Machine learning for particle physicists

IV. Neural networks for particle physics problems

Anja Butter

26th Vietnam School of Physics

Recap - How to train better networks

1. Data preprocessing
   - Rescaling, PCA
2. Network initialization
   - Glorot/HE, Normal/uniform
3. Optimizing the training procedure
   - Learning rate scheduling, momentum, Adagrad, Adam
4. Regularization
   - Via early stopping, additional loss, dropout, or Batchnorm
5. Hyperparameter tuning
   - get a feeling for the network, random search, Bayesian optimization

Neural networks in particle physics

Selected examples:

1. CNN for jet classification
2. Graph networks & Particle Net
3. Autoencoders for anomaly detection

1. CNN for jet classification

A short history of image based jet classification

[1407.5675] J. Cogan, M. Kagan, E. Strauss, and A. Schwartzman,
define jet image (classification with Fisher discriminant)

A short history of image based jet classification

[1407.5675] J. Cogan, M. Kagan, E. Strauss, and A. Schwartzman,
define jet image (classification with Fisher discriminant)

[1501.05968] L. G. Almeida, M. Backovic, M. Cliche, S. J. Lee, and M. Perelstein,
dense layers for image classification

A short history of image based jet classification

[1407.5675] J. Cogan, M. Kagan, E. Strauss, and A. Schwarztman,
define jet image (classification with Fisher discriminant)

[1501.05968] L. G. Almeida, M. Backovic, M. Cliche, S. J. Lee, and M. Perelstein,
dense layers for image classification

[1511.05190] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman, and A. Schwartzman,
CNN for images, W vs QCD

A short history of image based jet classification

[1407.5675] J. Cogan, M. Kagan, E. Strauss, and A. Schwarztman,
define jet image (classification with Fisher discriminant)

[1501.05968] L. G. Almeida, M. Backovic, M. Cliche, S. J. Lee, and M. Perelstein,
dense layers for image classification

[1511.05190] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman, and A. Schwartzman,
CNN for images, W vs QCD

[1603.09349] P. Baldi, K. Bauer, C. Eng, P. Sadowski, and D. Whiteson,
detector effects and pile up

A short history of image based jet classification

[1407.5675] J. Cogan, M. Kagan, E. Strauss, and A. Schwarztman,
define jet image (classification with Fisher discriminant)

[1501.05968] L. G. Almeida, M. Backovic, M. Cliche, S. J. Lee, and M. Perelstein,
dense layers for image classification

[1511.05190] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman, and A. Schwartzman,
CNN for images, W vs QCD

[1603.09349] P. Baldi, K. Bauer, C. Eng, P. Sadowski, and D. Whiteson,
detector effects and pile up

[1612.01551] P. T. Komiske, E. M. Metodiev, and M. D. Schwartz,
quark/gluon discrimination

A short history of image based jet classification

[1407.5675] J. Cogan, M. Kagan, E. Strauss, and A. Schwarztman,
define jet image (classification with Fisher discriminant)

[1501.05968] L. G. Almeida, M. Backovic, M. Cliche, S. J. Lee, and M. Perelstein,
dense layers for image classification

[1511.05190] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman, and A. Schwartzman,
CNN for images, W vs QCD

[1603.09349] P. Baldi, K. Bauer, C. Eng, P. Sadowski, and D. Whiteson,
detector effects and pile up

[1612.01551] P. T. Komiske, E. M. Metodiev, and M. D. Schwartz,
quark/gluon discrimination

[1701.08784] G. Kasieczka, T. Plehn, M. Russell, and T. Schell,
top tagging

A short history of image based jet classification

[1407.5675] J. Cogan, M. Kagan, E. Strauss, and A. Schwarztman,
define jet image (classification with Fisher discriminant)

[1501.05968] L. G. Almeida, M. Backovic, M. Cliche, S. J. Lee, and M. Perelstein,
dense layers for image classification

[1511.05190] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman, and A. Schwartzman,
CNN for images, W vs QCD

[1603.09349] P. Baldi, K. Bauer, C. Eng, P. Sadowski, and D. Whiteson,
detector effects and pile up

[1612.01551] P. T. Komiske, E. M. Metodiev, and M. D. Schwartz,
quark/gluon discrimination

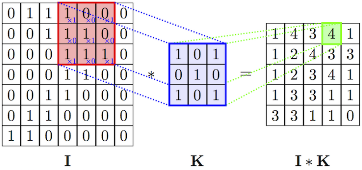
[1701.08784] G. Kasieczka, T. Plehn, M. Russell, and T. Schell,
top tagging

[1803.00107] S. Macaluso, D. Shih
improved top tagging

Types of image layers

- Previously: Dense Layer

- Convolutional Layer



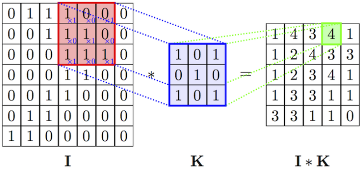
https://cambridgespark.com/content/tutorials/convolutional-neural-networks-with-keras/index.html

The network learns the parameters of K

Types of image layers

- Previously: Dense Layer

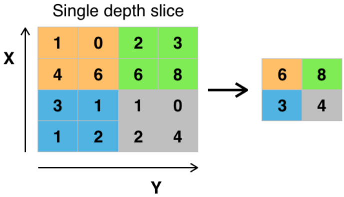
- Convolutional Layer



https://cambridgespark.com/content/tutorials/convolutional-neural-networks-with-keras/index.html

The network learns the parameters of K

- (Max)Pooling Layer

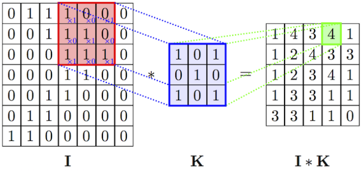


https://commons.wikimedia.org/wiki/File:Max_pooling.png

Types of image layers

- Previously: Dense Layer

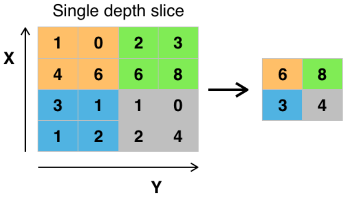
- Convolutional Layer



https://cambridgespark.com/content/tutorials/convolutional-neural-networks-with-keras/index.html

The network learns the parameters of K

- (Max)Pooling Layer

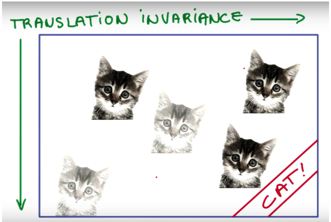


https://commons.wikimedia.org/wiki/File:Max_pooling.png

- Followed by flattening layer $(n \times n) \to n^2$

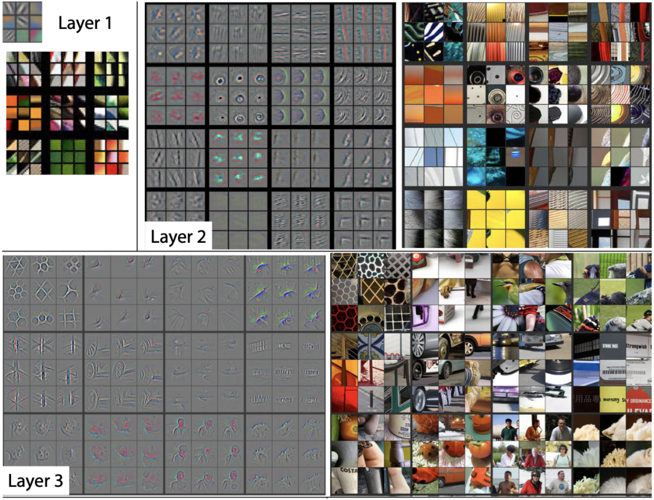
CNN features

- translational equivariance
  "K(shifted image) = shift K(image)"
- pooling $\to$ translational invariance
  NN(cat in left corner) = cat
  NN(cat in right corner) = cat
- weight sharing



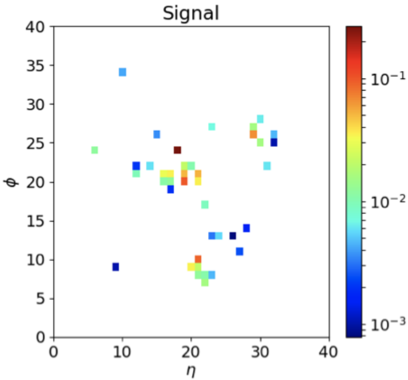
https://www.cc.gatech.edu/ san37/post/dlhc-cnn/

Intermediate features



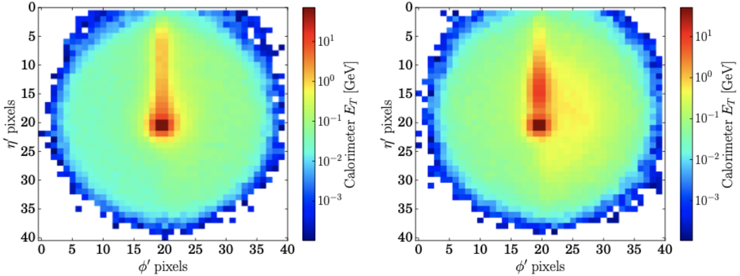
1311.2901 M. D. Zeiler, R. Fergus

A QCD jet image



1701.08784 G. Kasieczka, T. Plehn, M. Russell, and T. Schell

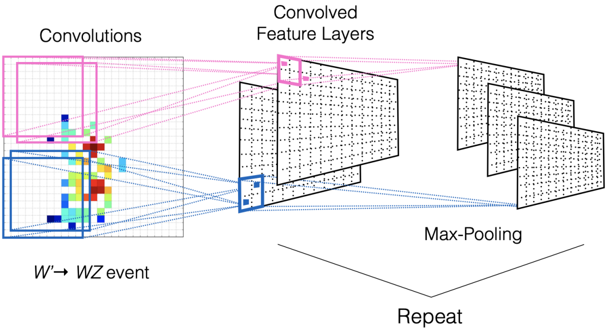
A QCD jet image



1701.08784 G. Kasieczka, T. Plehn, M. Russell, and T. Schell

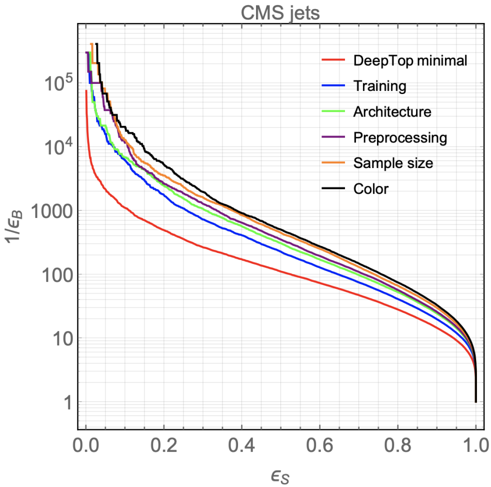
Average over 10000 QCD (left) and top (right) jets

Processing detector images



1511.05190 L. Oliveira, M. Kagan, L. Mackey, B. Nachman, A. Schwartzman

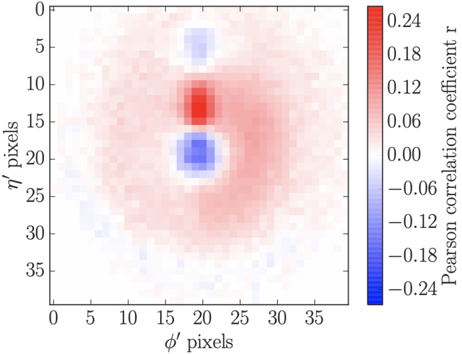
Performance



1803.00107 S. Macaluso, D. Shih

Looking into the black box

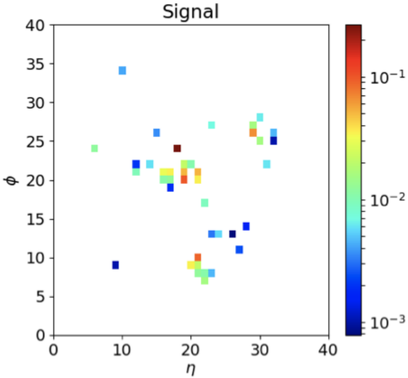
Pearson formula $r_{ij} = \dfrac{\sum_{images}(x_{ij} - \bar{x}_{ij})(y - \bar{y})}{\sqrt{\sum_{images}(x_{ij} - \bar{x}_{ij})^2}\sqrt{\sum_{images}(y - \bar{y})^2}}$



1701.08784 G. Kasieczka, T. Plehn, M. Russell, and T. Schell

2. Graph based networks

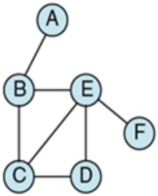
Our image ain't an image



Sparse! (mostly empty)
Can not include additional information eg. from tracker
Better represented by individual points

Graph representation

Graph consists of nodes and edges:



A
B
C
D
E
F

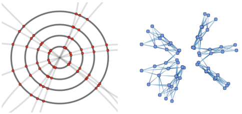
Vertex vector

	A	B	C	D	E	F
A	0	1	0	0	0	0
B	1	0	1	0	1	0
C	0	1	0	1	1	0
D	0	0	1	0	1	0
E	0	1	1	1	0	1
F	0	0	0	0	1	0

Adjacency matrix

commons.wikimedia.org

- tracks in detector
- objects in event
- particles in jet
- ...

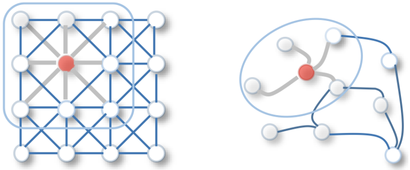


2007.13681J. Shlomi, P. Battaglia, J.-R. Vlimant

Graph network

pixels $\rightarrow$ point cloud
find k nearest neighbours $\rightarrow$ graph edges

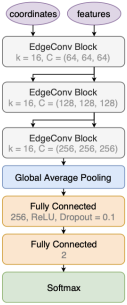
Edge convolution $$\boldsymbol{x}'_i = \frac{1}{k}\sum_{j=1}^{k} h_\Theta(\boldsymbol{x}_i, \boldsymbol{x}_{i_j} - \boldsymbol{x}_i)$$



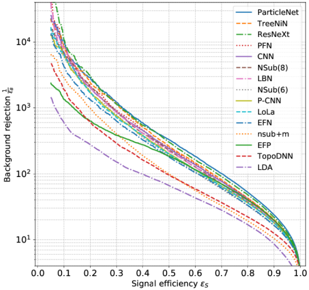
1901.00596 Z. Wu, et al.

ParticleNet

1902.08570 H. Qu, L. Gouskos



(a) ParticleNet



Excellent performance!

3. Autoencoder for anomaly detection

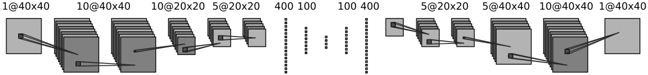
Do observed jets really look like QCD jets?

Anomaly detection

Unsupervised training

Idea : Network should learn what a QCD jet looks like

Solution : Autoencoder



Encoder: compress $image_{in} \rightarrow z$

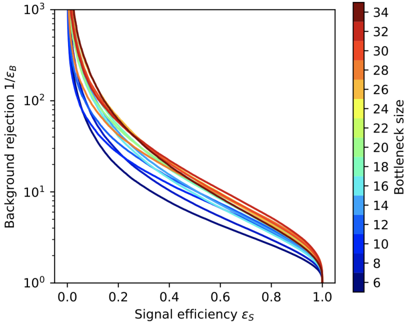
Decoder: build $z \rightarrow image_{AE}$

$$\mathcal{L} = \sum_{pixel}(image_{in} - image_{AE})^2$$

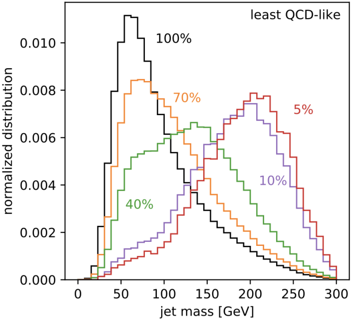
Testing with top jets

Use reconstruction loss to define how QCD like an event is

$$\mathcal{L} > D \rightarrow \text{signal (top)}$$



What is "QCD like"?



→ network focuses on jet mass

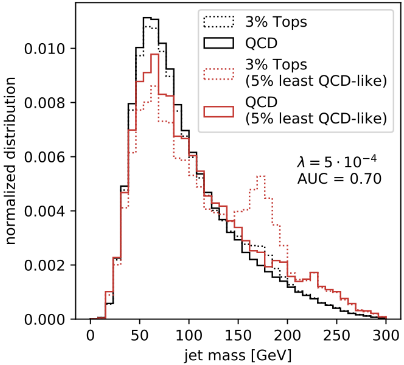
Can we force the network to ignore the jet mass?

Adversarial training

Train second network Φ to extract the mass from the output image
$\rightarrow$ new loss:

$$\mathcal{L} = \mathcal{L}_{auto} - \lambda \mathcal{L}_{adv}(M)$$

$$\mathcal{L}_{adv}(M) = [\Phi(|image_{in} - image_{AE}|) - M]^2$$



$\rightarrow$ Mass dependence successfully removed!
$\rightarrow$We will make use of adversarial training in the last lecture!

Summary

Machine learning has the potential to boost particle physics!

Examples:

1. CNN for jet classification
2. Graph networks & Particle Net
3. Autoencoders for anomaly detection
4. . . . and many more