Jet Images

Deep Learning Edition

Luke de Oliveira[1], Michael Kagan[2], Lester Mackey[3], Benjamin Nachman[2], Ariel Schwartzman[2]
[1]Stanford University, Institute for Computational and Mathematical Engineering
[2]SLAC National Accelerator Laboratory
[3]Stanford University, Department of Statistics

Stanford Institute for Computational & Mathematical Engineering | ICME

SLAC NATIONAL ACCELERATOR LABORATORY

Overview

- The notion of a *jet-image*
- Image pre-processing
- Unique considerations
- Deep Architectures + performance
- Isolating “the physics” — how can we learn from deep learning?

The *jet-image*

- Powerful, but simple analogy from M. Kagan et al. [1]
- Treat the detector as a camera — calorimeter towers as pixels, and depositions as intensity

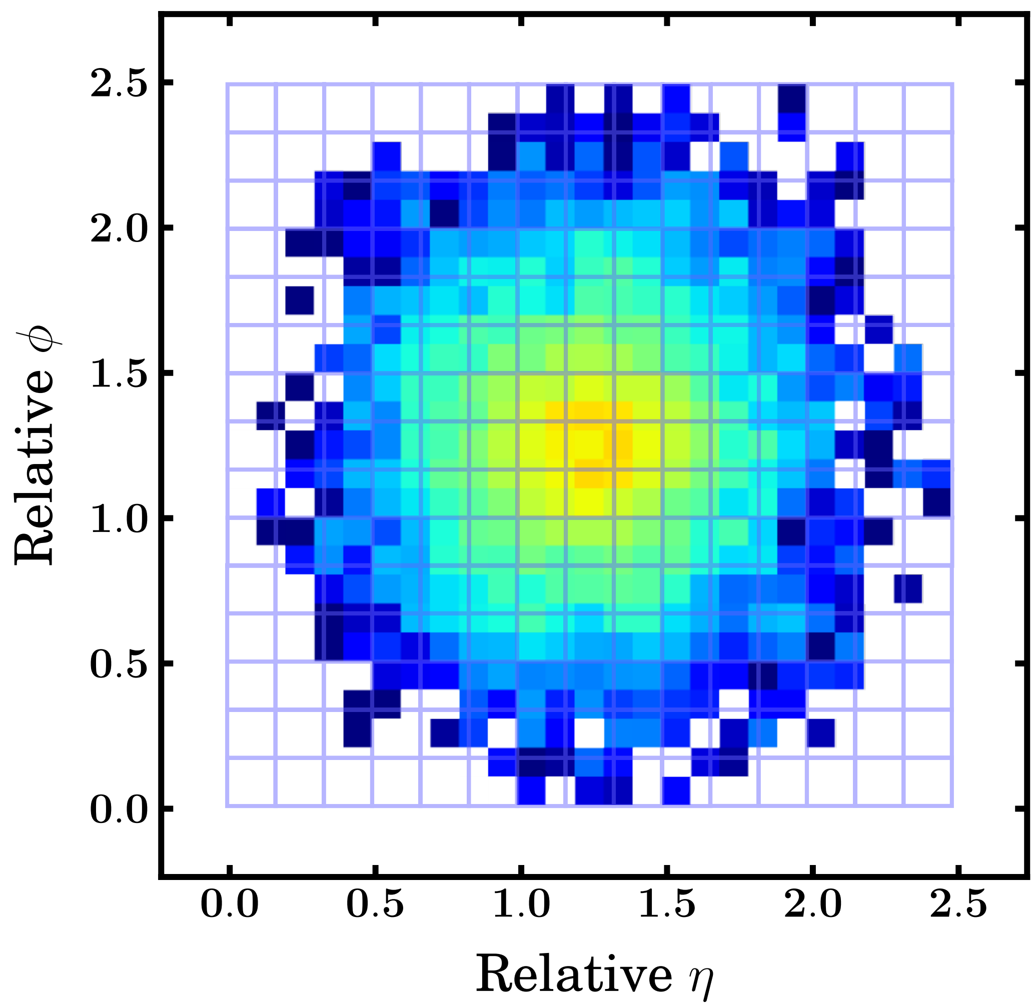


Image details

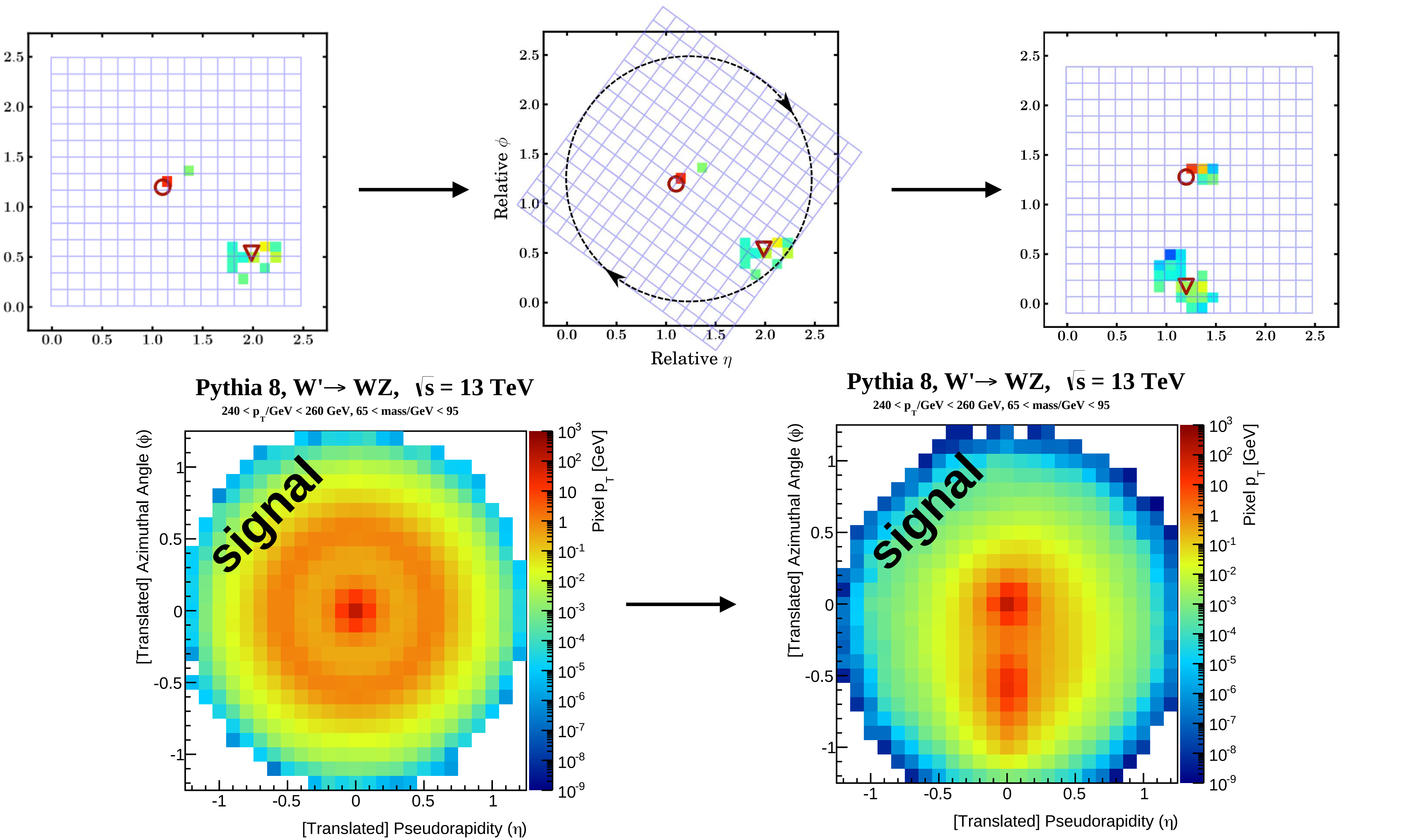
- Proof of concept
- Simulate pixelation conditions of the ATLAS detector
  - (0.1 x 0.1 pixels)
- Normalization / no-normalization
- $W' \rightarrow WZ$, *with* $W \rightarrow qq$, $Z \rightarrow \nu\,\nu$
- Require clustered subjet p_T > 5% of jet p_T

Image Preprocessing

- Account for symmetries of spacetime — see [2] for more details
- Want the image to be centered at $(\eta, \phi) = (0, 0)$
  - ϕ-rotation is around z-axis, so this is fine
  - η-rotation is a Lorentz-boost along the z-axis, so replace E_i with $p_{\mathrm{T},i} = E_i / \cosh(\eta_i)$
- Image is rotated so sub-leading subjet is at $-\pi/2$ — cubic spline interpolation for new grid
- Finally, flip so harder side is consistent

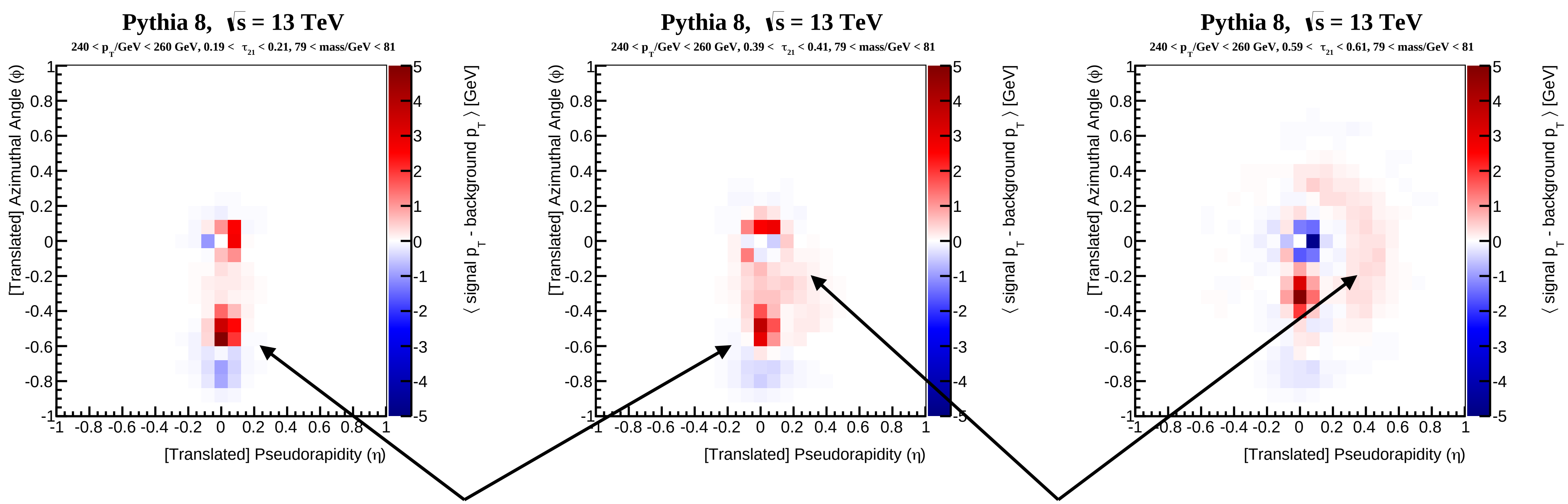
Image Preprocessing

- …in a picture…



Where is the physics?

You can **see** the physics!



The distance
between subjets
is slightly different

gluon jet background
is a color octet,
diffuse radiation

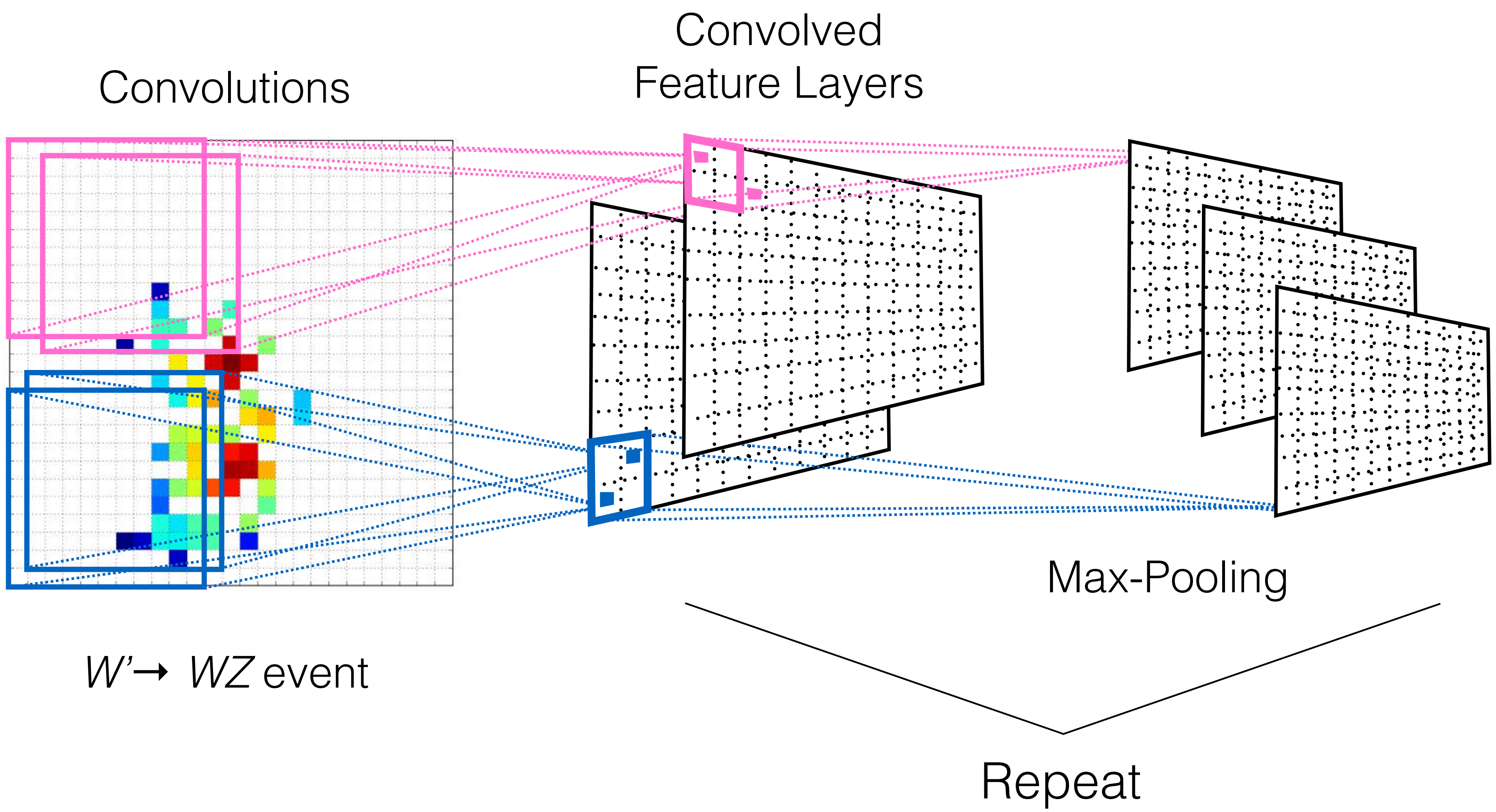
less pronounced second subjet

Unique Deep Learning Considerations

(or, why I find the problem interesting)

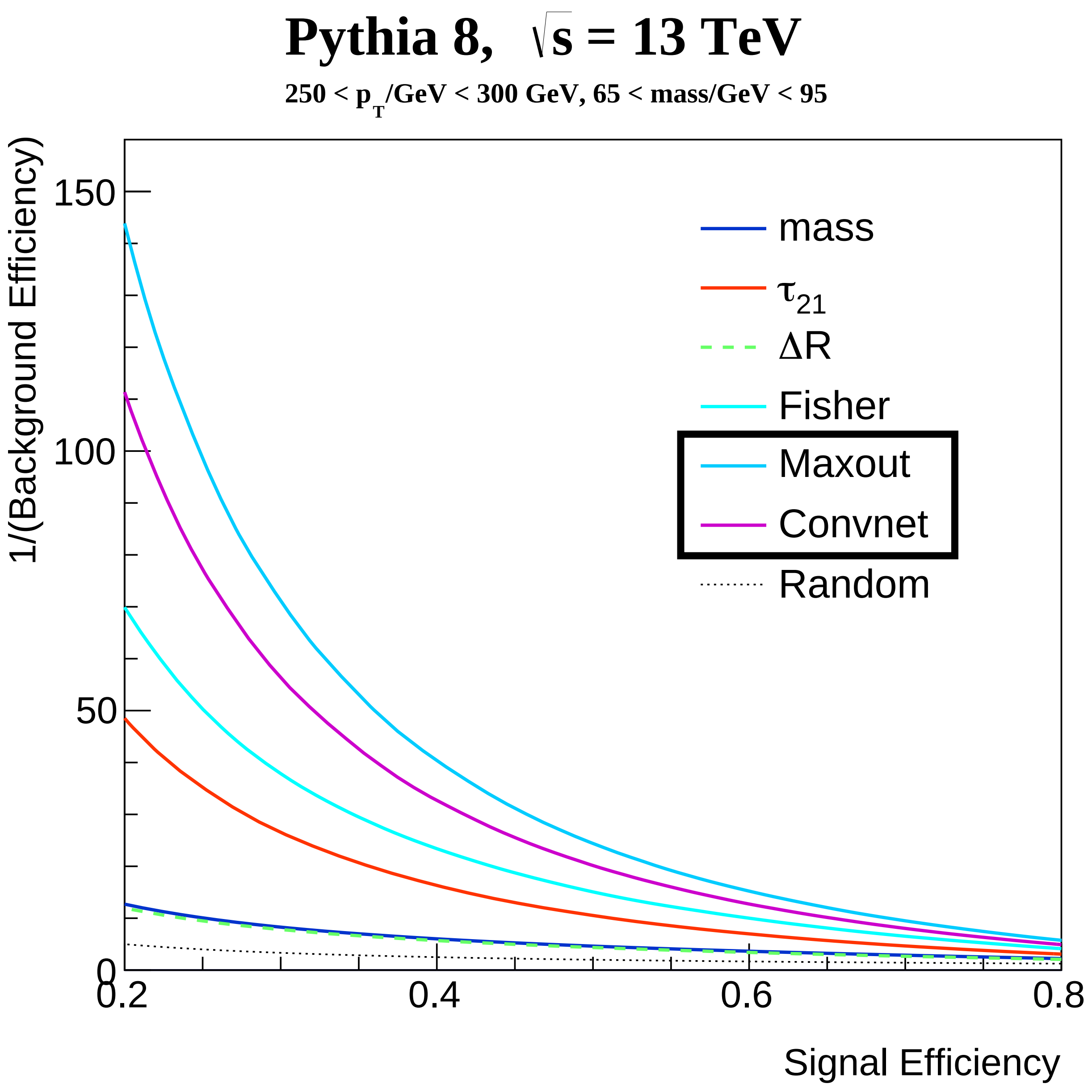
- Sparsity
- Invariance
- Can search for feature representations (rare luxury in ML)
- CNN filters — (11x11)! These are huge! Use FFT-based convolutions [3]
  - Why so big?
- Because of HEP knowledge, can "marginalize" known components

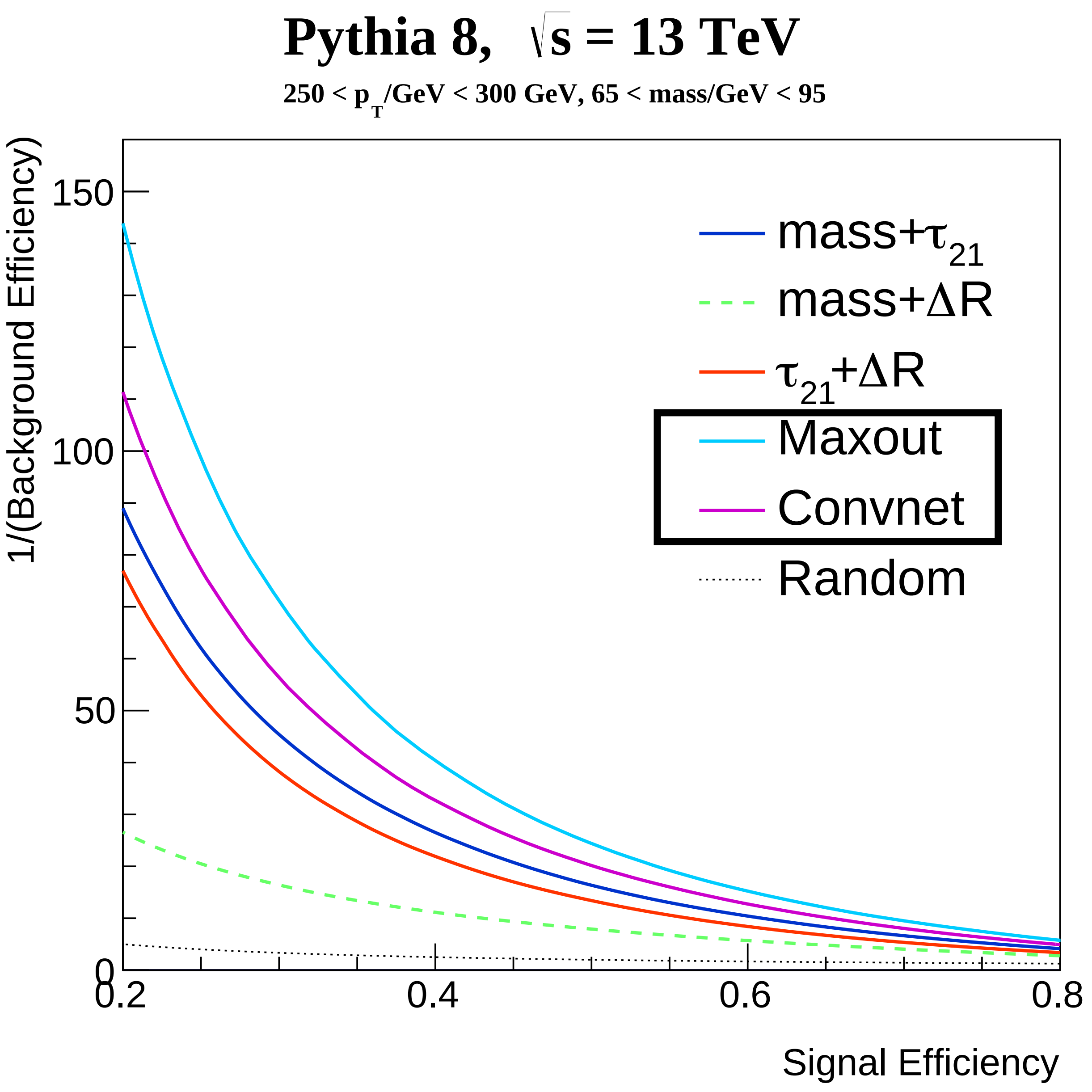
Visualization



$$X * w|_{(k,\ell)} = \sum_{w_{ij} \in w} w_{ij} X_{k+i,\ell+j}$$ …for a 1 channel image

Performance

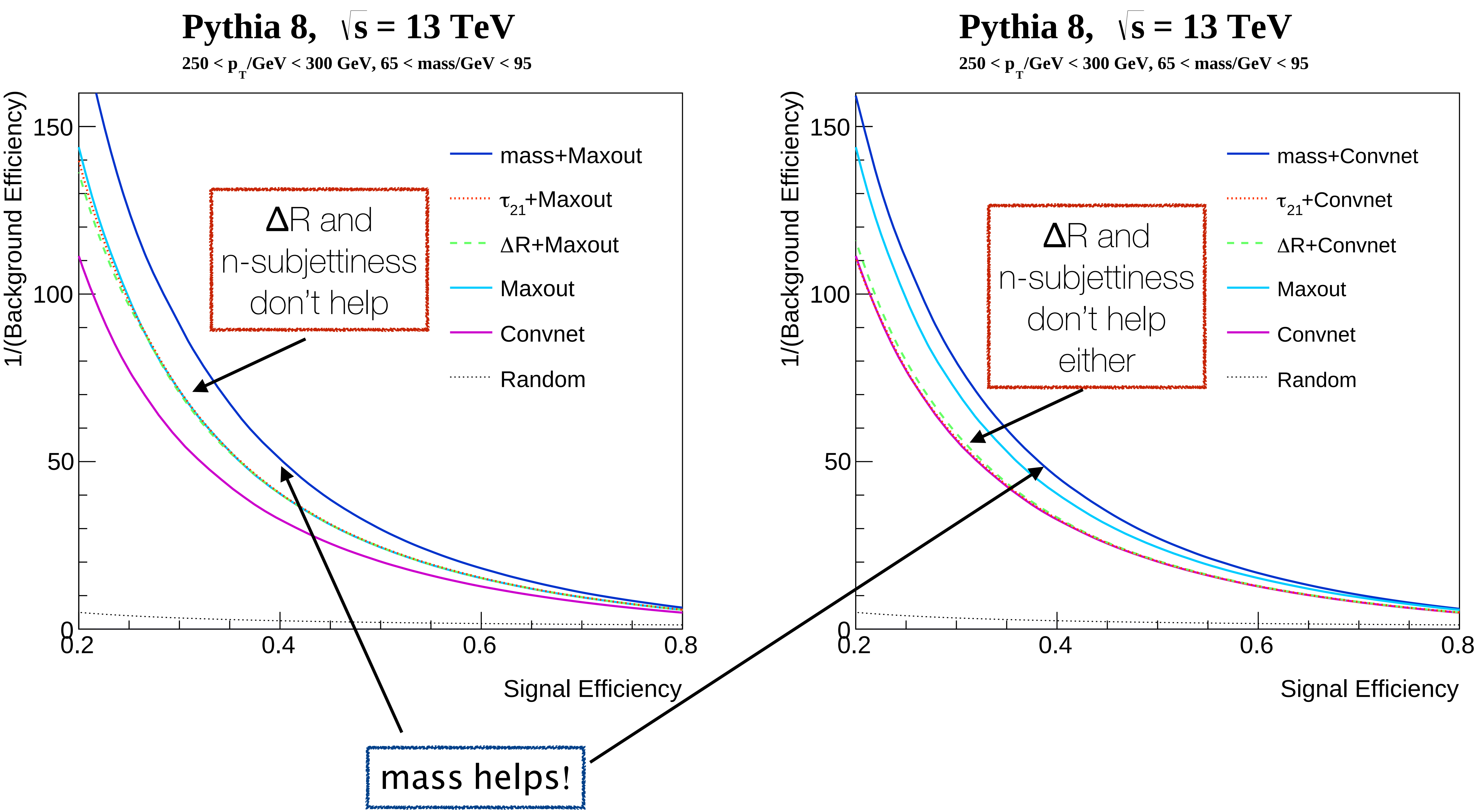




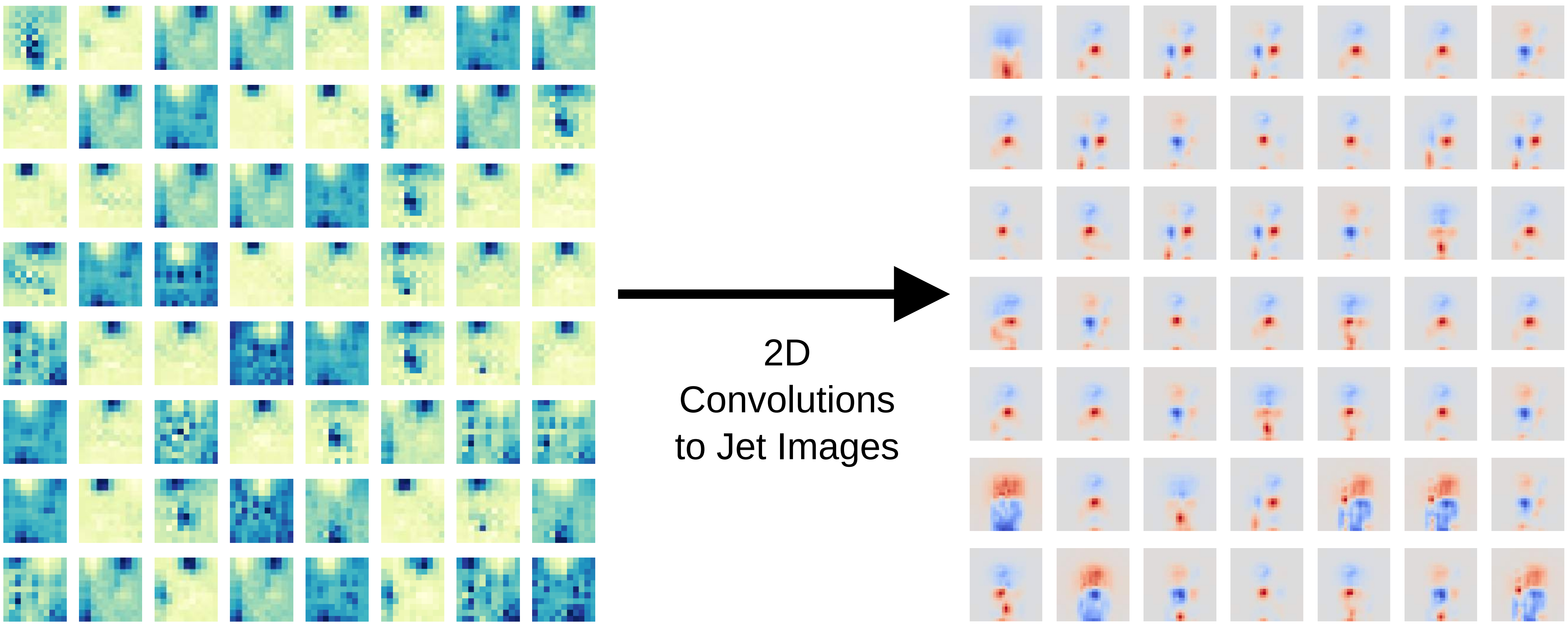
Understanding Performance

- Idea — Mass, **Δ**R, and *n*-subjettiness are *supposed* to capture all information
- Method — 2D likelihood of CNN and mass/**Δ**R/*n*-subjettiness
- Check — if **P(signal | DNN) ~ P(signal | DNN, mass)**, we have learned (to a first order) the effect of mass

The Joint Distribution

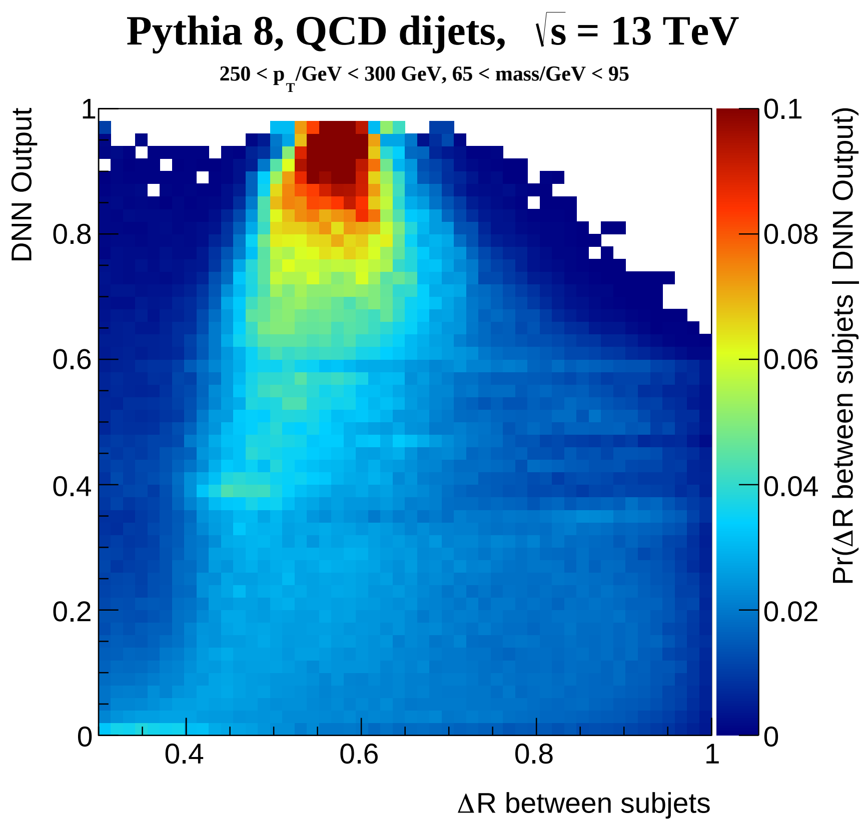


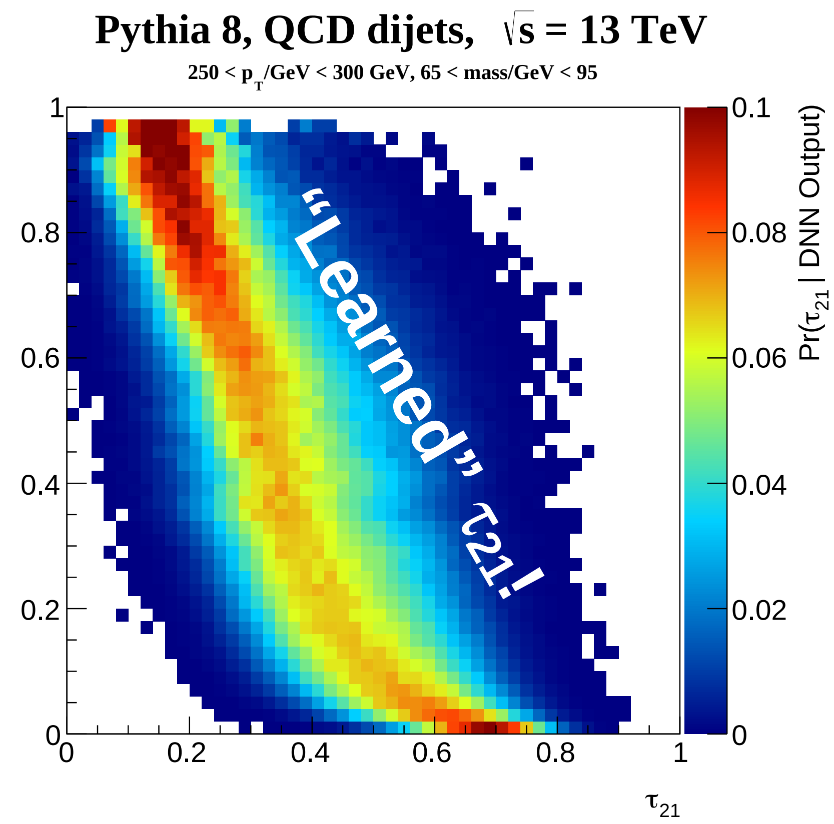
Visual Understanding

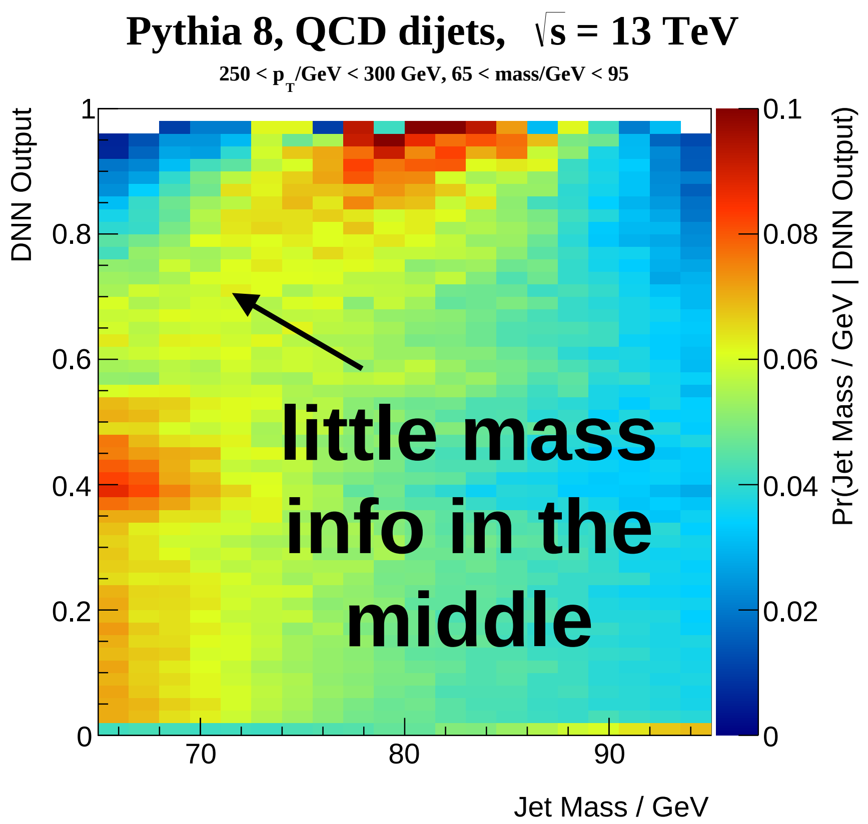


$$X_{\text{sig}} * w - X_{\text{bkg}} * w$$

What do we learn?







Identifying what we learn

- Can re-weight samples such that physics features have no discriminatory power, i.e.,

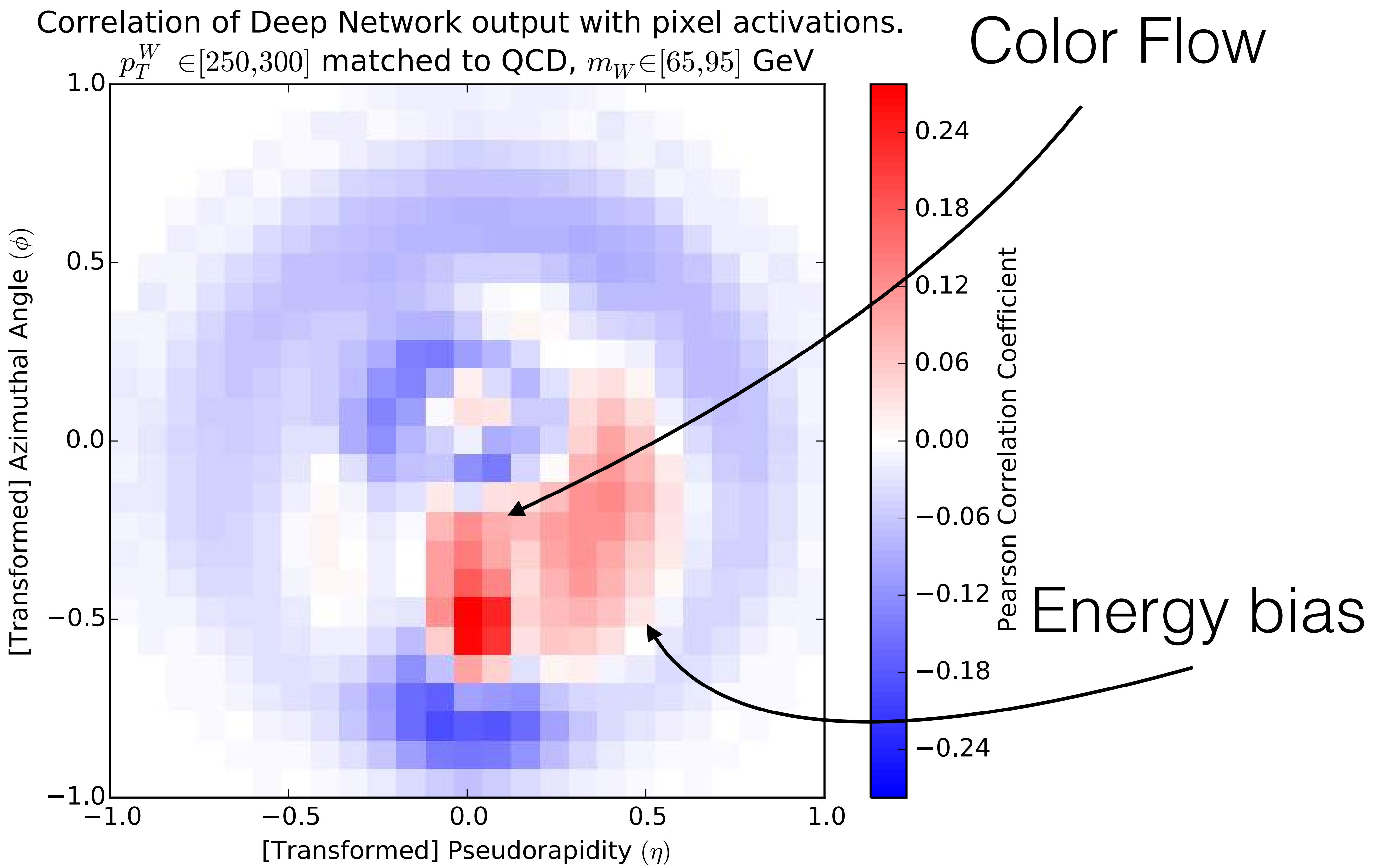
$$\frac{P(\mathrm{signal}|X, m, \tau_{21}, \Delta R, p_T)}{P(\mathrm{bkg}|X, m, \tau_{21}, \Delta R, p_T)} \approx 1$$

- For a given pixel p_{ij}, calculate the weighted Pearson Correlation Coefficient over the entire test set

$$\rho_W(p_{ij}, y_{\mathrm{DNN}})$$

- Can arrange this in an image

What do we learn *beyond*?



Concluding Remarks

- Treat the LHC as a camera — what can we learn about physics?
- Outperform the state-of-the-art —why?
- End-to-end learning — go to lowest level inputs
  - No loss of information
- Great example of ML + HEP collaboration

Citations

1. J. Cogan, M. Kagan, E. Strauss, A. Schwartzman, *Jet-Images: Computer Vision Inspired Techniques for Jet Tagging,* JHEP 02 (2015) 118. Preprint: 1407.5675 [hep-ph].

2. L. de Oliveira, M. Kagan, L. Mackey, B. Nachman, and A. Schwartzman, *Jet Image: Deep Learning Edition,* Preprint: 1511.05190 [hep-ph]

3. Nicolas Vasilache, Jeff Johnson, Michael Mathieu, Soumith Chintala, Serkan Piantino, Yann LeCun, *Fast Convolutional Nets With fbfft: A GPU Performance Evaluation,* 1412.7580 [cs.LG]