Managed by Fermi Research Alliance, LLC for the U.S. Department of Energy Office of Science

Imaging datasets

Gabriel Perdue on behalf of many folks,

2017 / May / 12

DS@HEP 2017

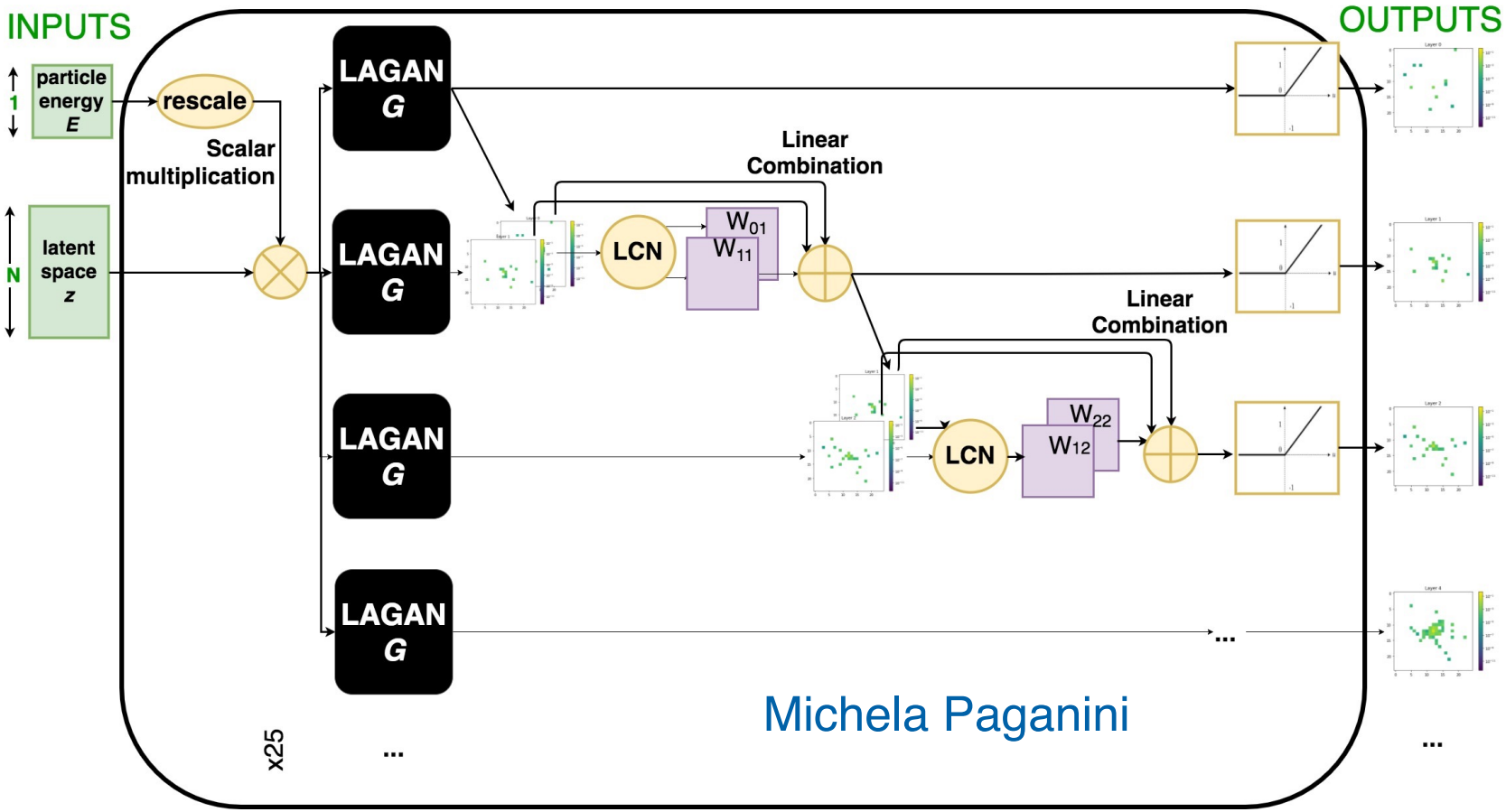
Imaging datasets

- Amir presented a nice framework for managing deep learning experiments on his cluster at UTA.
  - Assumes Keras but many nice features for managing large datasets and parallel streaming/serving data across the cluster.
  - Most people who tried were probably able to take the starter notebooks and hook up simple neural networks.
- Interesting progress on a number of other problems:
  - running pion/electron separation in the LCD crystal calorimeter,
  - energy regression for electrons in the LCD crystal calorimeter - from zero to a running solution (to be iterated on),
  - three different versions of the GANs running on the LCD data set,
  - some work on reproducing some results discussed here, e.g., the CaloGAN repository - the documentation is exemplary! - went from zero to running in ~10 minutes on a laptop.

Mea culpa

- We were working with very large datasets (multi TB, production style code more than macro style code) and ran into some technical difficulties associated with many users attempting to work with the data at once.
  - Order of magnitude more users on the cluster than it is used to.
- Despite some of the technical difficulties, we got a bunch of new code to incorporate into the framework (seed framework for an open data project?):
  - energy regression
  - GANs
- Will all be available in the notebooks and code repository soon.

Can use CaloGAN for LCD!

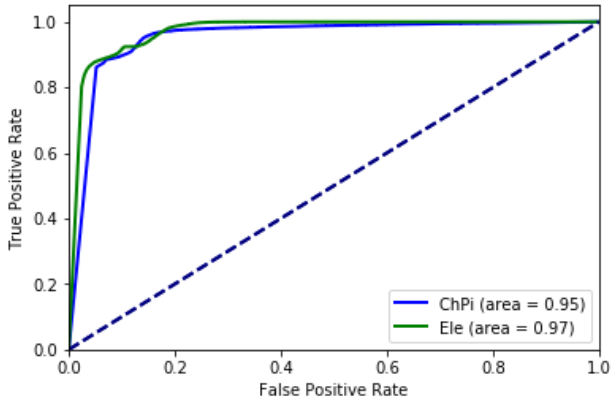


Note: need extra care because of high dimensionality (n, 25, 25, 25, 1) and sparsity

Electron vs. Charged Pion Discrimination

- Charged pions can fake electrons via charge exchange in ECAL
  - $\pi^{\pm}$ n → π^0 p, π^0 → γγ produces EM shower
- Modify 4-particle classifier ($e^{\pm}$, $\pi^{\pm}$, π^0,γ) in CaloDNN-Experiment to do $e^{\pm}$ vs. $\pi^{\pm}$ discrimination → initial results reasonable:

```
Initial Score: [0.72727116297200778, 0.32367509664948452]
Epoch 1/2
878/878 [==============================] - 61s - loss: 0.2967
- acc: 0.9459 - val_loss: 0.3039 - val_acc: 0.9451
Epoch 2/2
878/878 [==============================] - 61s - loss: 0.1194
- acc: 0.9722 - val_loss: 0.9029 - val_acc: 0.9273
Evaluating score on test sample...
Final Score: [0.90411072243408519, 0.94970497917751262]
```



- Next steps:
  - Generate large $e^{\pm}$ and $\pi^{\pm}$ samples with fixed energy and H/E < 0.2 (to select charged pions undergoing charge exchange in ECAL)
  - Repeat DNN classifier and compare to simple, feature-based MVA
  - Explore volumetric CNNs

Amir Farbin, Ben Hooberman, Maurizio Pierini, Ryan Reece, Jean-Roch Vlimant, Matt Zhang