Hamburg @ PAHN-Pan: AutoML

Gregor Kasieczka
(gregor.kasieczka@uni-hamburg.de)

PAHN-PaN Task Area 3

2019-08-22

CLUSTER OF EXCELLENCE
QUANTUM UNIVERSE

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

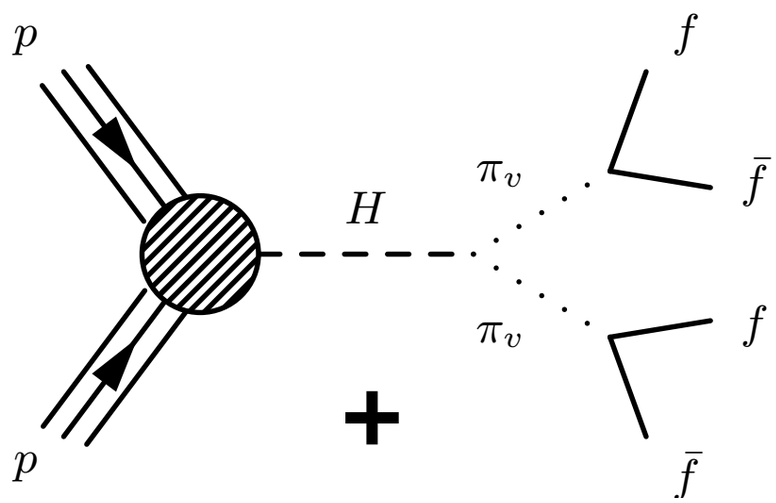
Emmy Noether-Programm
Deutsche Forschungsgemeinschaft
DFG

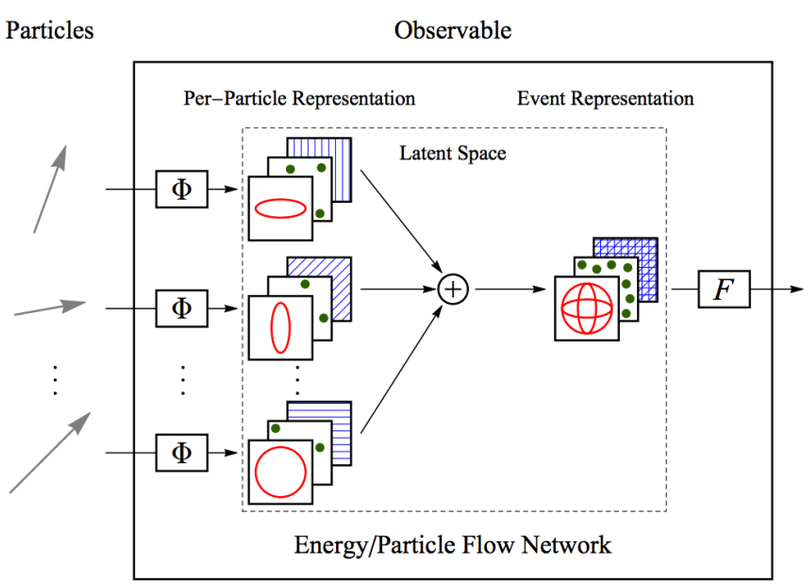
PIER
Partnership of Universität Hamburg and DESY

Bundesministerium für Bildung und Forschung

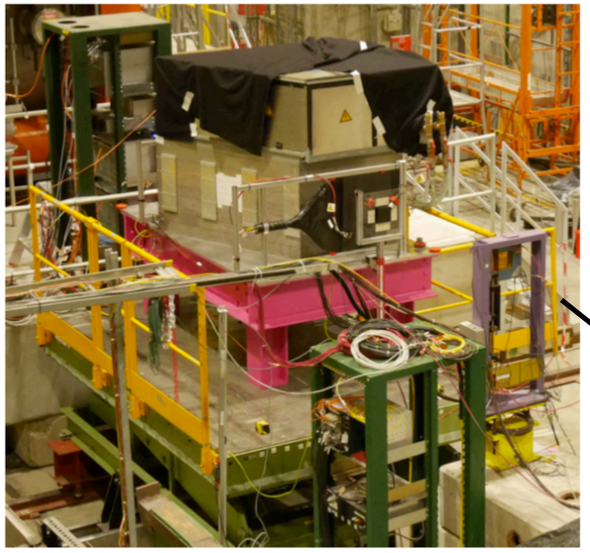
Applications

- Applications of deep learning in CMS analyses (new physics searches, top measurements, ttH) and CALICE calorimeter development

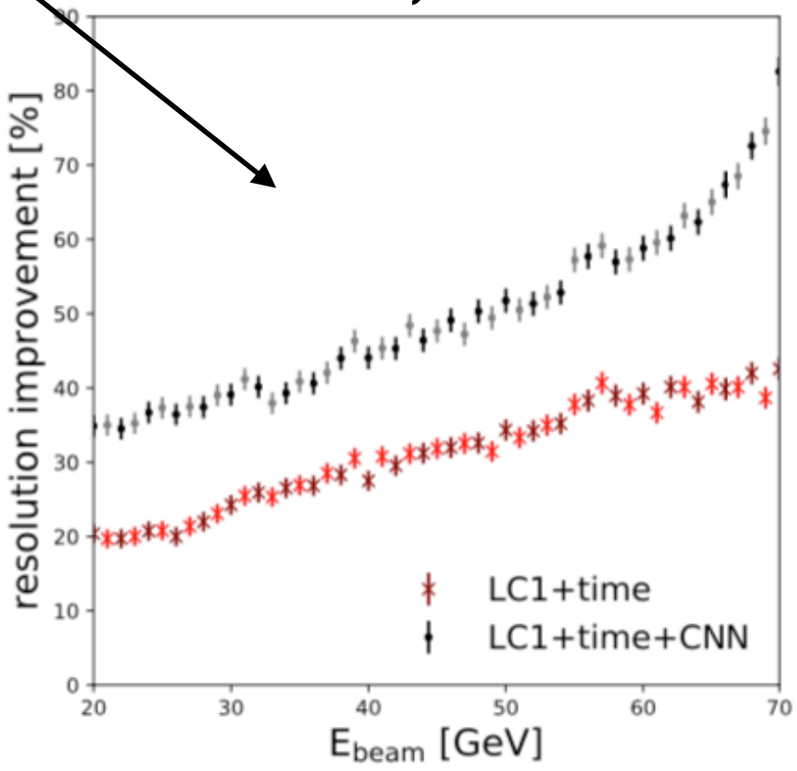




1810.05165



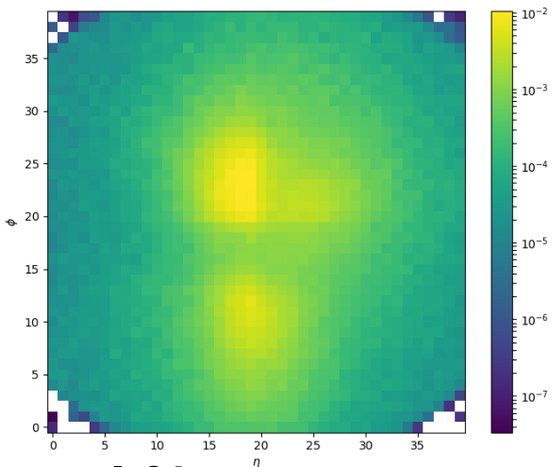
E. Buhmann, MSc Thesis



Boosted Hadronic Top Tagging (with Tilman Plehn)

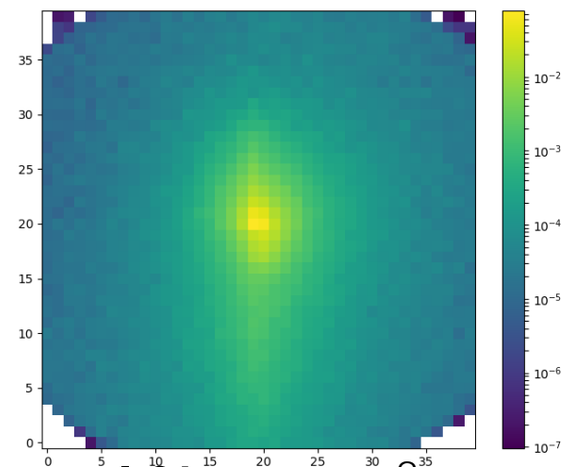
How to tag tops?

Top Quarks



10k images

QCD Jets



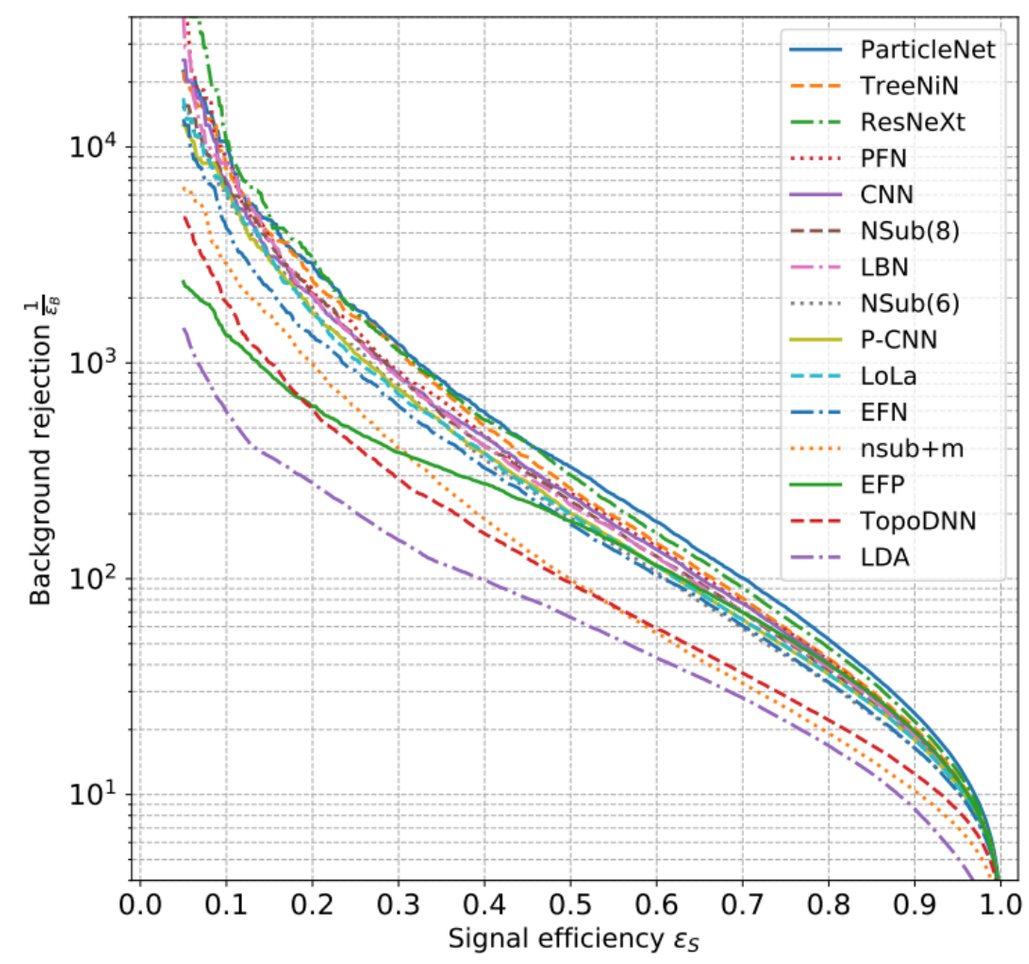
10k images

Image based (CNN) tagging: 1701.08784

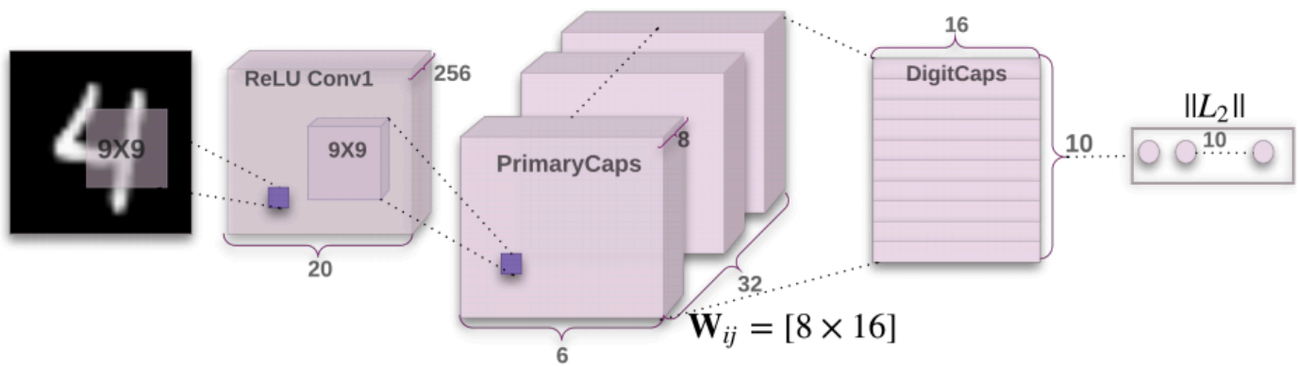
$$\tilde{k}_j \xrightarrow{\text{LoLa}} \hat{k}_j = \begin{pmatrix} m^2(\tilde{k}_j) \\ p_T(\tilde{k}_j) \\ w_{jm}^{(E)}\, E(\tilde{k}_m) \\ w_{jm}^{(d)}\, d_{jm}^2 \end{pmatrix}$$

Physics Architecture: 1707.08966

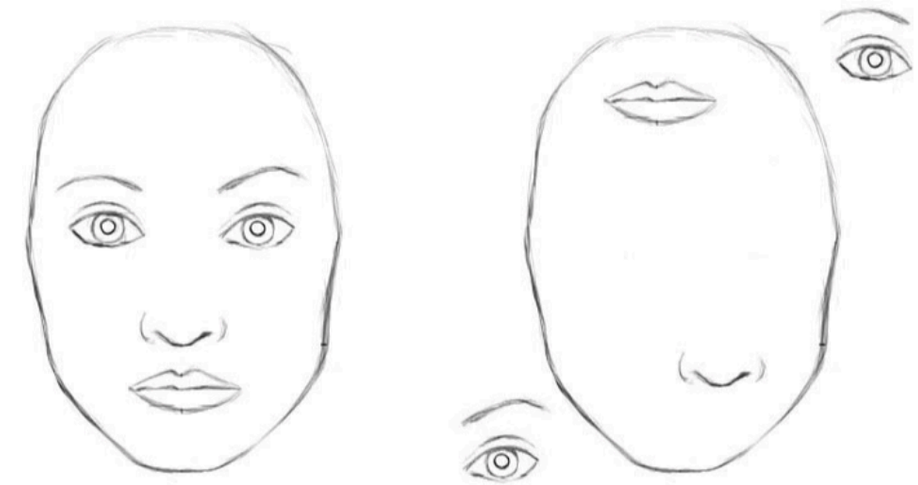
Community performance comparison (<u>dataset public!</u>): 1902.09914



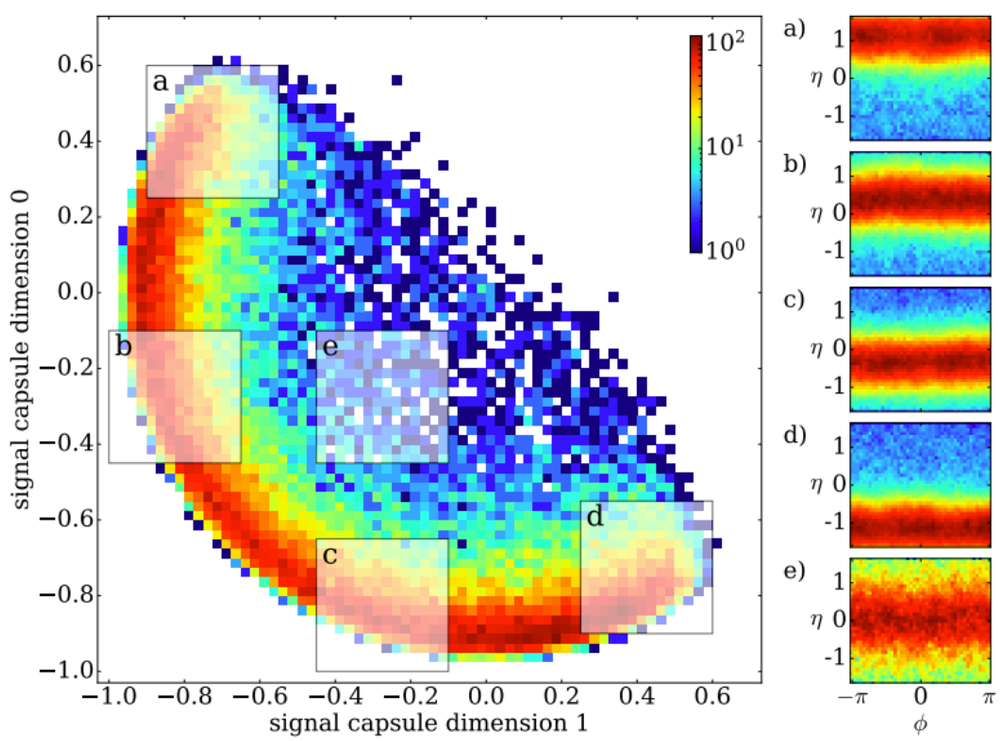
Capsules instead of Convolution



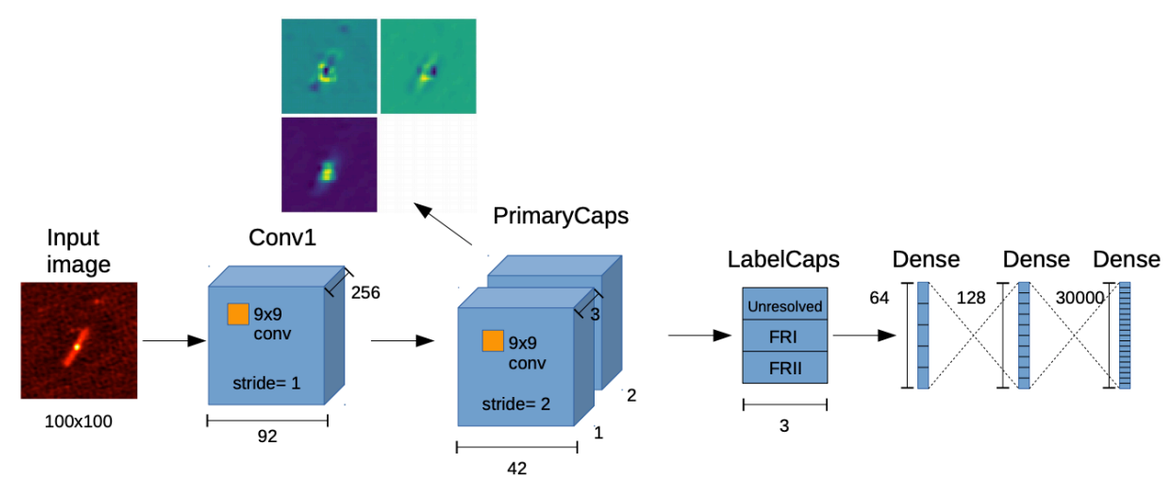
1710.09829



For LHC events (with Tilman Plehn): 1906.11265

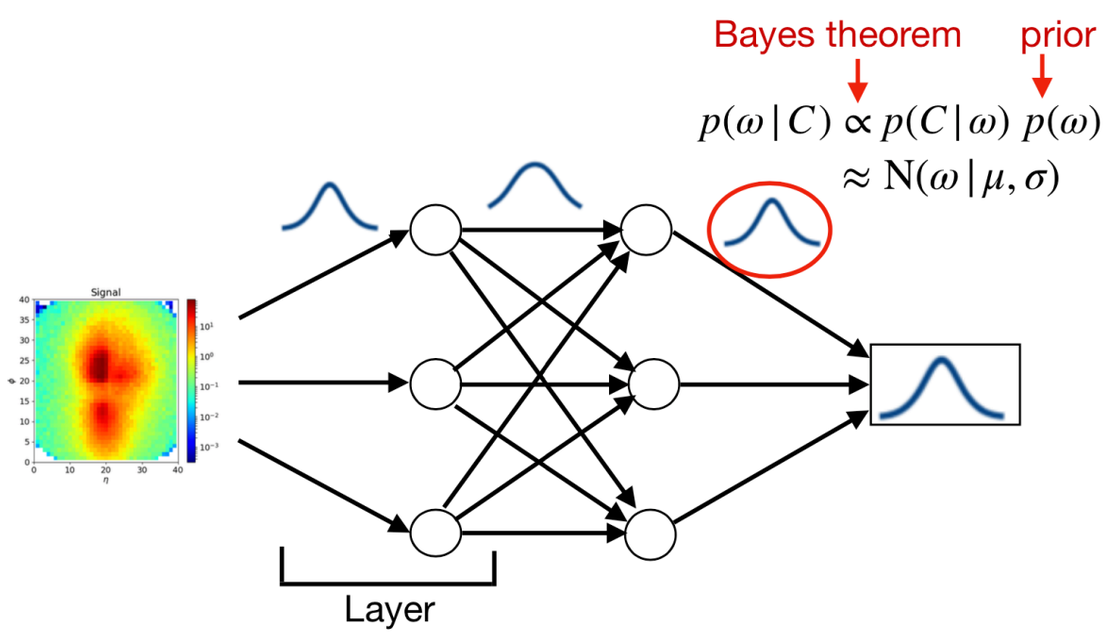


For radiogalaxies (with Marcus Brueggen): 1905.03274

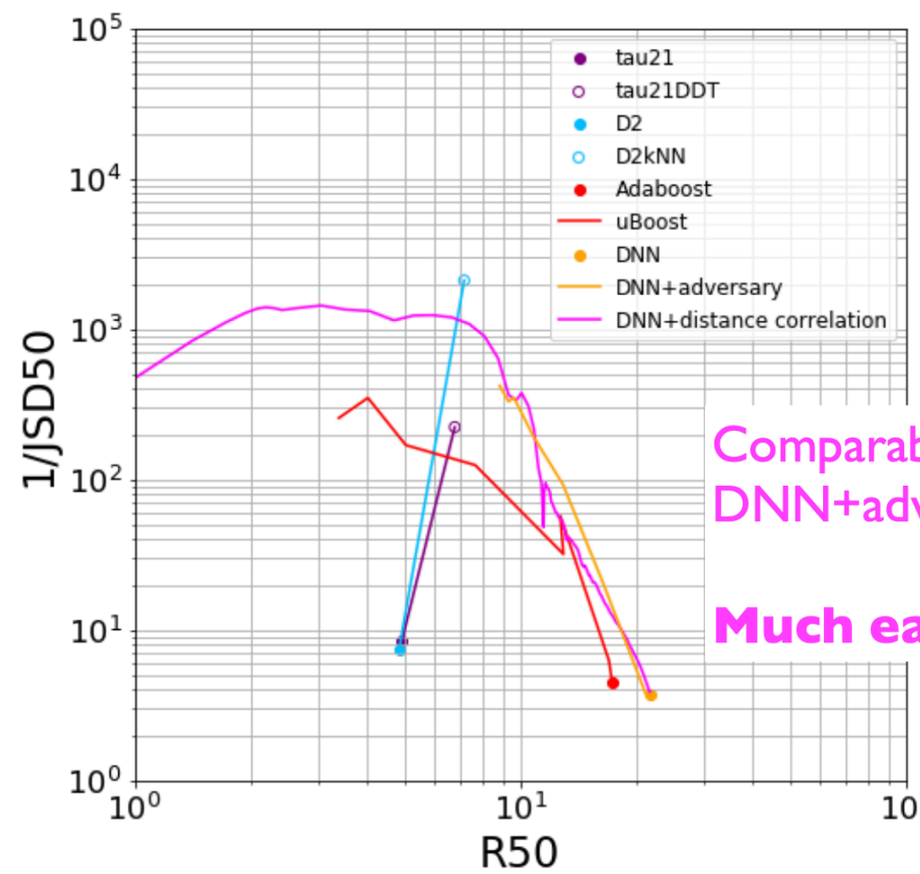


Dealing with Uncertainty

Bayesian Networks (with Tilman Plehn)
1906.11265



Better decorrelation and stability through Disco
(with David Shih, coming soon)

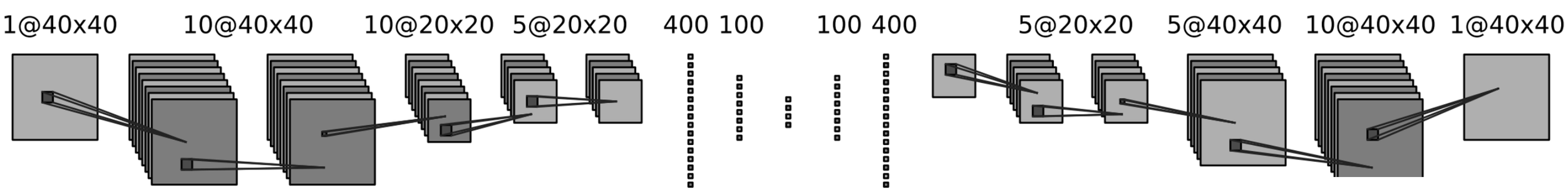


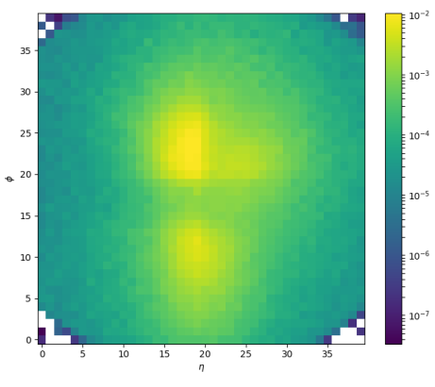
Comparable performance to DNN+adversary.

Much easier to train.

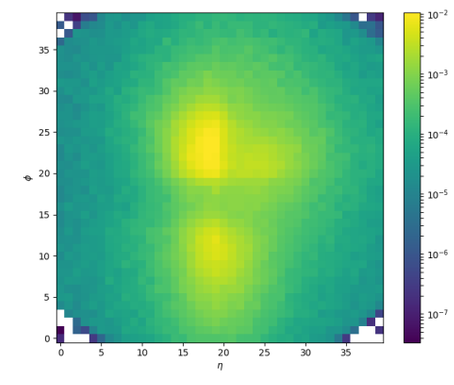
Model independent discovery

New physics searches based on anomaly finding with auto encoders (with Tilman Plehn), 1808.08979



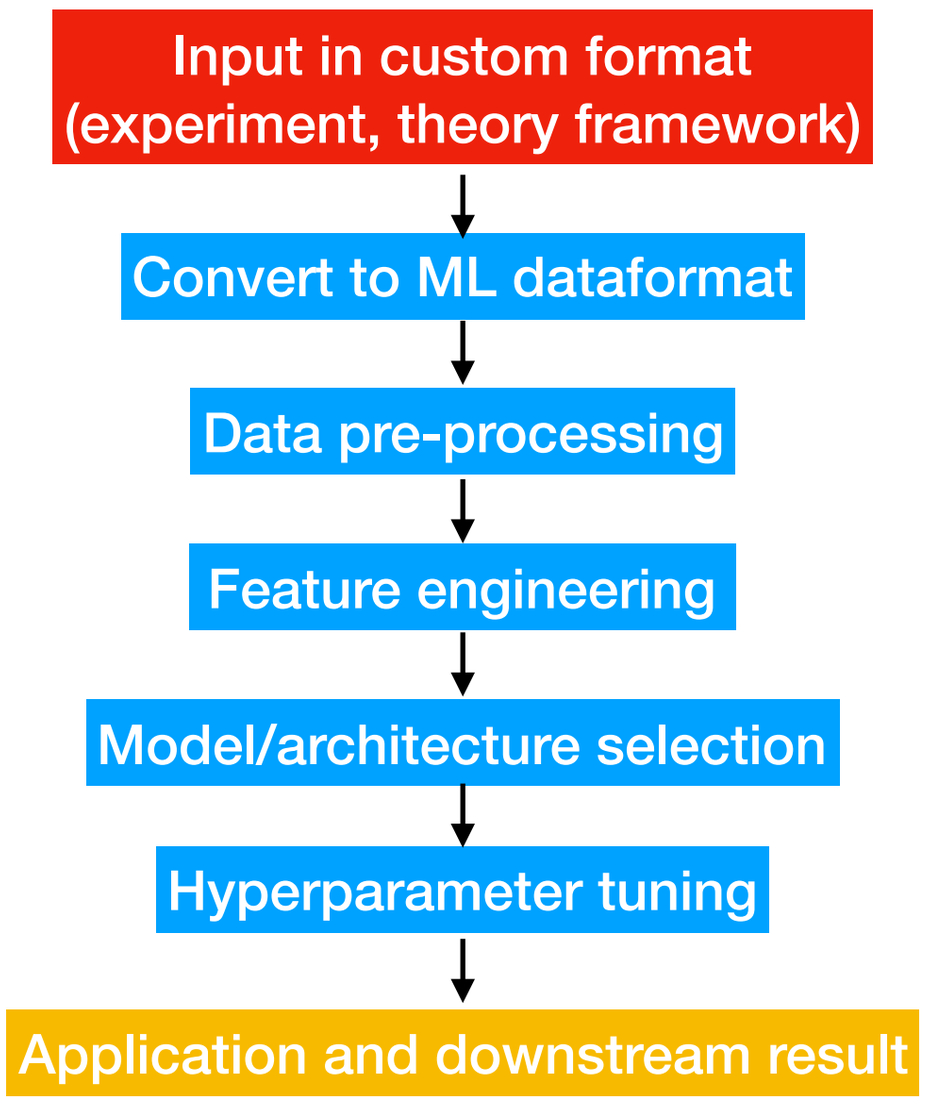


$$L_{\text{auto}} = \sum_{1600 \text{ pixels}} \left(k_T^{\text{norm,in}} - k_T^{\text{auto}} \right)^2$$



What next?

Machine Learning pipeline:



Large part of research time spent on these tasks

Similar for many of our applications

Automate these: AutoML

(the goal is not to produce the best model for each scenario but provide a reliable baseline with minimal user effort)

Some thoughts:

AutoML is a (much) bigger research effort outside of our domain

Do not compete with, find a way to build a fully automated toolchain (as a service?)

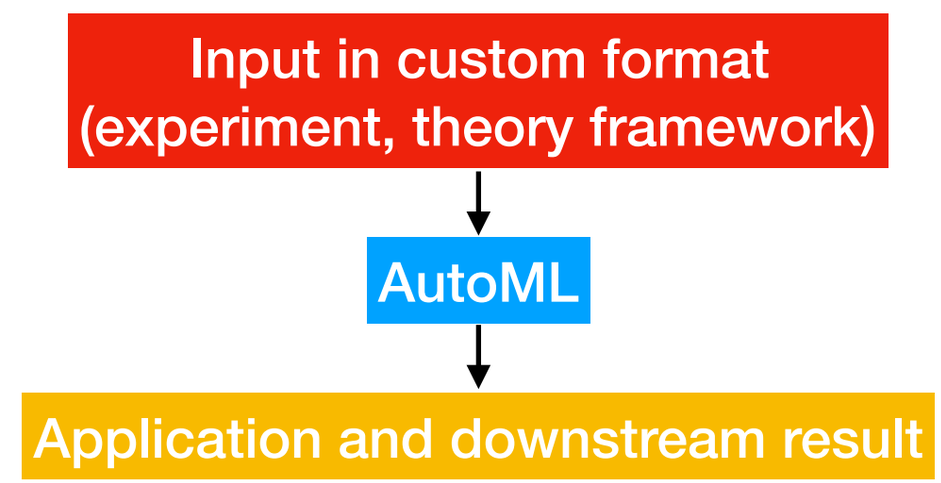
Make sure it works for all our examples (limit to supervised tasks)

Build our domain expertise into the automated model selection (our data is inherently different from phots and unstructured measurements)

Invite other NFDI consortia to use, “only” need to provide input conversion

Interplay with technical developments in PAHN-PaN

Goal:



Thank you!