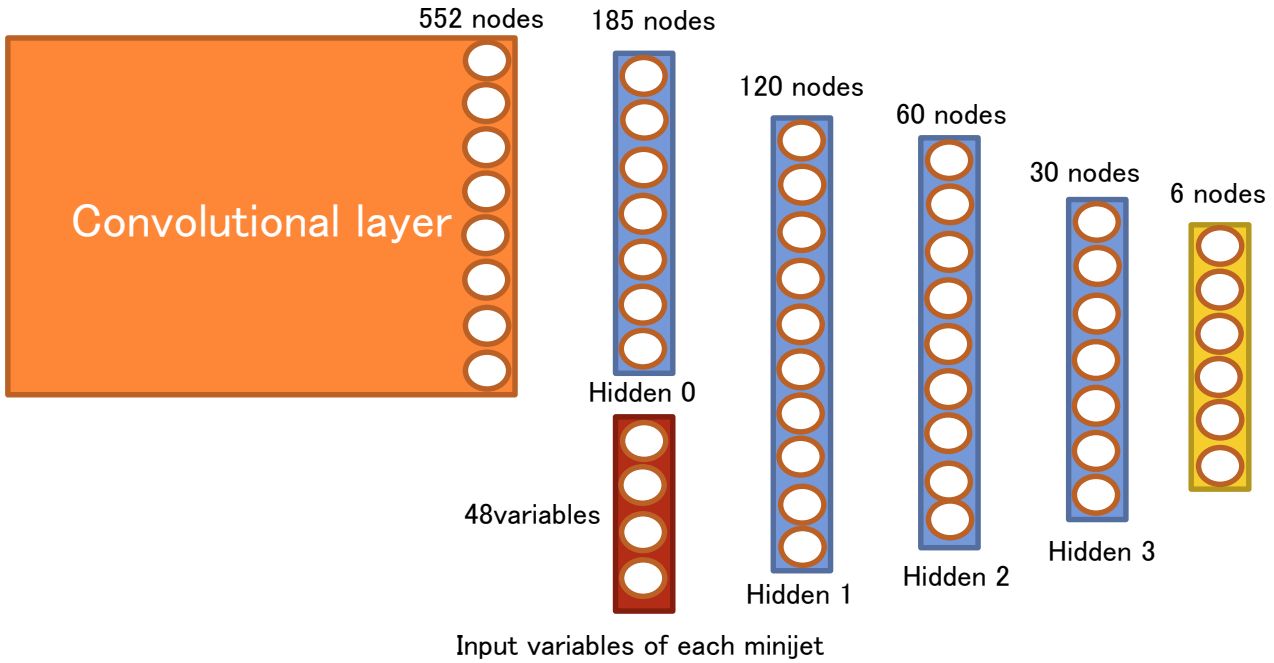


Network Architecture of CNN



- Structure of CNN
- Simple structure of feedforward neural network with 1 hidden layer
- There is no reason for the determination of num. of nodes on each layer

Preliminary structure of the jet clustering neural network

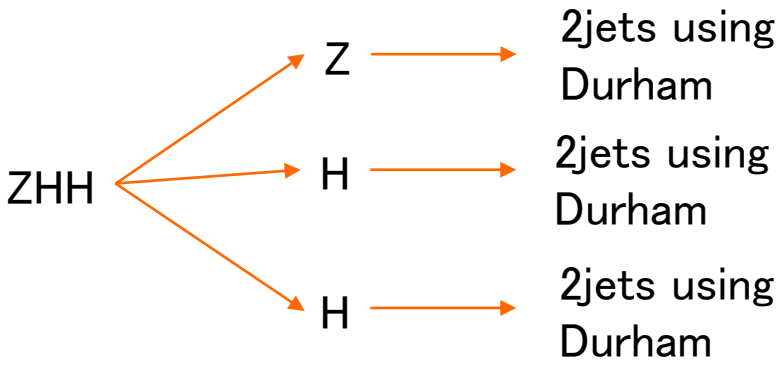


Probability of the jet which a minijet should be assigned to

- Structure of Jet clustering neural network
- CNN + Deep Neural network
  - CNN + feedforward neural network with 3(+1) hidden layers!
  - I have checked 3 hidden layers is better than 2 hidden layers, but not yet checked 1 hidden vs. 3 hidden well
  - NN doesn't work for more than 3 hidden layers so far…
- CNN will provide information of each event
- There is no reason for the determination of num. of nodes on each layer

NOTATION

- Create “answer” jets: perfect Durham jet clustering



- Numbering jets
  - Simply, energy ordering of the jets
  - So far, there seems no dependence of the jet direction!

Network training

- Start from 30 minijets in an event
  - To save CPU time for training
  - Durham jet clustering is used to reduce num. of minijets to 30
- ZHH→(qq)(bb)(bb) events: 6 jet assignment
- ～1500 events are used for training
  - Num. of epoch is O(1000)
- Check minijet assignment efficiency between Durham and NN

Preliminary results

- Using ～1500 ZHH→(qq)(bb)(bb) events
- Check assignment efficiency using training events
- How many tracks can be assigned correctly

	jet1	jet2	jet3	jet4	jet5	jet6
NN	6242	5819	5383	5191	4899	4191
Durham	6185	5778	5383	5149	4825	3954
Total tracks	7387	6920	6288	5891	5423	4571

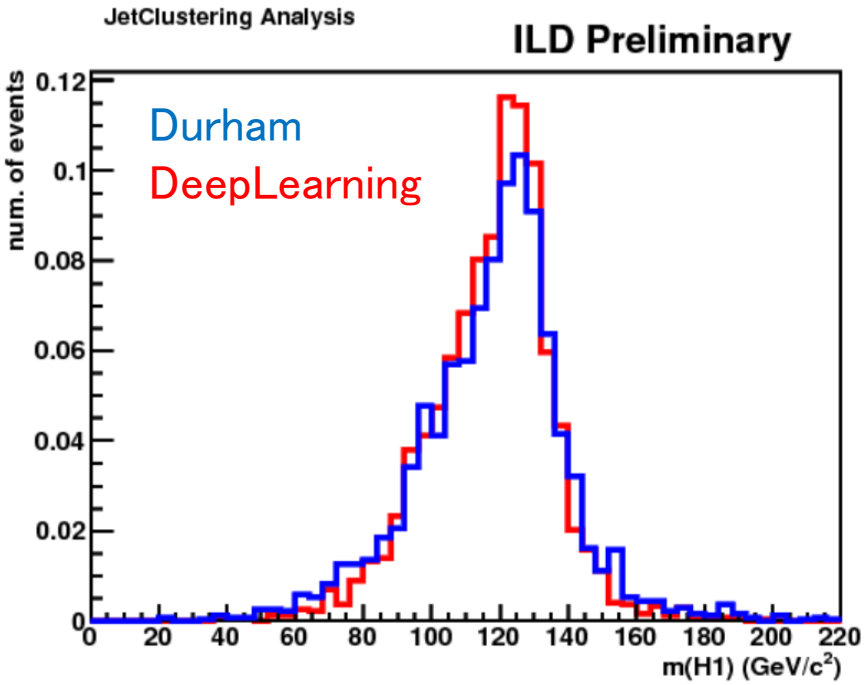
- Energy fraction of main color singlet state
  - Mean over 1500 events

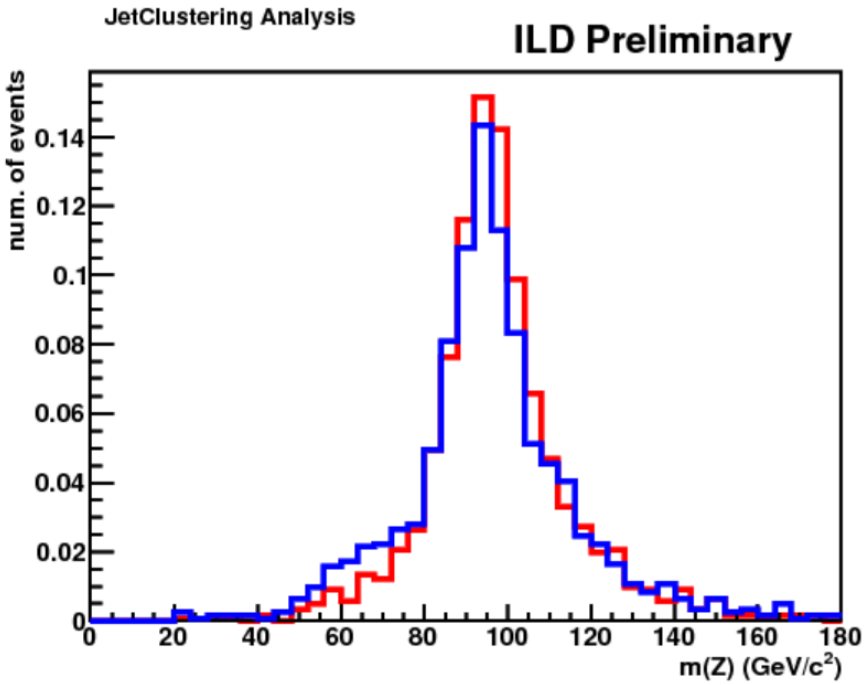
%	jet1	jet2	jet3	jet4	jet5	jet6
NN	95.8	95.0	95.1	94.7	95.1	93.7
Durham	94.5	93.5	94.1	93.0	92.9	88.0

- Efficiency goes up to ～95% level

Mass distribution

- Assume color information is known





- Start to go good!!
  - But not enough… need >95% efficiency…
  - High mass region has room to correct?

Over-Fitting problem

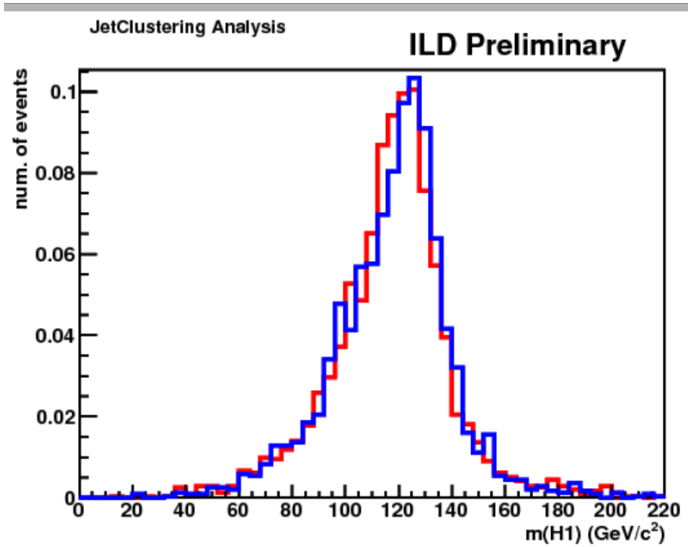
- Biggest problem is over-fitting!
  - Cannot apply to test samples

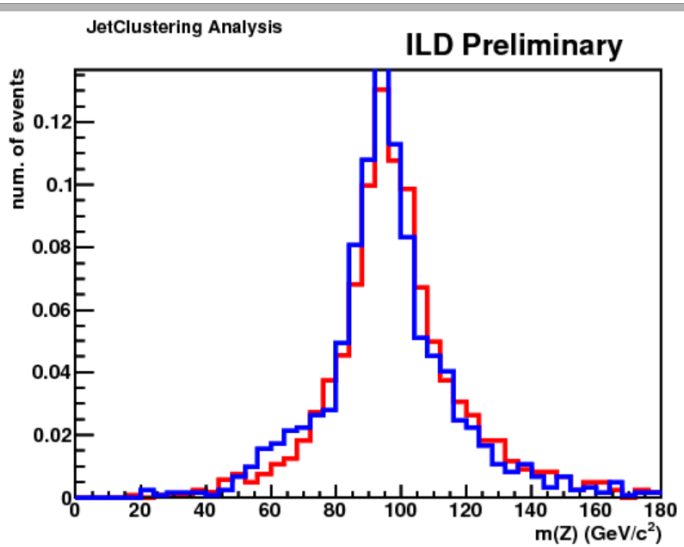
Efficiency & Mass distribution of test events

- Energy fraction of main color singlet state
  - Mean over 1500 events

%	jet1	jet2	jet3	jet4	jet5	jet6
NN	93.8	91.4	92.2	92.9	91.9	85.0
Durham	94.3	93.0	93.3	94.1	92.9	85.9

- Mass distribution
  - Assume color information is known





- Still over-fitting…

PROSPECTS

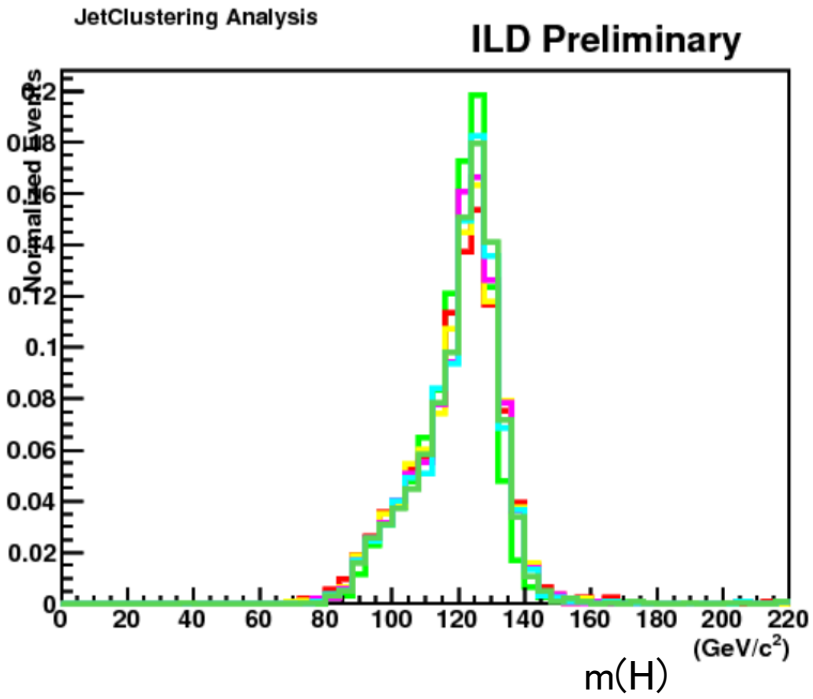
- Over-fitting problem: will reduce when more events are used for training
  - But, I don't know how many events are enough…
    - 5K? 10K? 50K? 100K? Or more?
    - Already getting hard to train… GPU or very big cluster computing is necessary
  - Though, now have reached very stable network training!
    - There were so many efforts…

- Include color information for network training: already introduce one idea
  - Loosen the condition(loss function): $y_i + y_j = 1.0$: i,j are color singlet partner jets each other
  - But, does not work… →so far, seems NN cannot recognize such effect
  - I think cnn will be able to catch such information
  - Needs new idea

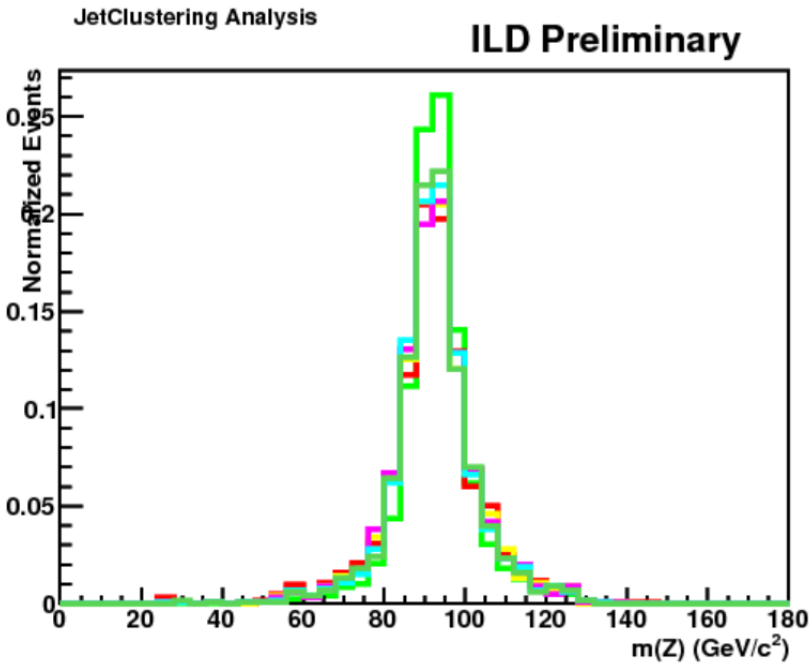
- Improvement:
  - At least, maybe needs >98% minijet assignment efficiency…
    - Very difficult to improve with avoiding over-fitting
    - I think I can make NN with 100% efficiency for training sample
  - Network architecture? Input variables? cnn?
  - Num. of minijets?

BACKUPS

Effect of minijets



PFO
20jets
30jets
40jets
50jets
60jets



	20jets	30jets	40jets	50jets	60jets
Mis-clustered minijet rate(%)	15.2	16.3	17.2	18.0	18.5

KEYWORDS

- Memo of keywords for DeepLearning
  - I feel very important:
    - Kernel Principal Component Analysis(PCA)
    - Batch Normalization
  - For avoiding over-fitting:
    - L2 Normalization
    - Dropout
  - For CNN:
    - Restricted Linear Unit(ReLU)
  - Other:
    - Softmax