Automatic Colorization for Jet Clustering

Masakazu Kurata

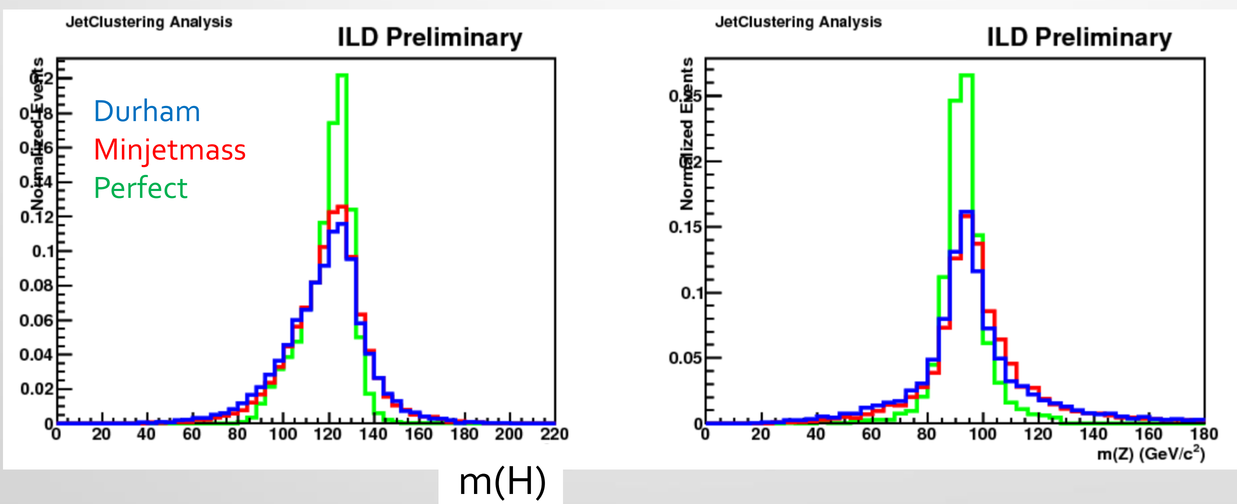
KEK

ALCW2018

05/28/2018-06/01/2018

Introduction

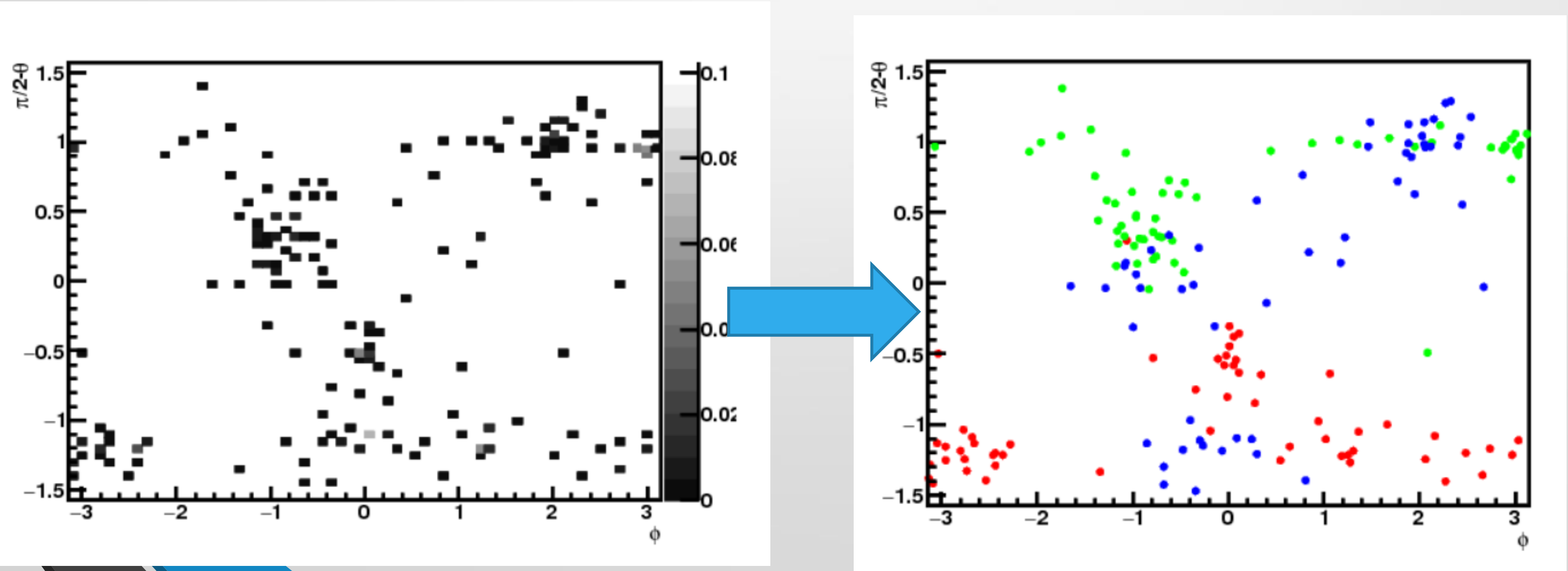
- Jet clustering is one of the main key to obtain better physics results
  - Physics results are strongly limited by mis-clustering
  - To obtain correct jets leads to improve the mass resolution of the resonances
- Present jet clustering is far from good tool for reconstructing jets
  - e.g. Higgs self-coupling@500GeV(ZHH): ～40% improvement if perfect!



- Staging: even at 250GeV, clustering is very important
  - Separation of ZH/ZZ/WW in hadronic events

Make the most of CNN

- Tried supervised learning(Feedforward neural network): LCWS17
- One of the problem is how we can absorb the difference between events
  - Needs very high track(minijet) assignment efficiency to improve mass resolution
  - For very high efficiency, from NN view, all the events look "exception"… → infinite number of nodes & infinite number of events is necessary?
- CNN can relatively absorb position shift & distortion of (jet) shape
  - So, CNN meets this?
- One idea: "Automatic Colorization" using CNN
  - Gray scale → Color

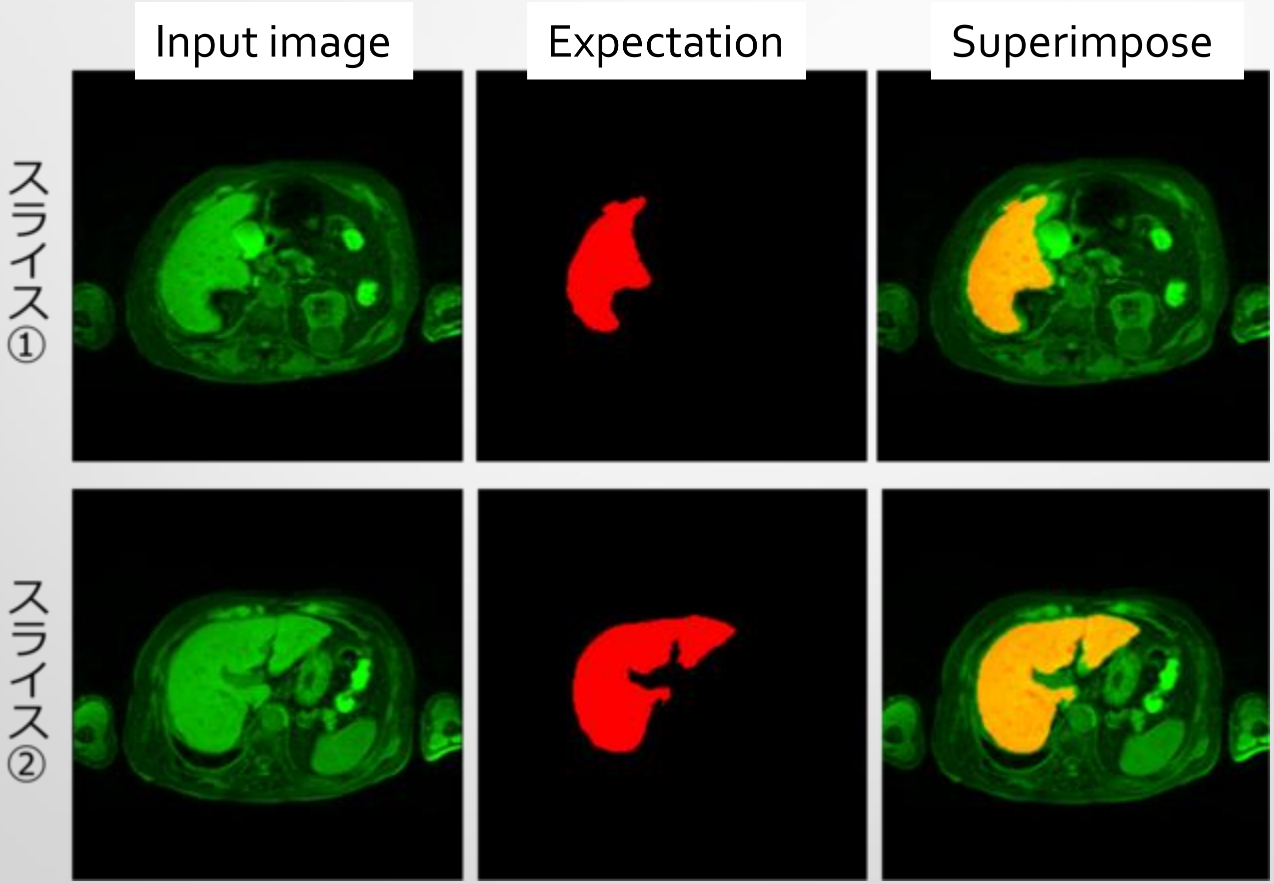


Example

Grey scale	Expectation	Truth

or

- We can estimate the region of what we want to know
- This calls “semantic segmentation”
- Example:



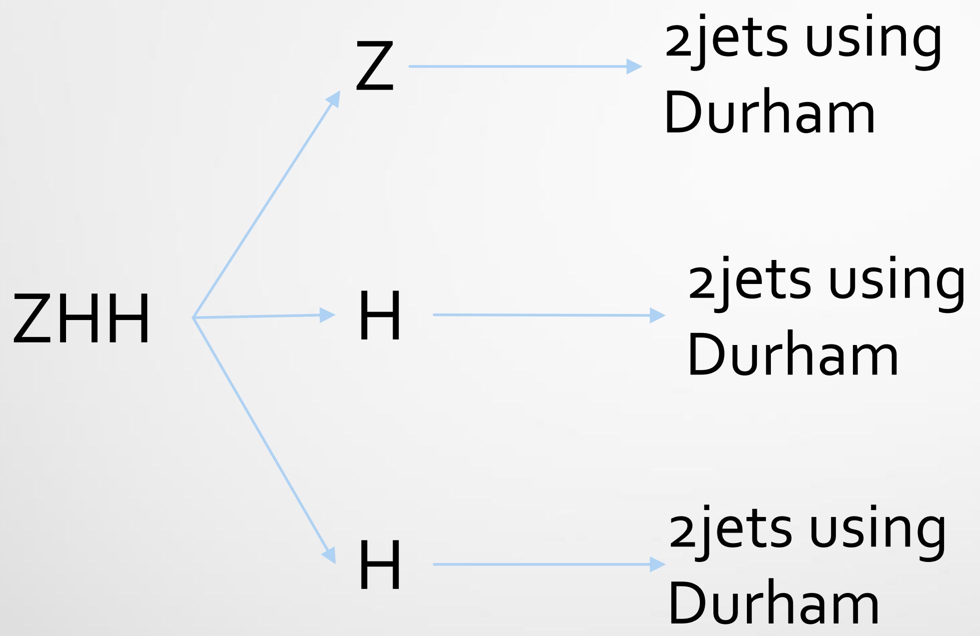
- Can we apply these kinds of techniques for jet clustering?

Use CNN for automatic colorization

- For jet clustering, we need the global and local information for each event
  - Global: Where is the large energy located?
  - Local: Correlation between neighbors or large energy area?
- Using CNN, we will extract both of the features
  - Encorder-Decorder type CNN is used (calls as u-network)
  - Already ～30 layers in CNN!
- Add Conditional Random Fields for improvement
  - 1-2% improvement can be seen in semantic segmentation
  - How about this?

notation

- Supervised learning - Create “answer” jets: perfect Durham jet clustering



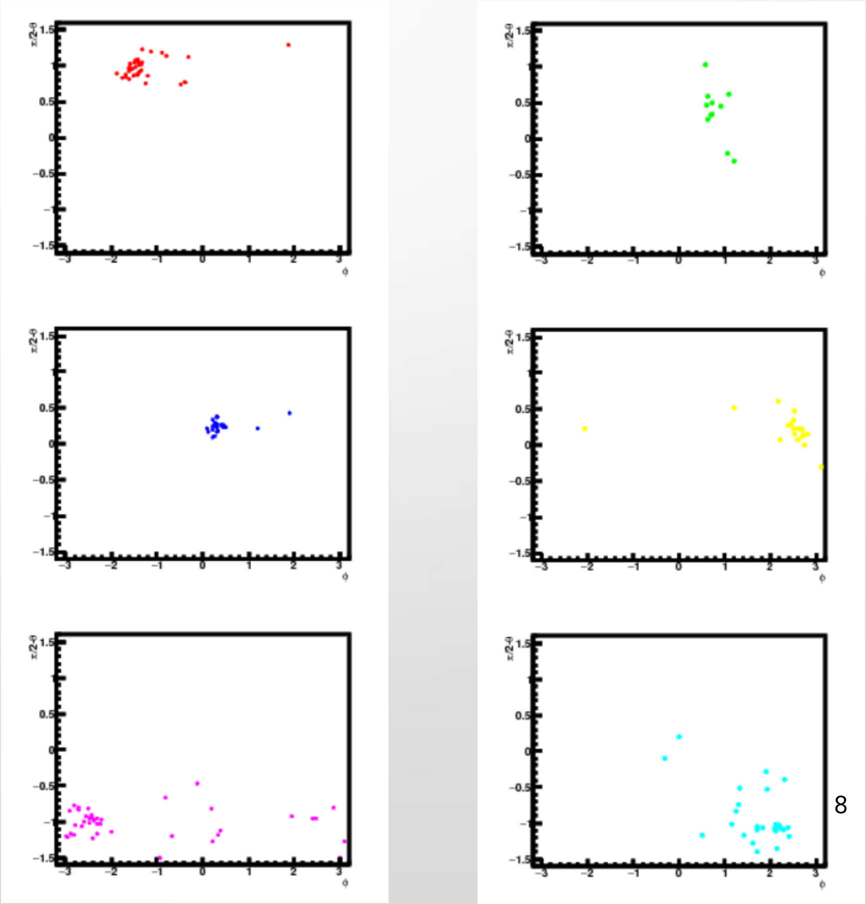
So far, do not consider color singlet state: number of jets is 6

ZHH→(qq)(bb)(bb)→6jets

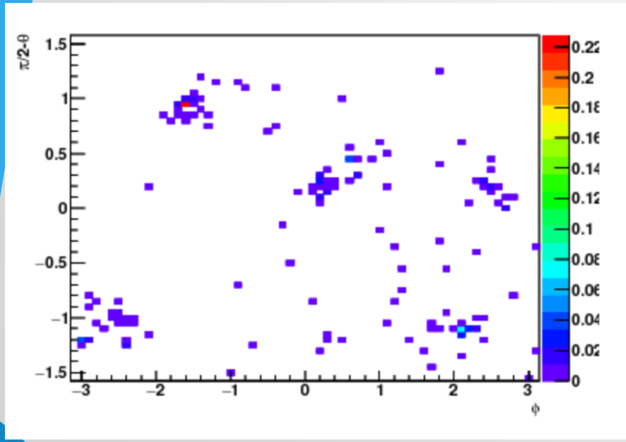
Trial

- Using energy map of each event, estimate color of each track
  - Do not consider color-singlet state

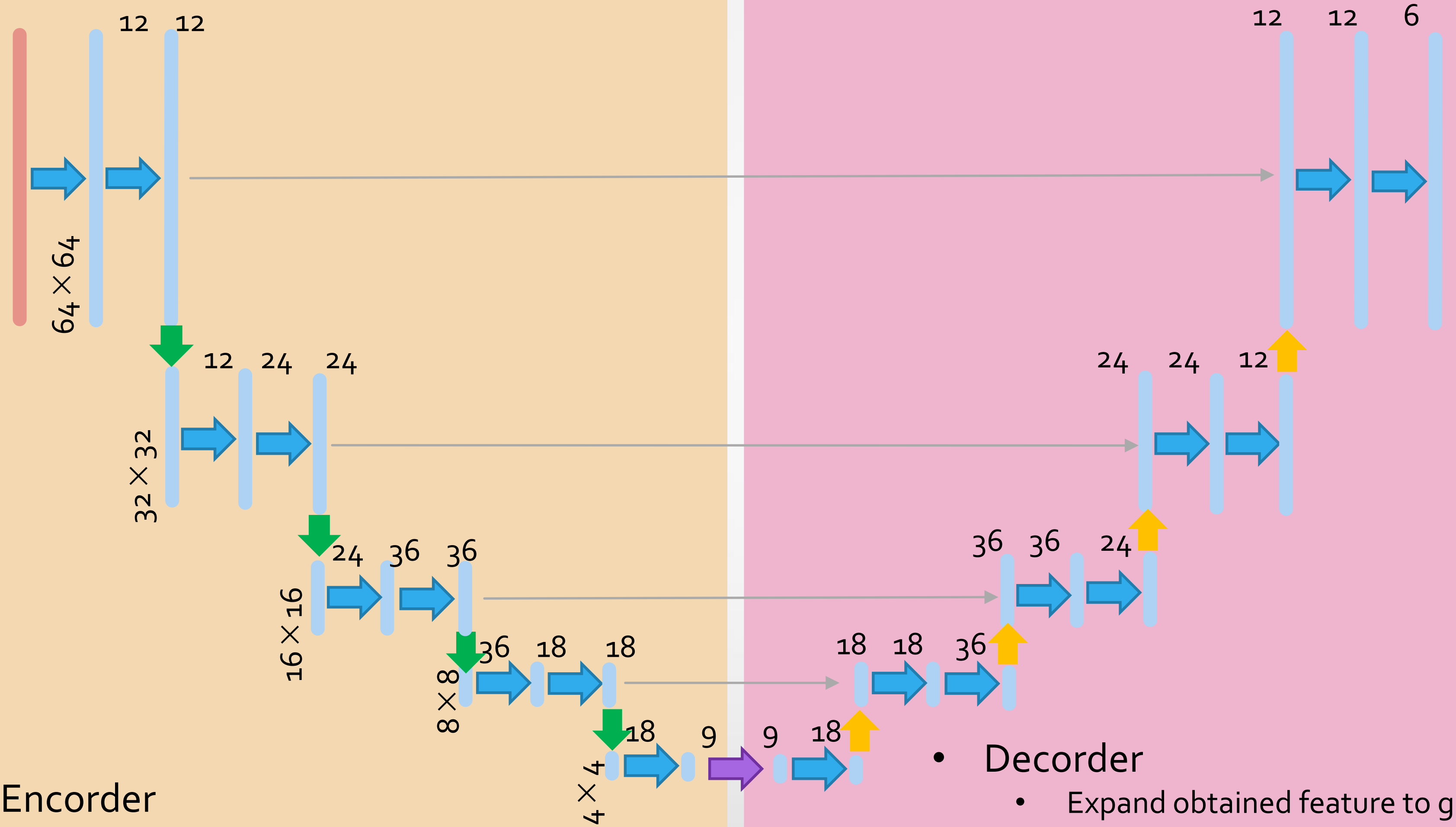
Output(64×64pixel figure)



Input(64×64pixel figure)



U-network

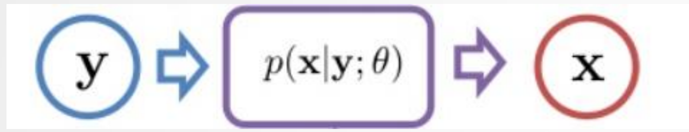


- Encorder
  - Extract local & higher order feature
  - Downsample to make network robust for distortion & shift effect
  - Lost position information

- Decorder
  - Expand obtained feature to global
  - Upsample to recover position information
  - Merge encoder nodes to get precise position information

Conditional Random Fields

- Inference of a latent variable from measured variables
  - x: measured variables → energy, momentum, charge, etc.
  - y: latent variable → the jet which a particle is coming from



  - We have to estimate good conditional probability distribution
- Use Conditional Random Fields:
  - Estimate parameters to maximize the Boltzmann probability:

$$p(x|y;\theta) = \frac{1}{Z}\exp(-E(x))$$

$$E(\mathbf{x}) = \sum_{(i,j)\in\mathcal{E}} f_{ij}(x_i, x_j) + \sum_{i=1}^{N} f_i(x_i)$$

- Same structure as Ising model (because inspired by Ising model):

$$H(\sigma) = -J \sum_{i=1,\ldots,L} \sigma_i \sigma_{i+1} - h \sum_i \sigma_i$$

- Optimization procedure is very similar to Ising model
  - Mean field approximation, Gibbs sampling, etc.

Preliminary architecture

- CRF is used for post-processing of CNN

$$E(\mathbf{x}) = \sum_{(i,j)\in\mathcal{E}} f_{ij}(x_i, x_j) + \sum_{i=1}^{N} f_i(x_i)$$

Constraint for pairwise tracks

Output of u-network for each track

- In first term, we will be able to impose physics constraints
  - Now, simplest case: impose Durham distance measure:

$$f_{ij} = \omega \cdot \exp(-\alpha \frac{2Min\left(E_i^2, E_j^2\right)}{Evis^2}(1 - cos\theta))$$

  - Based on the fact that jet products will fly colinearly
- We can impose any physics constraint
  - I don't know what is good...
  - Vertex constraint?
  - Other distance measure(anti-kt?)
  - Something else?

Over fit check

- This is still test stage, so cannot check overfitting well
- Just estimate using loss function(small is better):

$$L = -\frac{1}{N}\sum_{jet}\sum_{track}\frac{E_{track}}{E_{jet}}\mathrm{Log}(y_{track})$$

- Cross entropy is weighted by track energy to introduce the importance
  - Energetic tracks should be colored correctly
- If no overfitting, L is almost same between test and train

Without CRF Num. of training events	Loss Train	Loss Test
140	0.185	1.78
11500	0.502	0.646

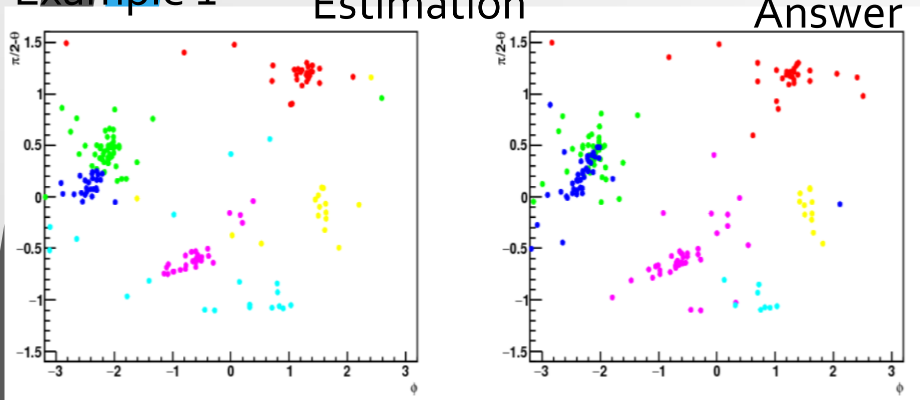
With CRF Num. of training events	Loss Train	Loss Test
900	0.308	1.25
11500	0.485	0.642

- CRF looks better performance?
- Over fitting will vanish if num. of training events is O(10000)
  - Reaches 11500 evts., but still overfitting...
  - Performance gradually degrades, still poor performance
  - So need to optimize the network size to recover the performance

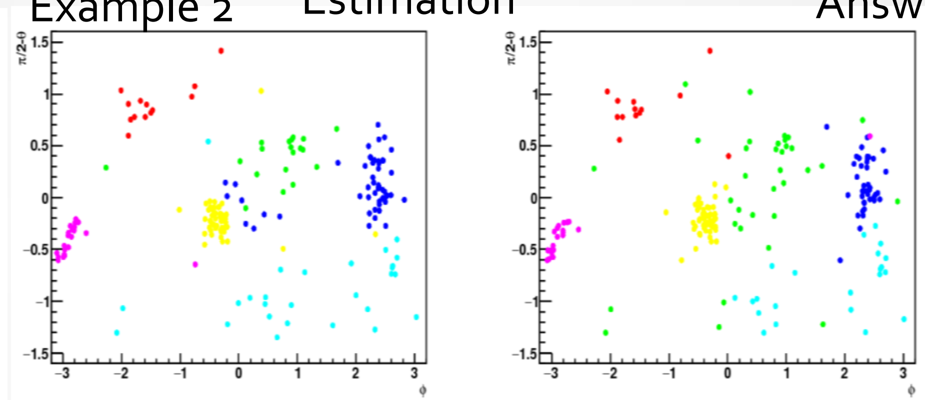
Examples(good)

- Using test samples

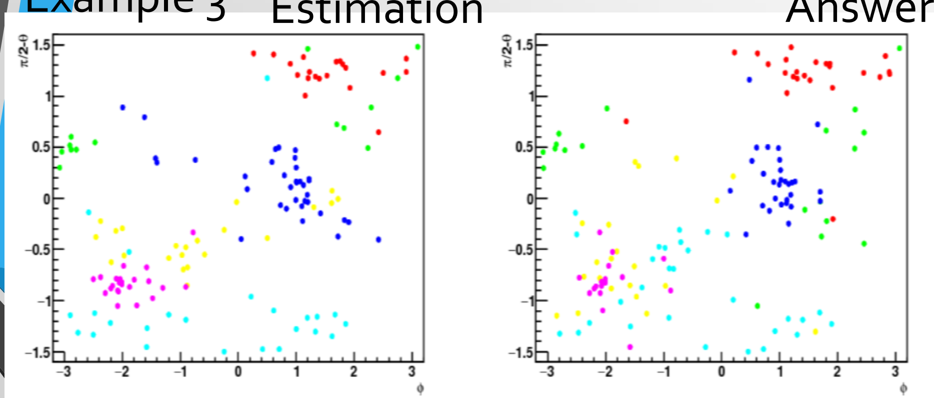
Example 1 Estimation Answer



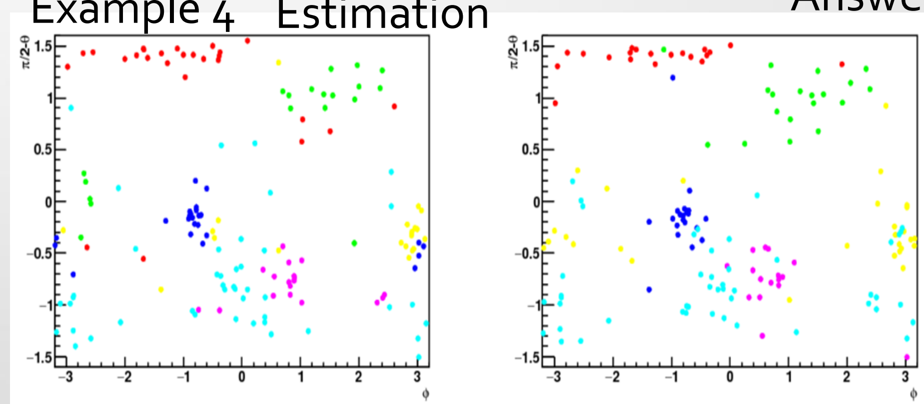
Example 2 Estimation Answer



Example 3 Estimation Answer



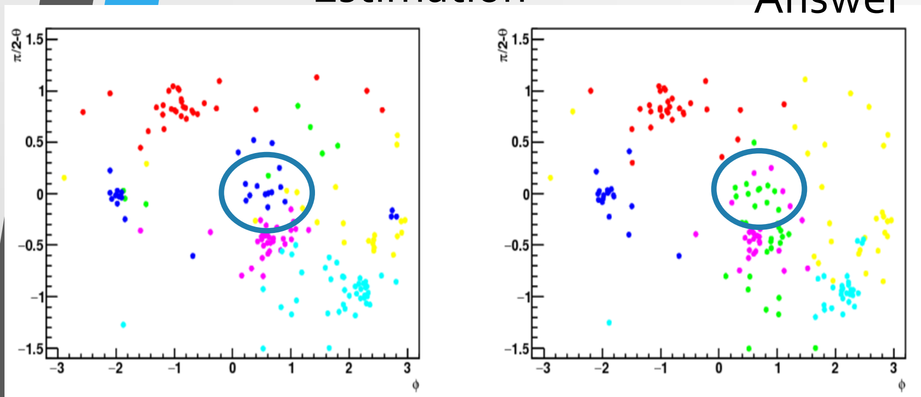
Example 4 Estimation Answer



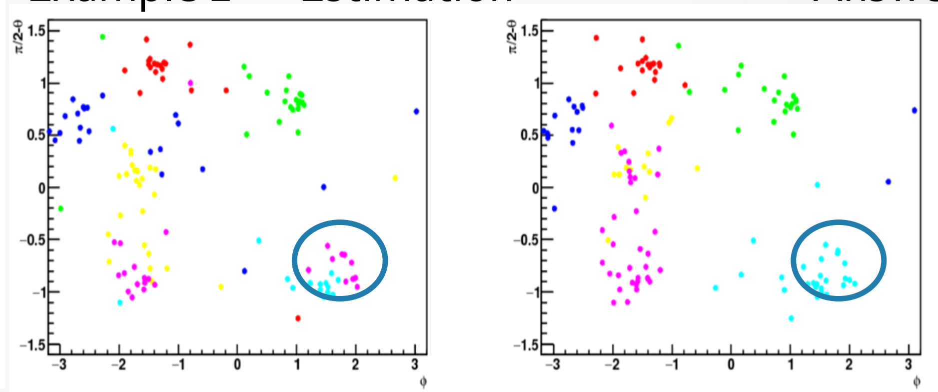
Examples(bad)

- Using test samples

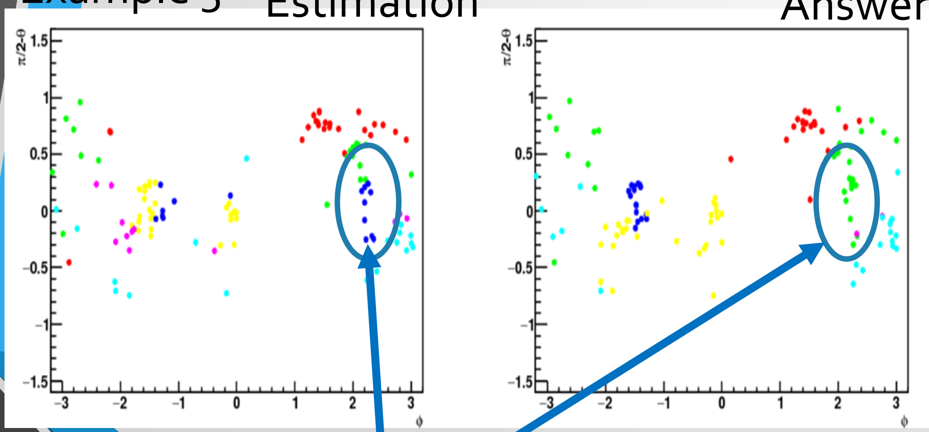
Example 1 Estimation Answer



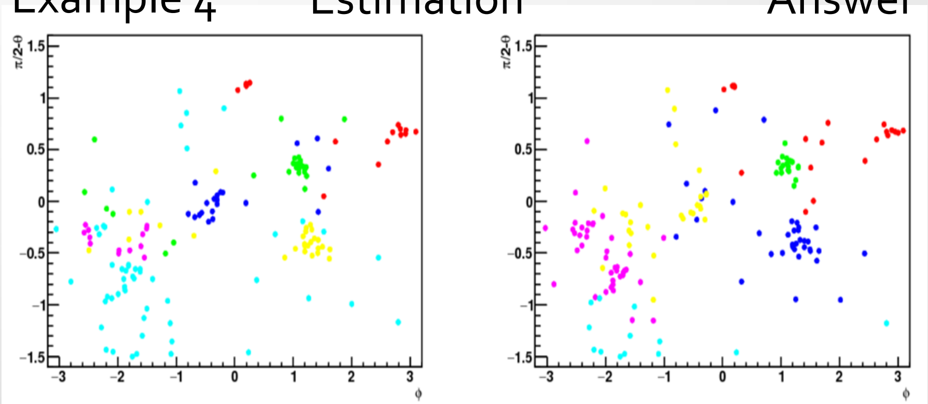
Example 2 Estimation Answer



Example 3 Estimation Answer



Example 4 Estimation Answer



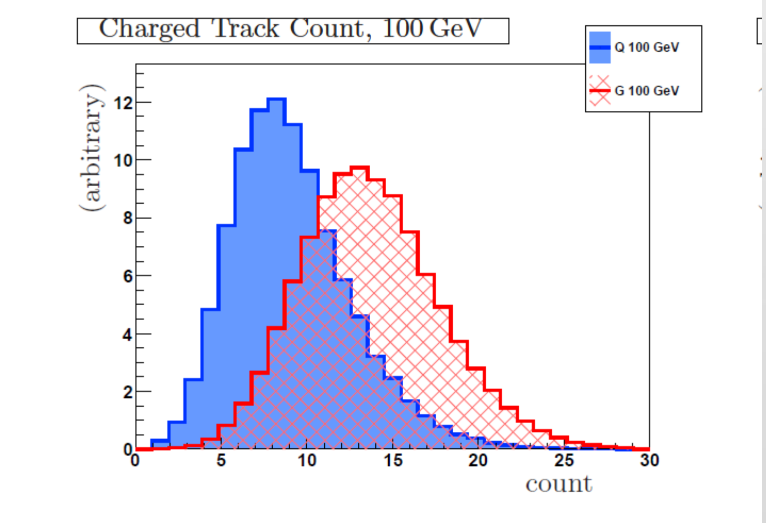
- This is typical mis-colorize
  - Looks insensitive in ϕ axis

Problems & Prospects

- Periodic condition?: regard ϕ direction as continuing infinitely
  - Convolute with periodic condition in all the layers: natural from collider physics point of view

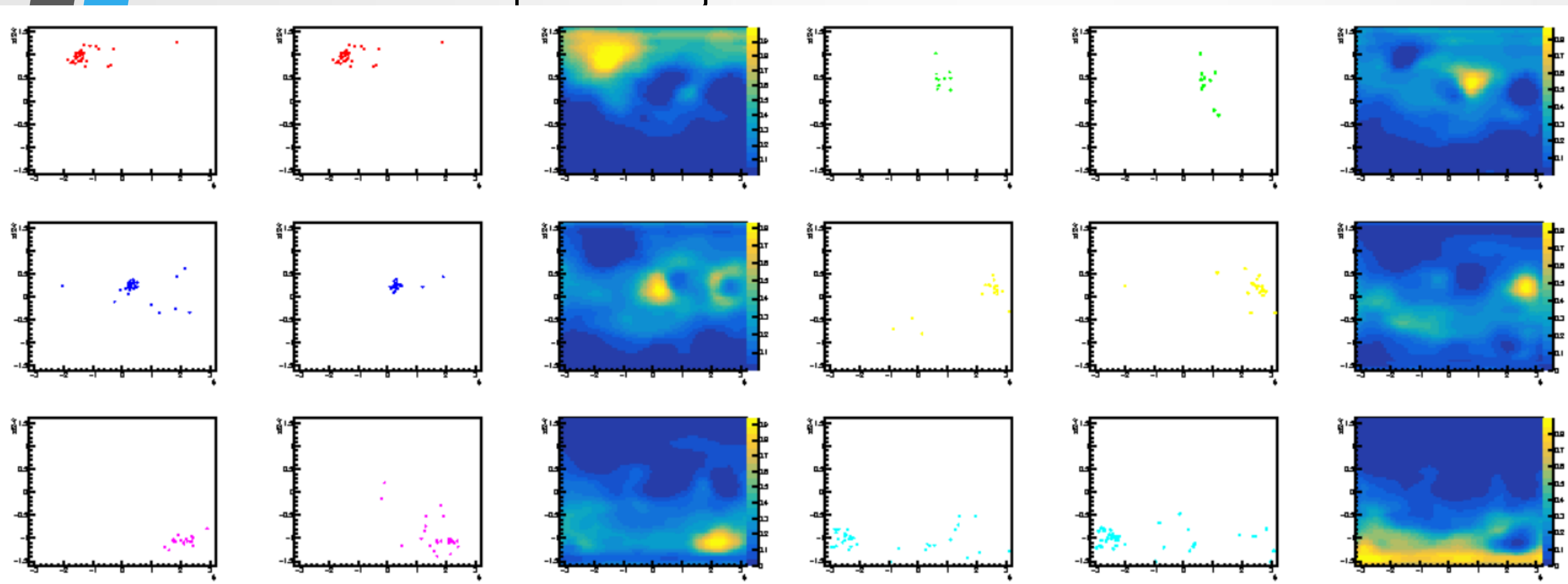
  ⇒ This effect makes ϕ direction insensitive?
  - Now checking the effect of periodic condition
- How about multi-input?
  - Apparently, only energy map is very difficult to make better jet clustering!
  - Charge, PID(=track mass), D0, Z0, etc…
  - e.g.) number of charged tracks in a jet has some power to separate quark/gluon jets
  - Investigate now

arXiv: 1211.7038



Prospects

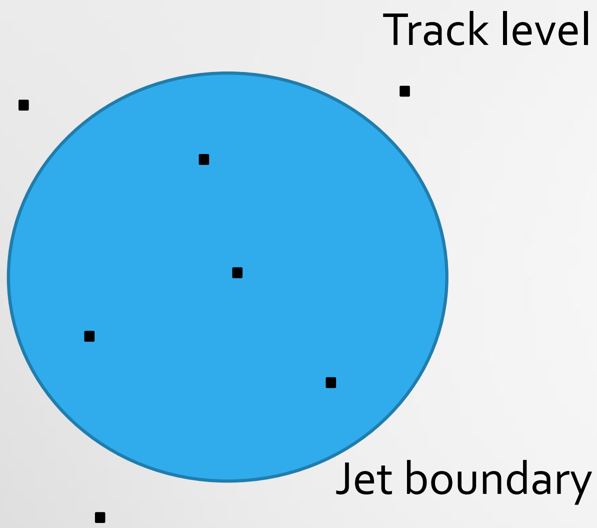
- From semantic segmentation point of view, we can estimate probability of each point
  - Can create heat map for each jet

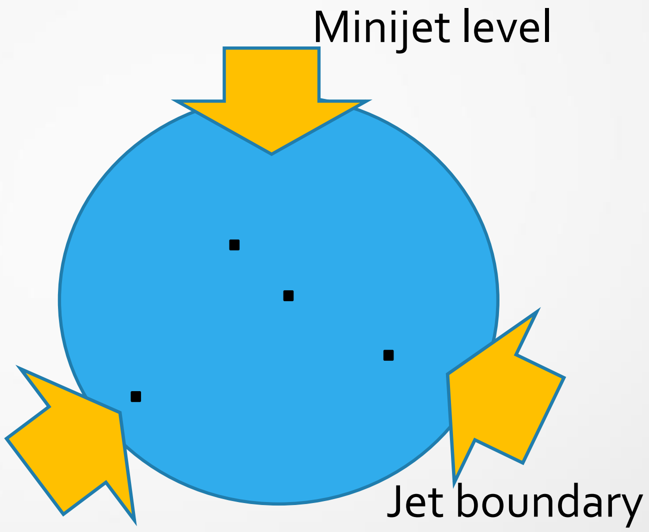


- So we can estimate the probability of minijet using these heat map
  - I think(of course?), some physics effect is necessary to obtain excellent efficiency

Prospects

- It is expected that each minijet move inside the jet boundary which is estimated from CNN





- From Deep learning side, there are some points for improvement
  - Optimize network and hyper parameters
  - There are still many ideas which is worth trying

Summary and Outlook

- Introduce present technique which is developing in machine learning field
  - Try the idea of automatic colorization using Convolutional Neural Network.
  - Introduce Conditional random fields, which was inspired by physics (Ising model), to improve the result
- So far, some events look OK, while some events have bad colorization
  - Bad colorization: Typically, insensitive in ϕ direction- due to considering periodic condition too much? Or due to not learning empty pixels?
  - Just starting point – there is much room to improve
  - Need sophisticate way to combine deep learning and physics naturally
    - Conditional Random Fields is one way to incorporate physics effect
    - Using minijets is another way

backups