Identifying QCD transition using Deep Learning

Kai Zhou

Collaborators: Long-Gang Pang, Nan Su,
Hannah Peterson, Horst Stoecker, Xin-Nian Wang

arXiv:1612.04262

Outline

- Introduction to Deep Learning
- Simple example of DL and application in Physics
- QCD EoS-meter in HIC using DL
- Summary and outlook

What is deep learning?

Artificial Intelligence (AI)

Machine Learning (ML)

Deep Learning (DL)

Booming in theory and applications!

1. Big Data
2. GPU parallel
3. New architecture

2006

Geoffrey Hinton

Supervised learning---most commen DL

RL

Reinforcement Learning
(or Deep Q-learning)

Playing Games

RNN/LSTM/GRU

Recurrent Neural Network
Long Short Term Memory
Gated Recurrent Unit

Natural Language processing

CNN

Convolution Neural Network

Image recognition/ classification

Fancy applications (RL)

AlphaGo (by Google DeepMind) beat human master

Self-driving cars Robot control Playing Games like flappybird, Starcraft

RNN/LSTM for natural language generation

Chinese Poetry Generation with Planning based Neural Network

Zhe Wang[†], **Wei He**[‡], **Hua Wu**[‡], **Haiyang Wu**[‡], **Wei Li**[‡], **Haifeng Wang**[‡], **Enhong Chen**[†]
[†]University of Science and Technology of China, Hefei, China
[‡]Baidu Inc., Beijing, China

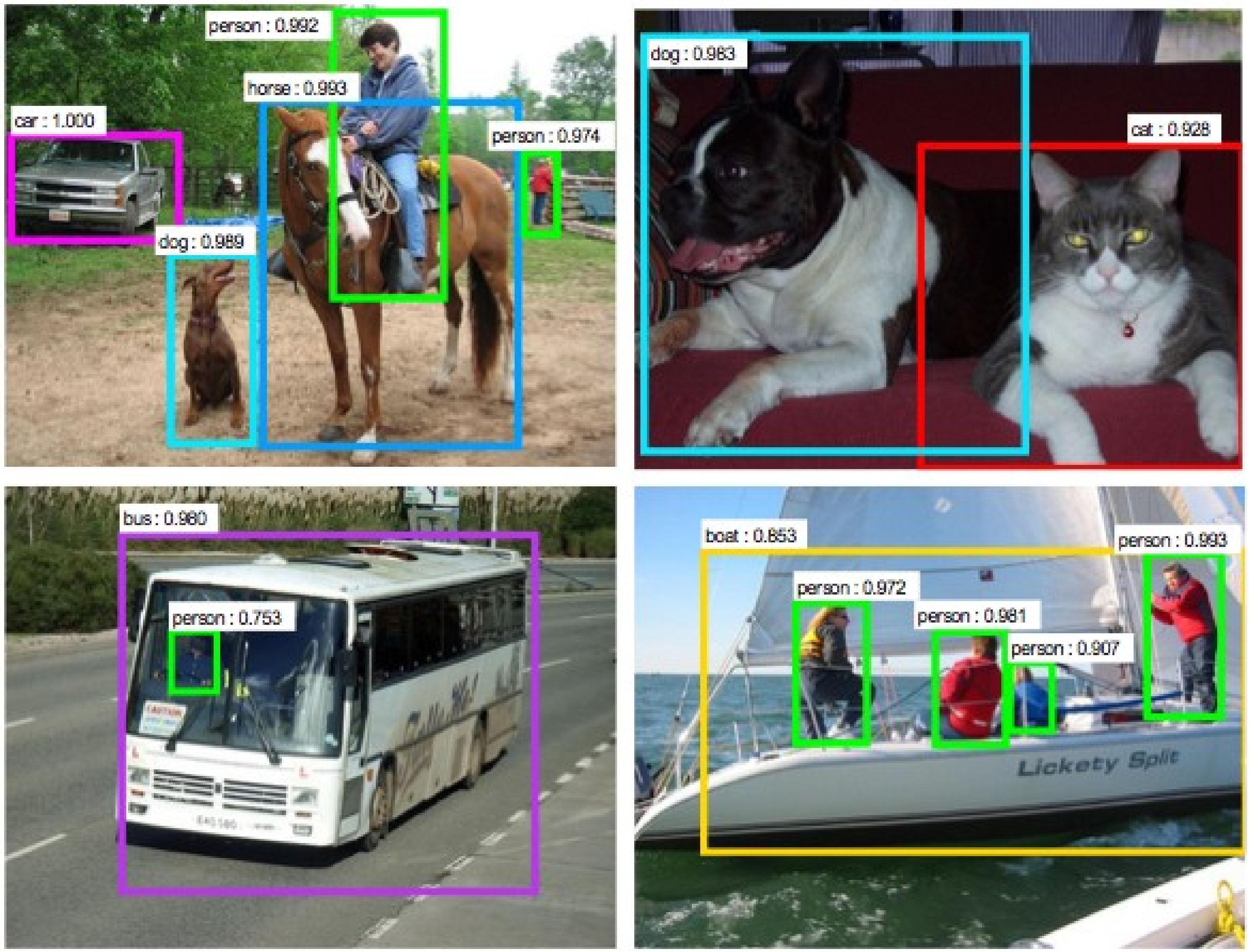
秋夕湖上	秋夕湖上
By a Lake at Autumn Sunset	By a Lake at Autumn Sunset
一夜秋凉雨湿衣，	荻花风里桂花浮，
A cold autumn rain wetted my clothes last night,	The wind blows reeds with osmanthus flying,
西窗独坐对夕晖。	恨竹生云翠欲流。
And I sit alone by the window and enjoy the sunset.	And the bamboos under clouds are so green as if to flow down.
湖波荡漾千山色，	谁拂半湖新镜面，
With mountain scenery mirrored on the rippling lake,	The misty rain ripples the smooth surface of lake,
山鸟徘徊万籁微。	飞来烟雨暮天愁。
A silence prevails over all except the hovering birds.	And I feel blue at sunset .

Table 6: A pair of poems selected from the blind test. The left one is a machine-generated poem, and the right one is written by Shaoti Ge, a poet lived in the Song Dynasty.

arXiv: 1610.09889v1

Object detection using CNN

Example detection results of Faster R-CNN



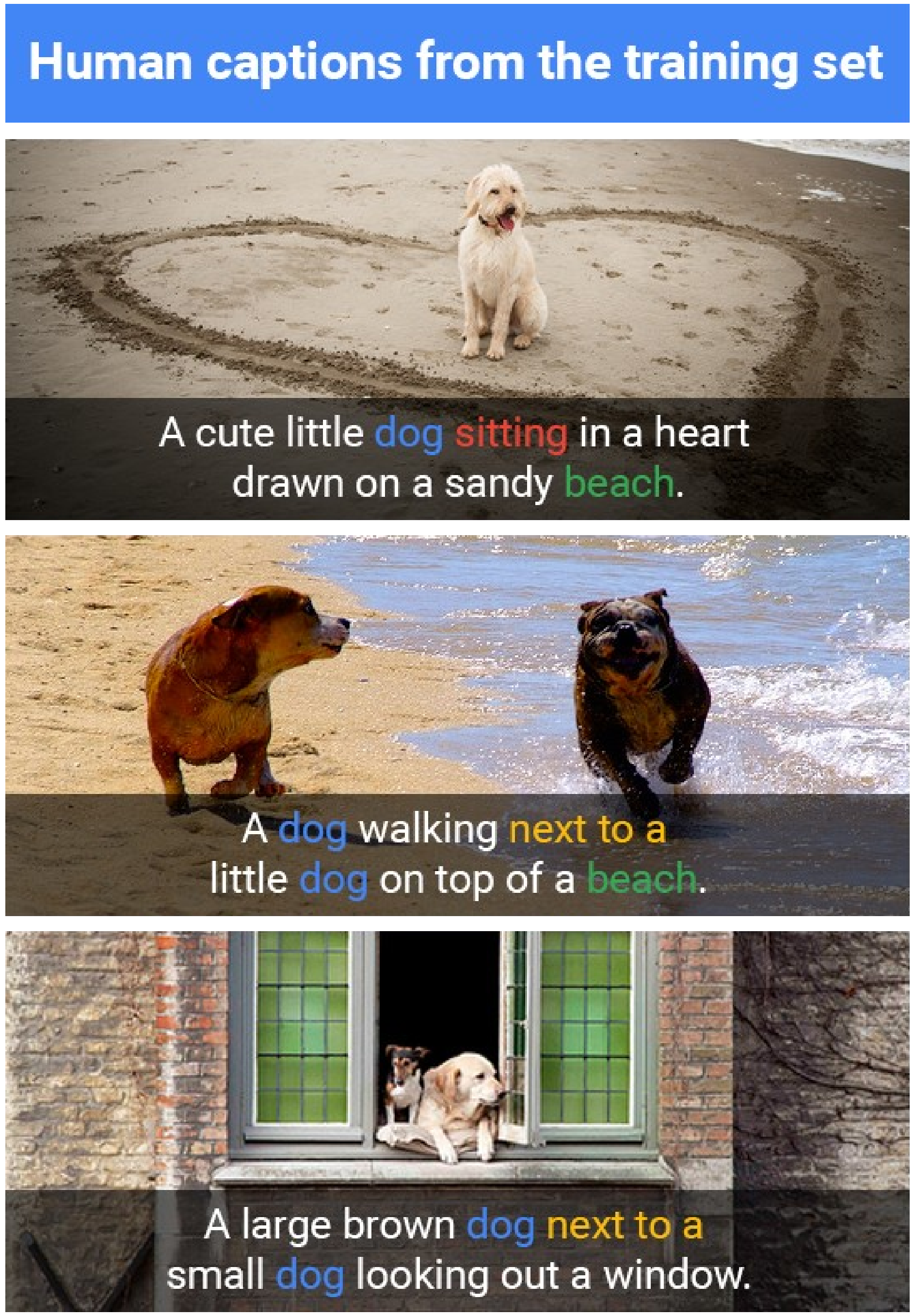
Shaoqing Ren, Kaiming He, Ross Girshick, & Jian Sun. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks". NIPS 2015.

Slides from KaiMing, He's recent talk

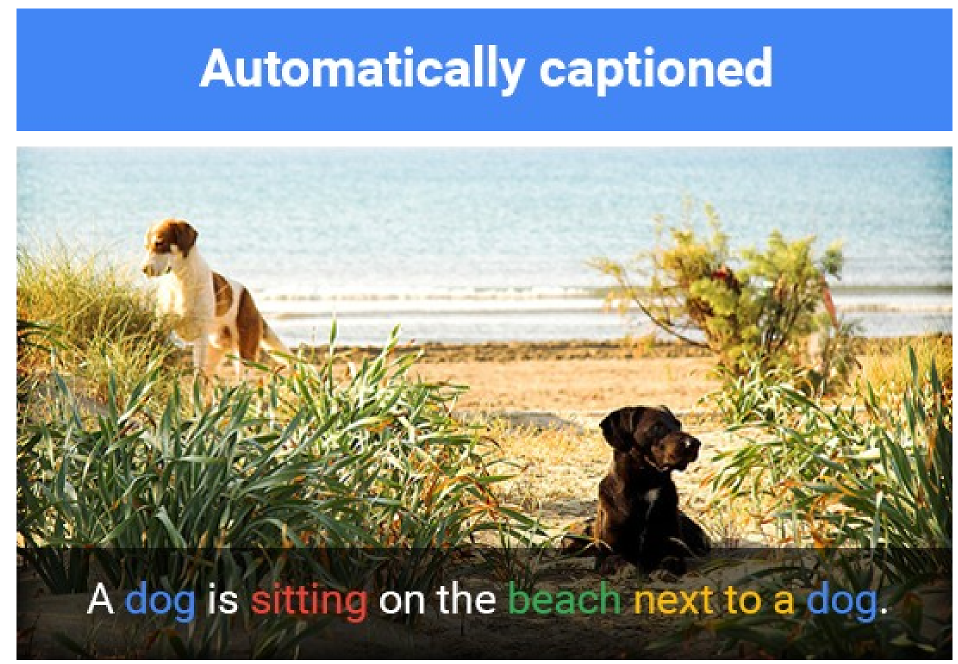
Artistic style transfer using CNN

arXiv:1508.06576 style can be modelled and transferred!

Hybrid CNN + LSTM for image captioning



By Google Brain



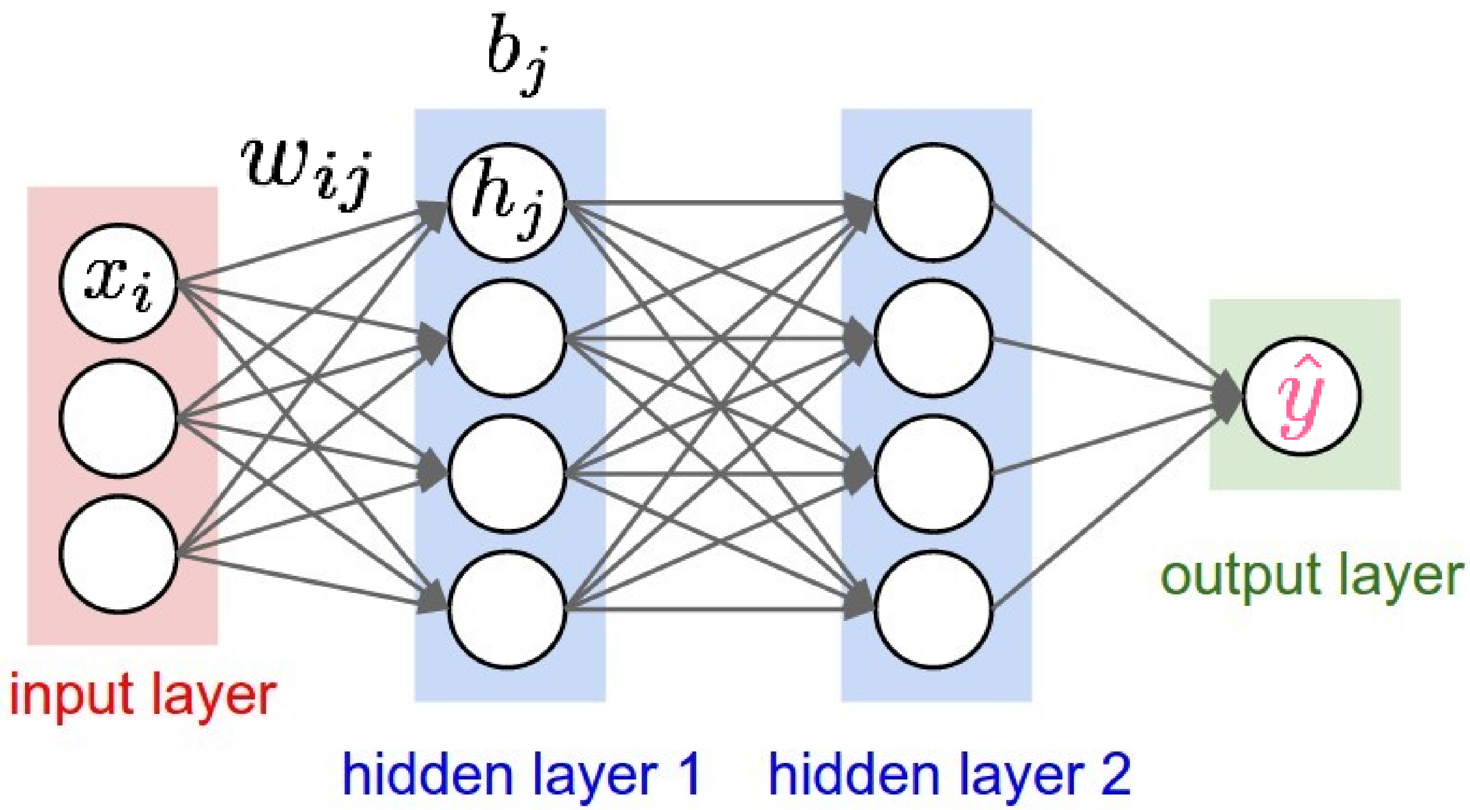
Using concepts learned from similar scenes in the training set.

An example of DL structure :

“hello world” example of deep neural network

Fig from CS231N, Stanford

1, Feed-Forward



$$z_j = \sum_{i=1}^{N} x_i w_{ij} + b_j$$

$$h_j = \sigma(z_j)$$

Linear operations:rotating, boosting,...

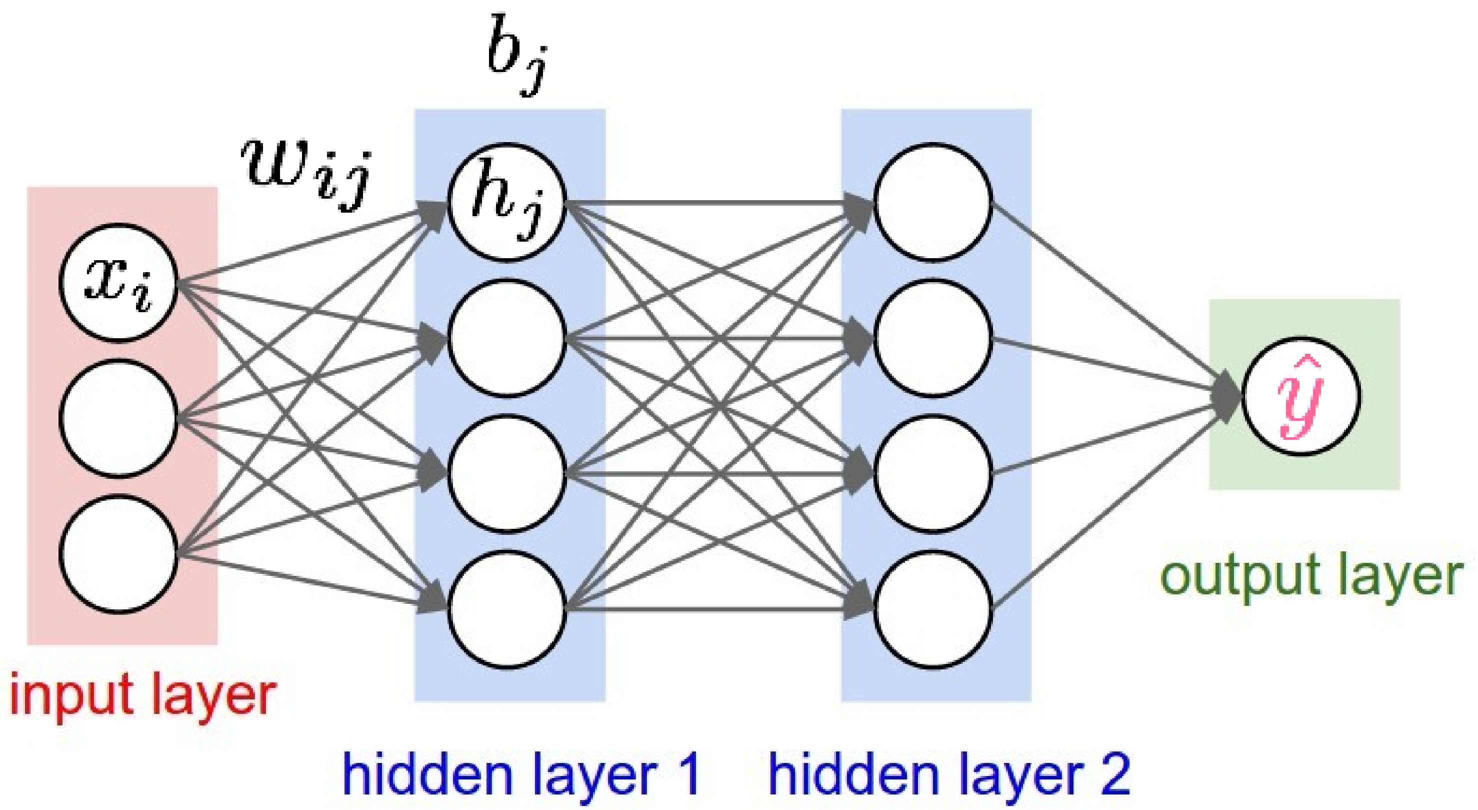
increasing or decreasing dimensions

non-linear activation function: correlation/links

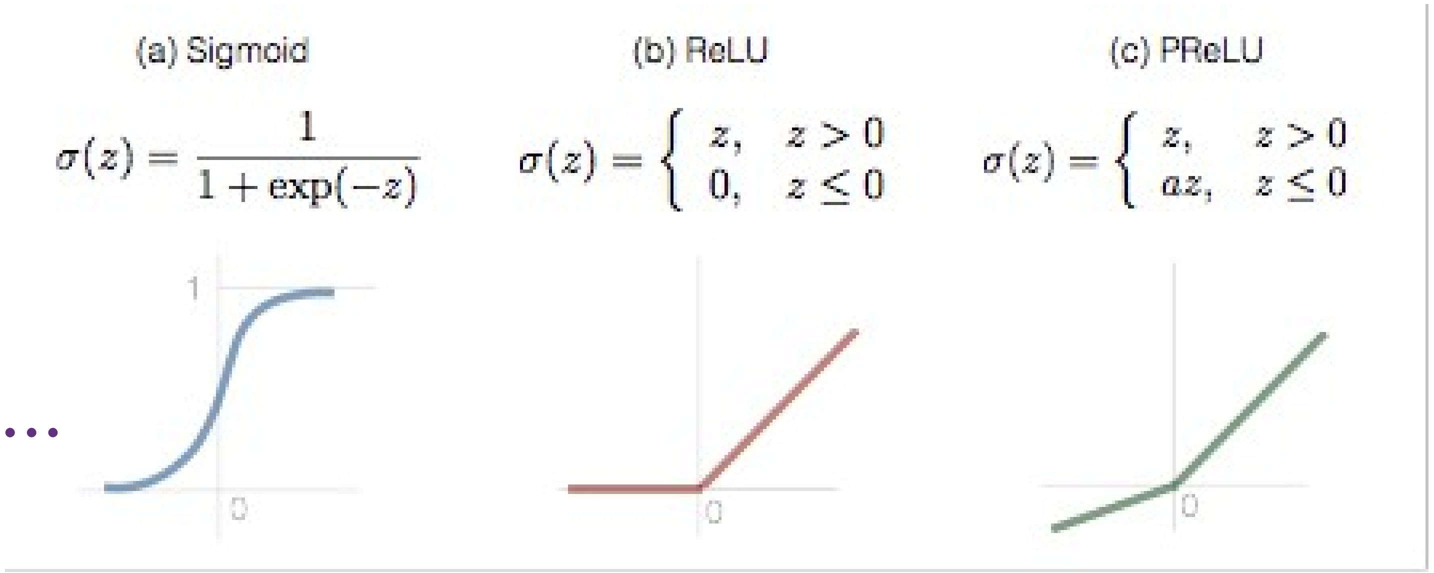
“hello world” example of deep neural network

Fig from CS231N, Stanford

1, Feed-Forward



$$z_j = \sum_{i=1}^{N} x_i w_{ij} + b_j$$



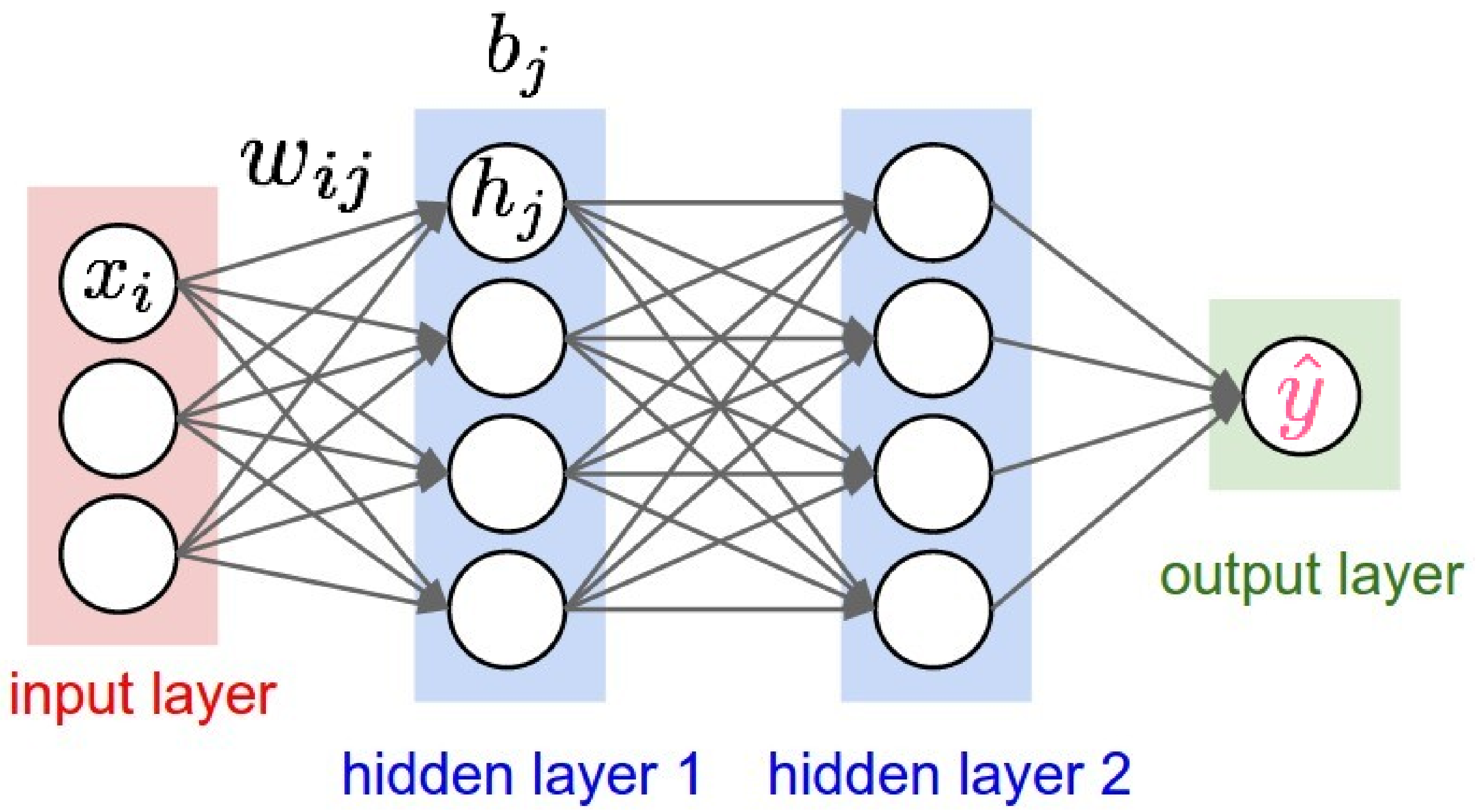
Linear operations:rotating, boosting,...

increasing or decreasing dimensions

“hello world” example of deep neural network

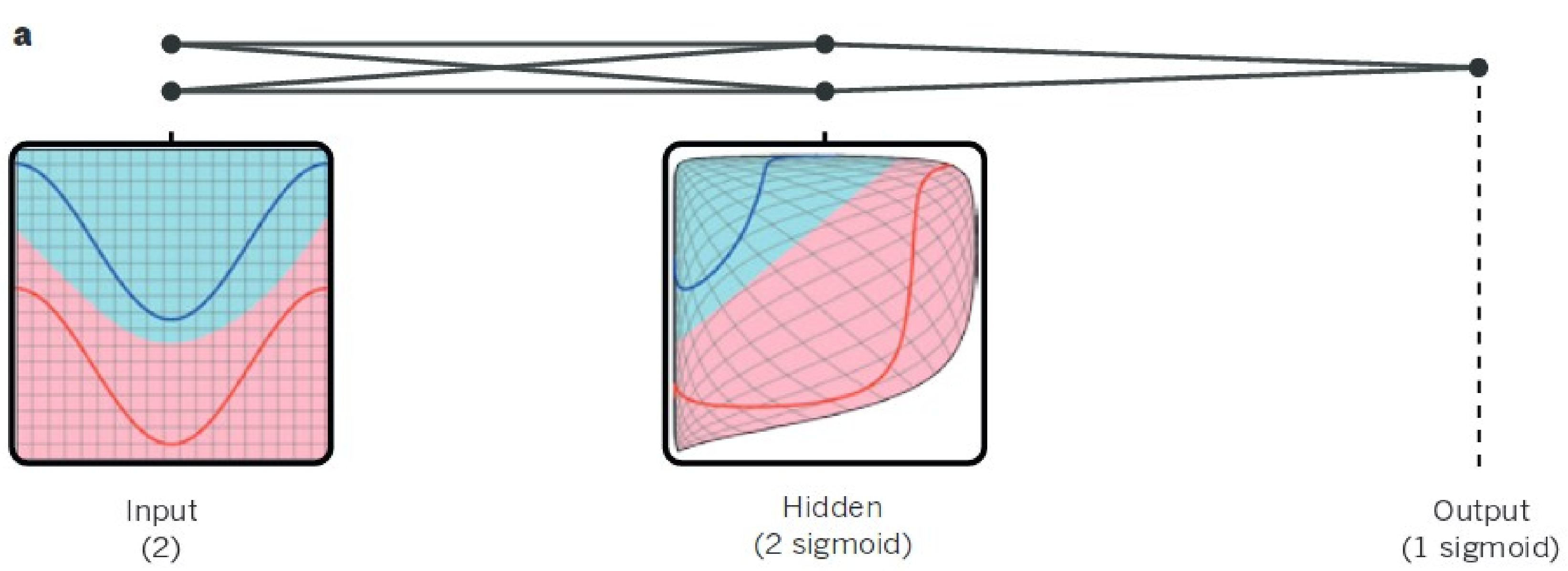
Fig from CS231N, Stanford

1, Feed-Forward



Transformation

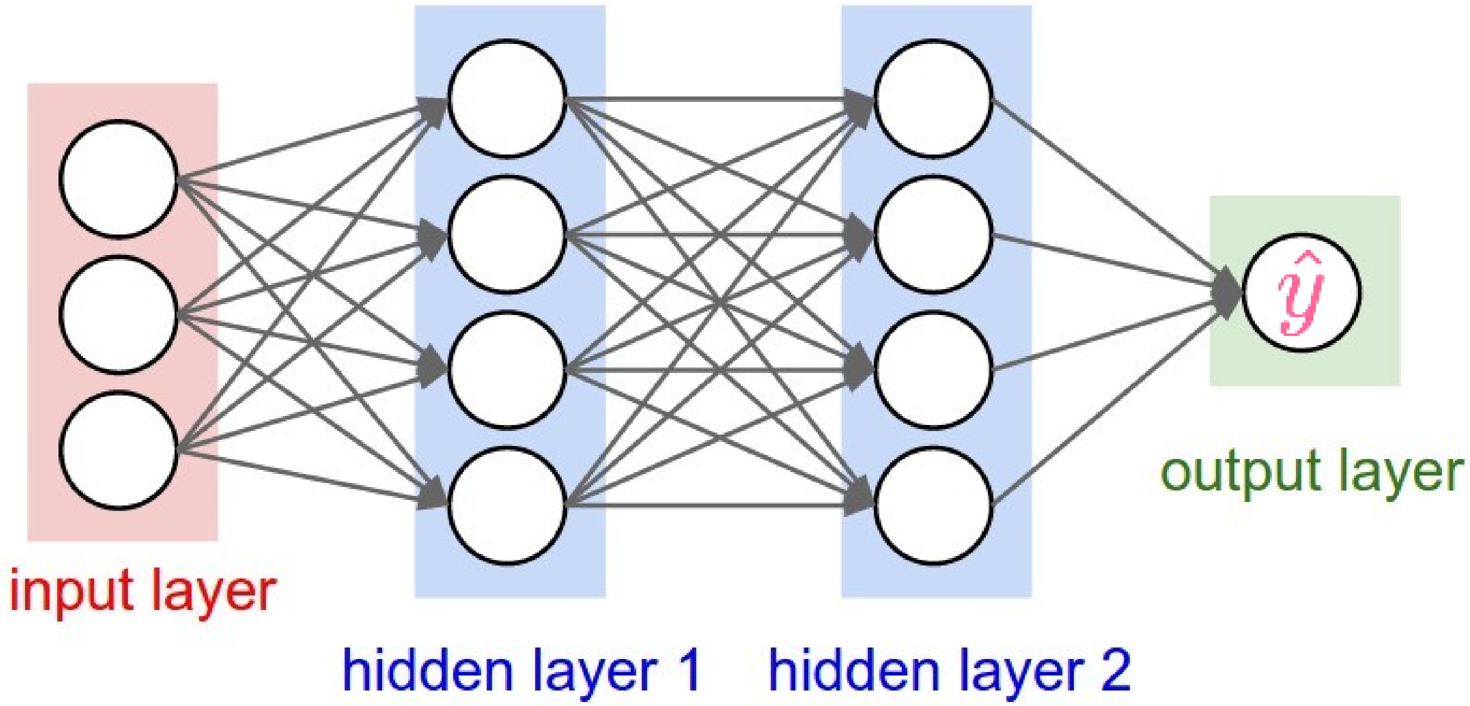
of feature space :



"hello world" example of deep neural network

Fig from CS231N, Stanford

2, Back Propagation



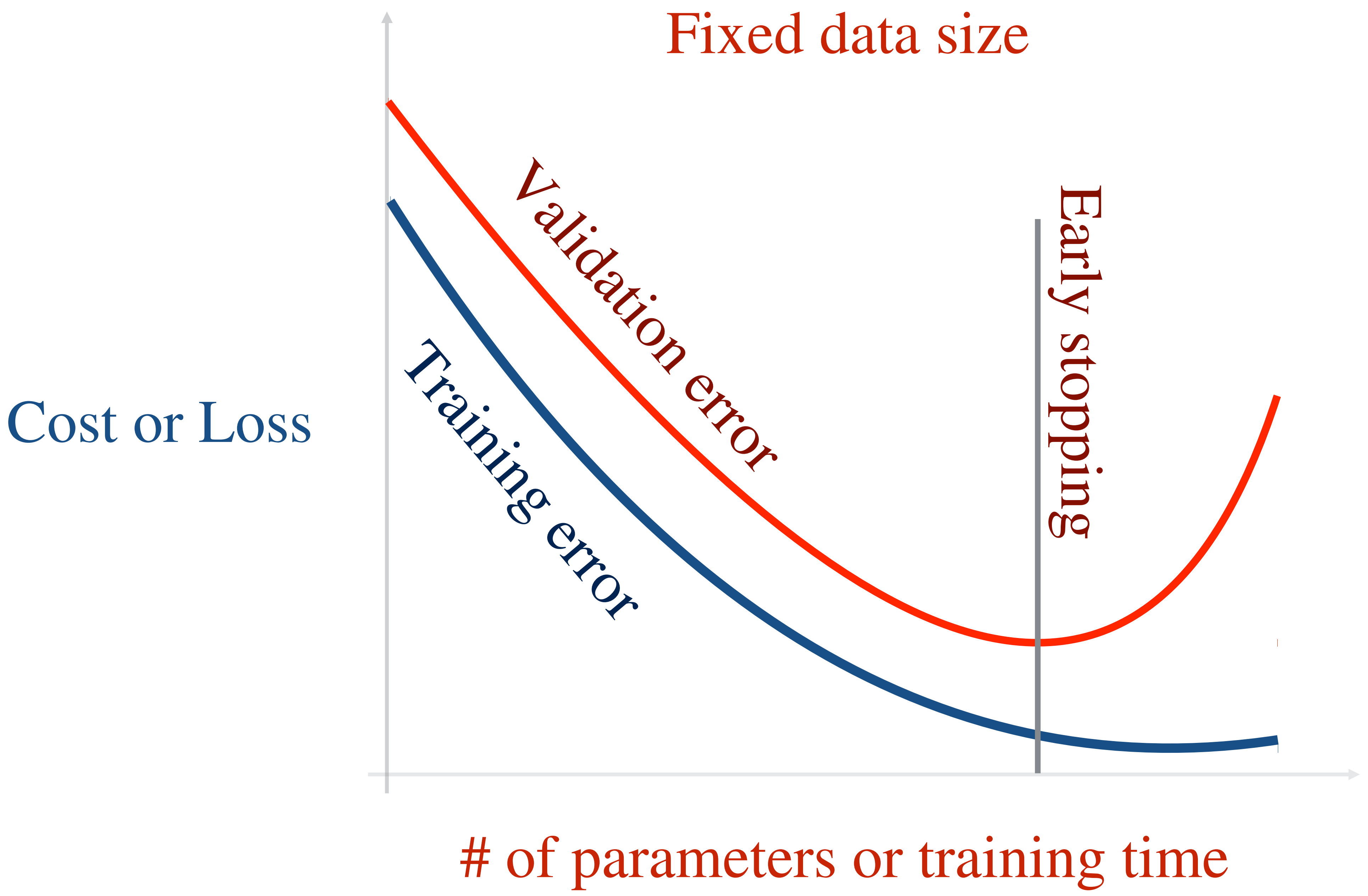
$$l(\theta) = \sum_i (\hat{y}_i - y_i)^2$$

Mean square error (simplest loss function)
with $\hat{y}_i$ the predicted value and y_i true value
$\theta =$ set of all the trainable parameters

$$\theta' = \theta - \epsilon \frac{\partial l(\theta)}{\partial \theta}$$

SGD for parameter-update to minimize loss function

Overfitting problem in fully connected network



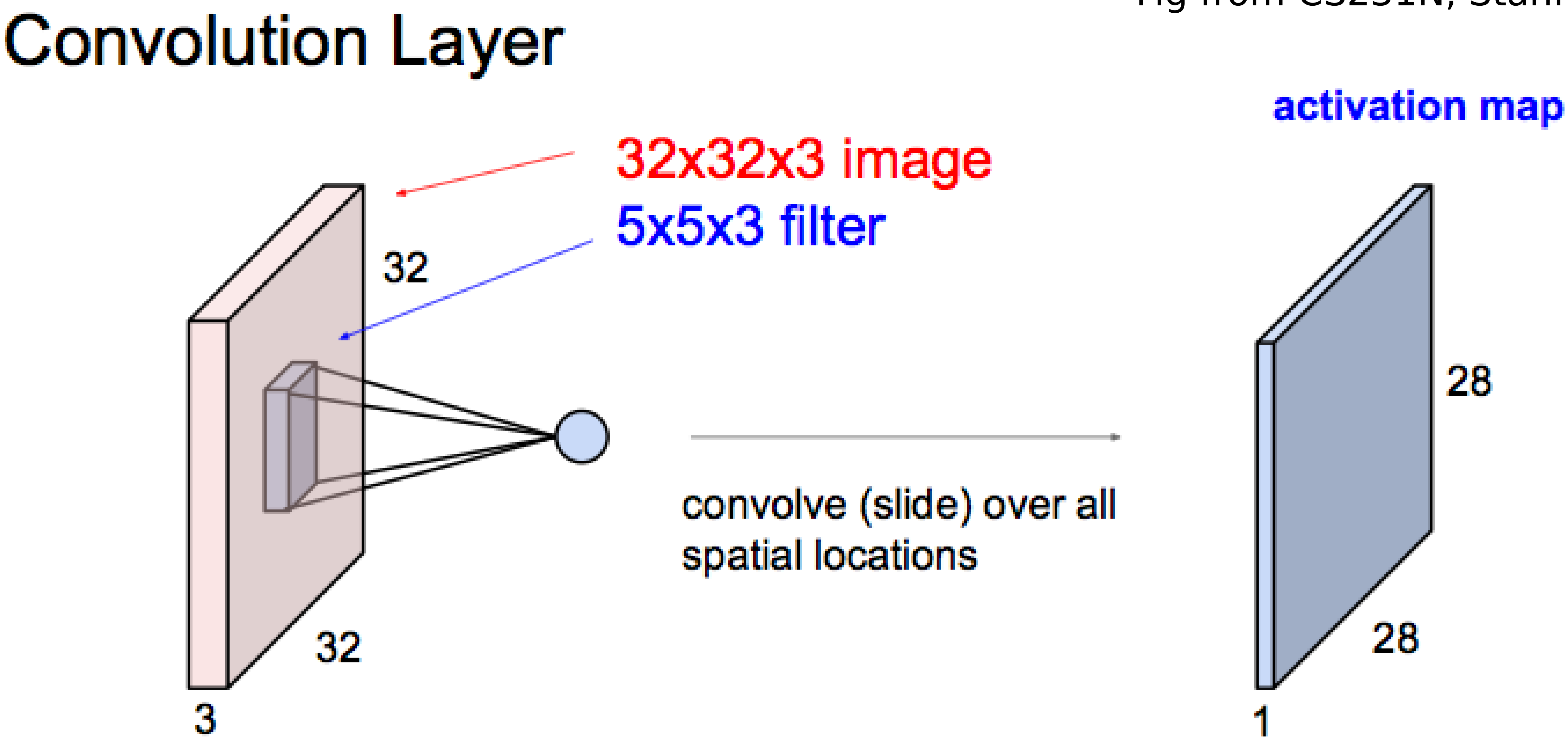
Too many parameters may easily over-fit to training dataset

Ways to reduce overfitting

1. Early stopping
2. Increase training dataset by
   a. preparing more data.
   b. data augmentation (crop, scale, rotate, flip …).
3. Reduce number of parameters
   a. Dropout: randomly discarding neurons.
   b. Drop connection: randomly discarding connections.
   c. CNN: locally connected to a small chunk of neurons in the previous layer.
   d. Go deeper.
4. Regularization, weight decay …

Convolution neural network

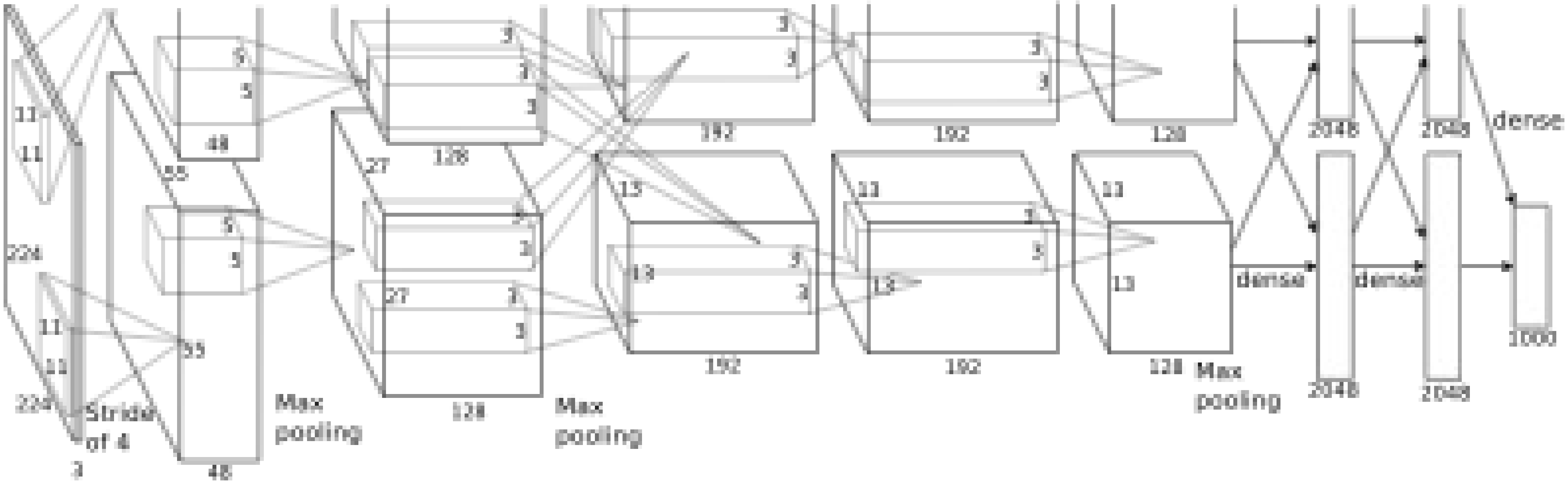
Fig from CS231N, Stanford



Advantage: scaling, rotating, translation invariant features can be learned since only subregion is connected to the filter/kernel which scan the whole input to feel the 2-d structure and local statistics, and Weight shared.

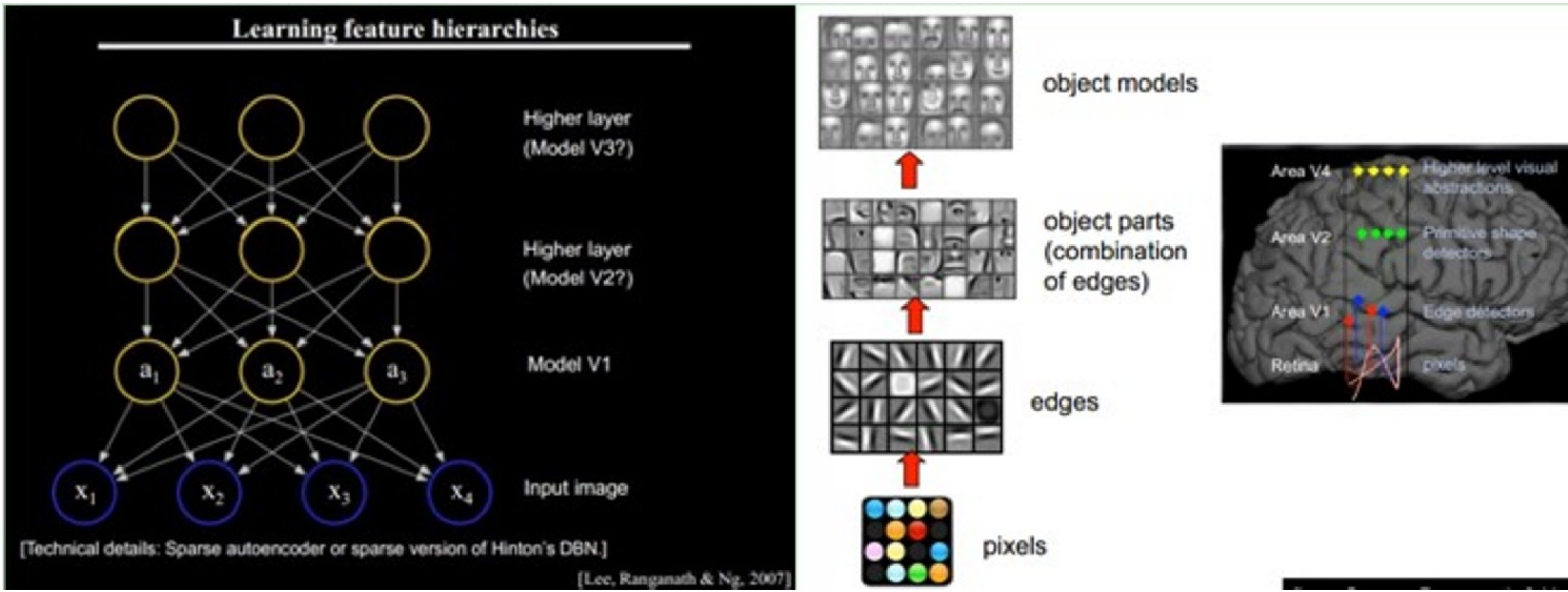
An example for CNN structure

AlexNet, Krizhevsky et al. 2012



- 8 layers, 60 million parameters
- Data augmentation and Dropout are used to reduce overfitting, first use ReLU
- Removing any hidden layer results in 2% loss.

What CNN did after the training?



Distributed representations are learned !

Open Source Libraries

Keras + TensorFlow in the present study

Keras is a high level neural network library, written in Python and capable of running on top of either TensorFlow or Theano.

```
# Build one fully connected neural network (100->64->10 neurons) in Keras

from keras.models import Sequential
from keras.layers import Dense, Activation

model = Sequential()
model.add(Dense(output_dim=64, input_dim=100))
model.add(Activation("relu"))
model.add(Dense(output_dim=10))
model.add(Activation(“softmax"))
model.compile(loss='categorical_crossentropy', optimizer='sgd',
metrics=['accuracy'])
```

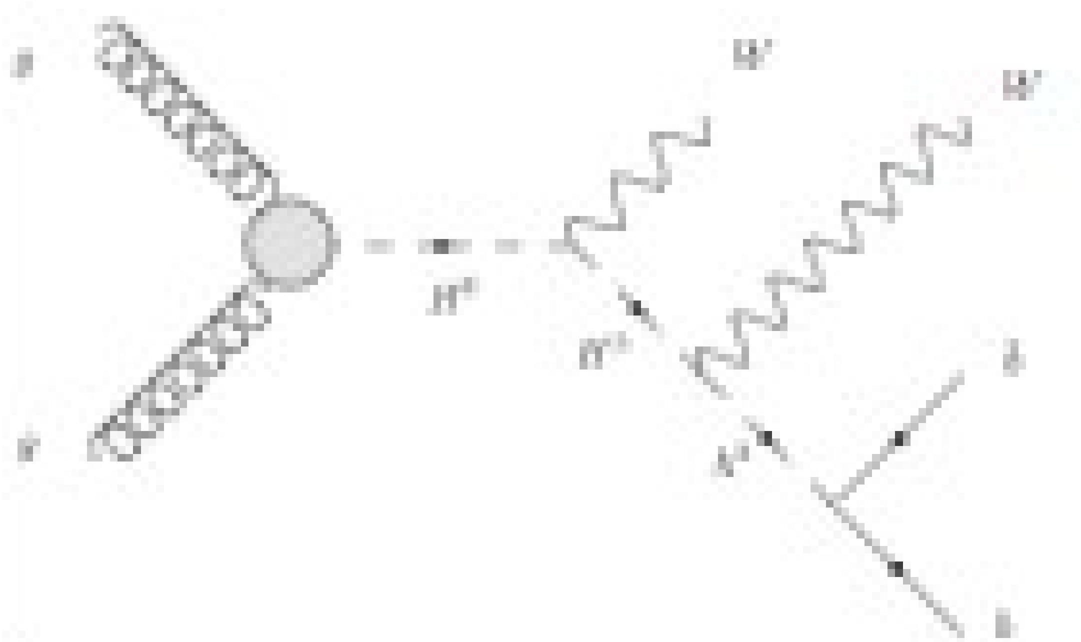
Deep learning in Physics (Particle Physics)

Searching for Exotic Particles in High-Energy Physics with Deep Learning

P. Baldi, P. Sadowski, and D. Whiteson

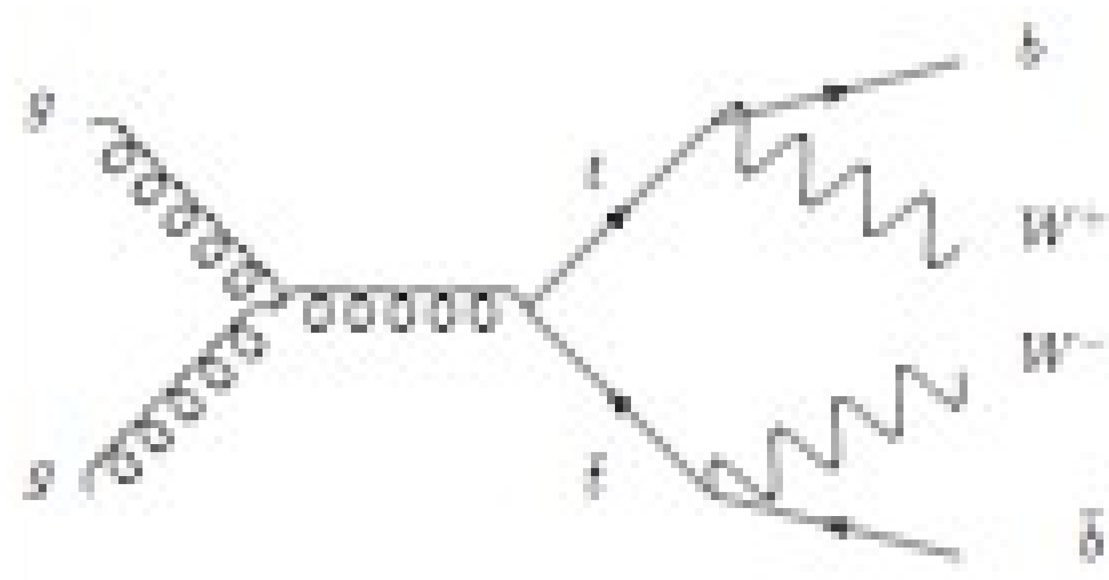
Nature Communications

Higgs benchmark



(a)

Signal



(b)

Background

Deep learning in Physics (HeavyIon Jet)

P. Baldi, K. Bauer, C. Eng, P. Sadowski, and D. Whiteson, *Jet Substructure Classification in High-Energy Physics with Deep Neural Networks*, *Phys. Rev.* **D93** (2016), no. 9 094034, [arXiv:1603.09349].

D. Guest, J. Collado, P. Baldi, S.-C. Hsu, G. Urban, and D. Whiteson, *Jet Flavor Classification in High-Energy Physics with Deep Neural Networks*, arXiv:1607.08633.

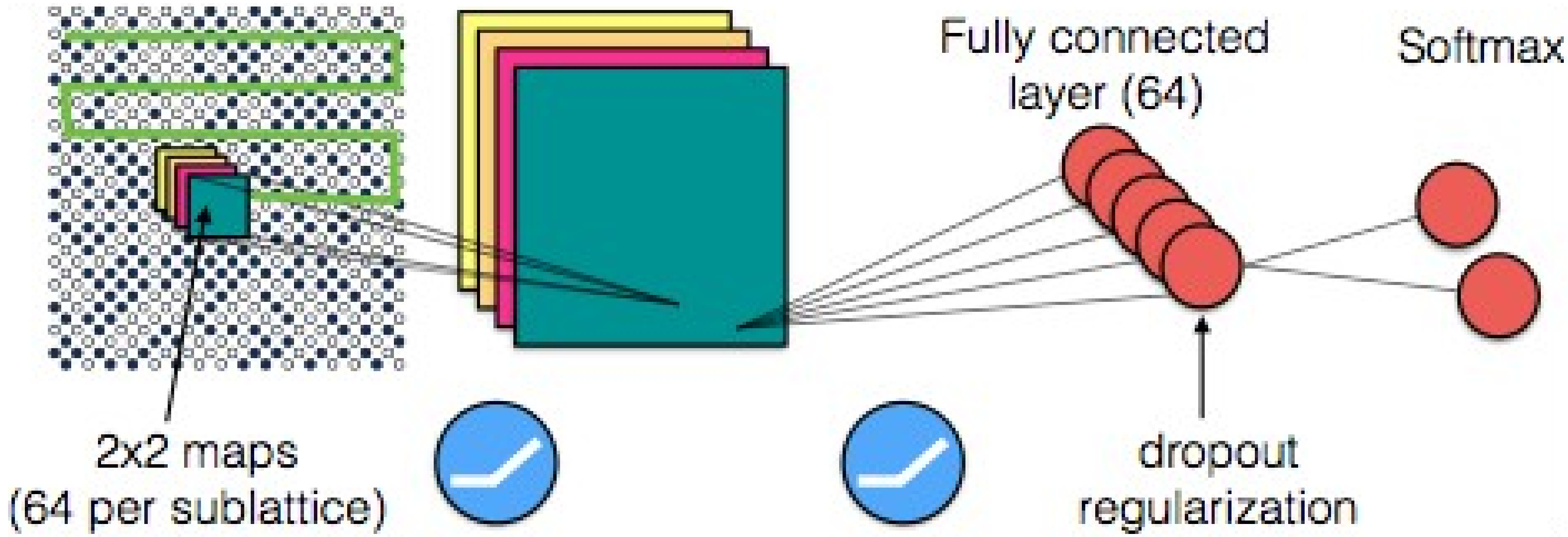
J. S. Conway, R. Bhaskar, R. D. Erbacher, and J. Pilot, *Identification of High-Momentum Top Quarks, Higgs Bosons, and W and Z Bosons Using Boosted Event Shapes*, arXiv:1606.06859.

J. Barnard, E. N. Dawe, M. J. Dolan, and N. Rajcic, *Parton Shower Uncertainties in Jet Substructure Analyses with Deep Neural Networks*, arXiv:1609.00607.

Deep learning in Physics (Cond-Mat Ising)

Machine learning phases of matter

Juan Carrasquilla[1] and Roger G. Melko[2,1]



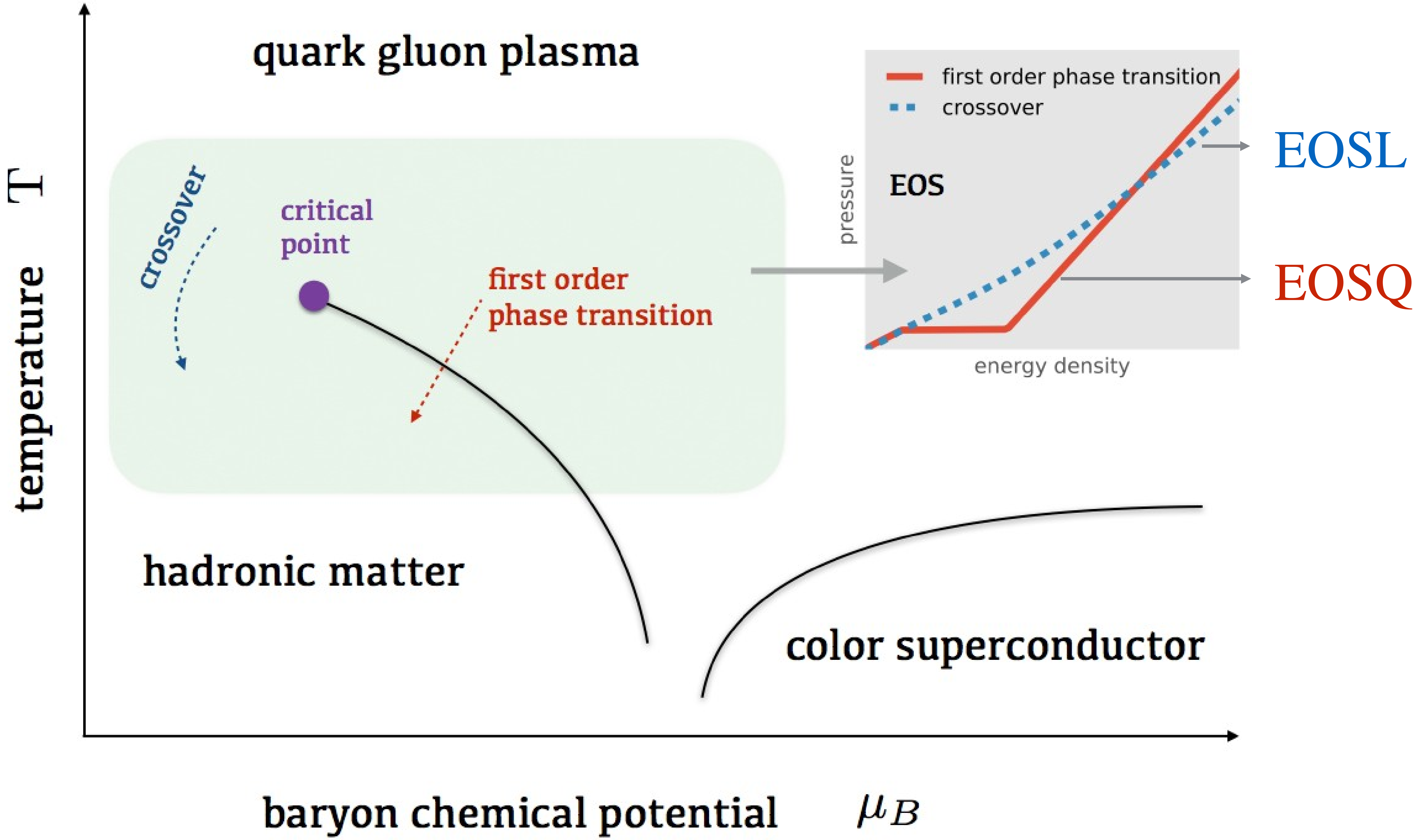
Extract order parameter from the spin configurations,

can encode phases and discriminate Phase Transition for

spin, Coulomb and even topological phases.

Identifying QCD transition in HIC using DL

QCD transition: 1st order or crossover?



Uncertainties in the simulation of HIC

Initial stage

Pre-Equilibrium dynamics?
Fast thermalization/tau0?
Initial entropy deposition?
Baryon stopping?
Initial state fluctuations and flow?

$$\nabla_\mu T^{\mu\nu} = 0$$
$$\nabla_\mu N^\mu = 0$$

EoS and transport coefficients
(shear viscosity, bulk viscosity)

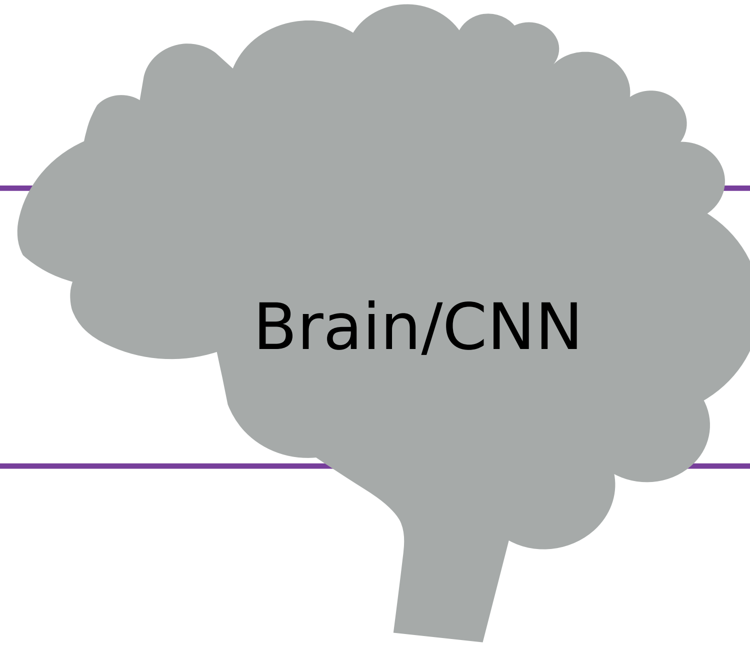
$$p^\mu \partial_\mu f + F \cdot \partial_p f = C$$

some unfixed scattering cross sections

Particle momentum distribution

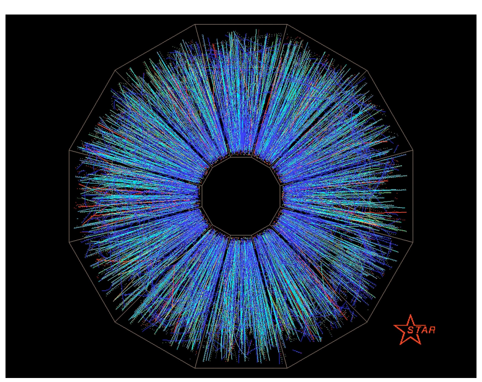
Big uncertainty when comparing with Exp. data

New perspective --- Deep Learning

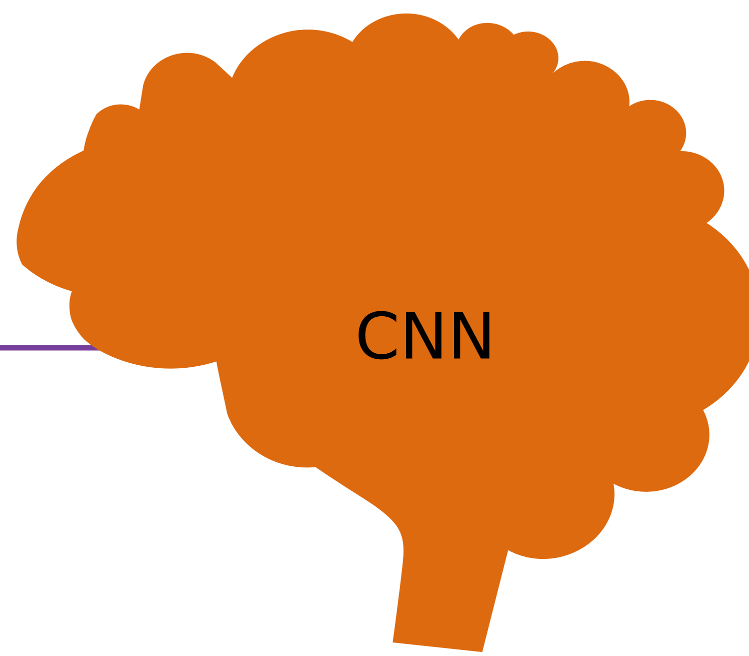


Dog

Cat



$\rho(p_T, \Phi)$



Crossover (0,1)

?

1st order (1,0)

Key idea for our prototype study

Supervised learning using deep convolution neural network with huge amount of labeled training data (spectra, EoS type) from event-by-event relativistic hydrodynamic simulations.

Training dataset

$\rho(p_T, \Phi)$ for charged pions at mid-rapidity $\rho(p_T, \Phi) \equiv \frac{dN_i}{dY p_T dp_T d\Phi} = g_i \int_\sigma p^\mu d\sigma_\mu f_i$

TRAINING DATASET	$\eta/s = 0$		$\eta/s = 0.08$	
	EOSL	EOSQ	EOSL	EOSQ
Au-Au $\sqrt{s_{NN}} = 200\,\mathrm{GeV}$	7435	5328	500	500
Pb-Pb $\sqrt{s_{NN}} = 2.76\,\mathrm{TeV}$	4967	2828	500	500

- CLVisc: (3+1D viscous hydrodynamic program parallelized on GPU using OpenCL) with AMPT initial condition (~22000 events, doubled by left-right flipping, 10% for validation).

- τ_0 is 0.4 fm for Au+Au and 0.2 fm for Pb+Pb collisions

- T_dec=0.137 GeV

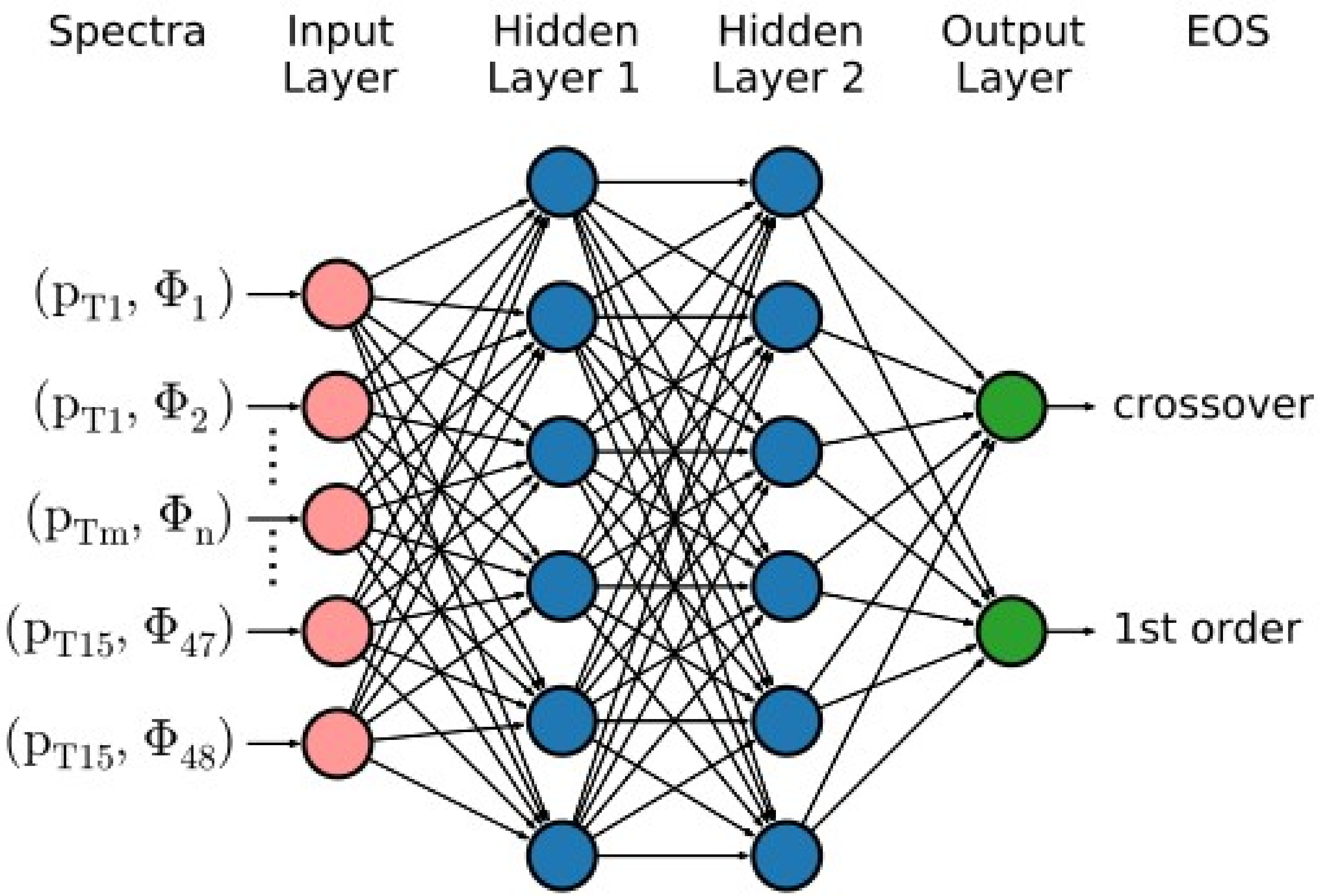
Testing dataset

TESTING DATASET GROUP 1 : iEBE-VISHNU + MC-Glauber						
Centrality:	$\eta/s=0$		$\eta/s=0.08$		$\eta/s=0.16$	
10-60%	EOSL	EOSQ	EOSL	EOSQ	EOSL	EOSQ
Au-Au $\sqrt{s_{NN}}=200\,\mathrm{GeV}$	300	250	250	150	200	250
Pb-Pb $\sqrt{s_{NN}}=2.76\,\mathrm{TeV}$	500	650	200	195	350	400

TESTING DATASET GROUP 2 : CLVisc + IP-Glasma		
Au-Au $\sqrt{s_{NN}}=200\,\mathrm{GeV}$	EOSL	EOSQ
b$\lesssim$8 fm & $\eta/s=0$	4165	4752

- iEBE-VISHNU: another viscous hydro package with different numerical solver for the partial differential equations and with different initial conditions, eta/s.
- τ_0 is 0.6 fm for all the testing dataset.
- T_dec in [0.115GeV, 0.142GeV] for iEBE-VISHNU

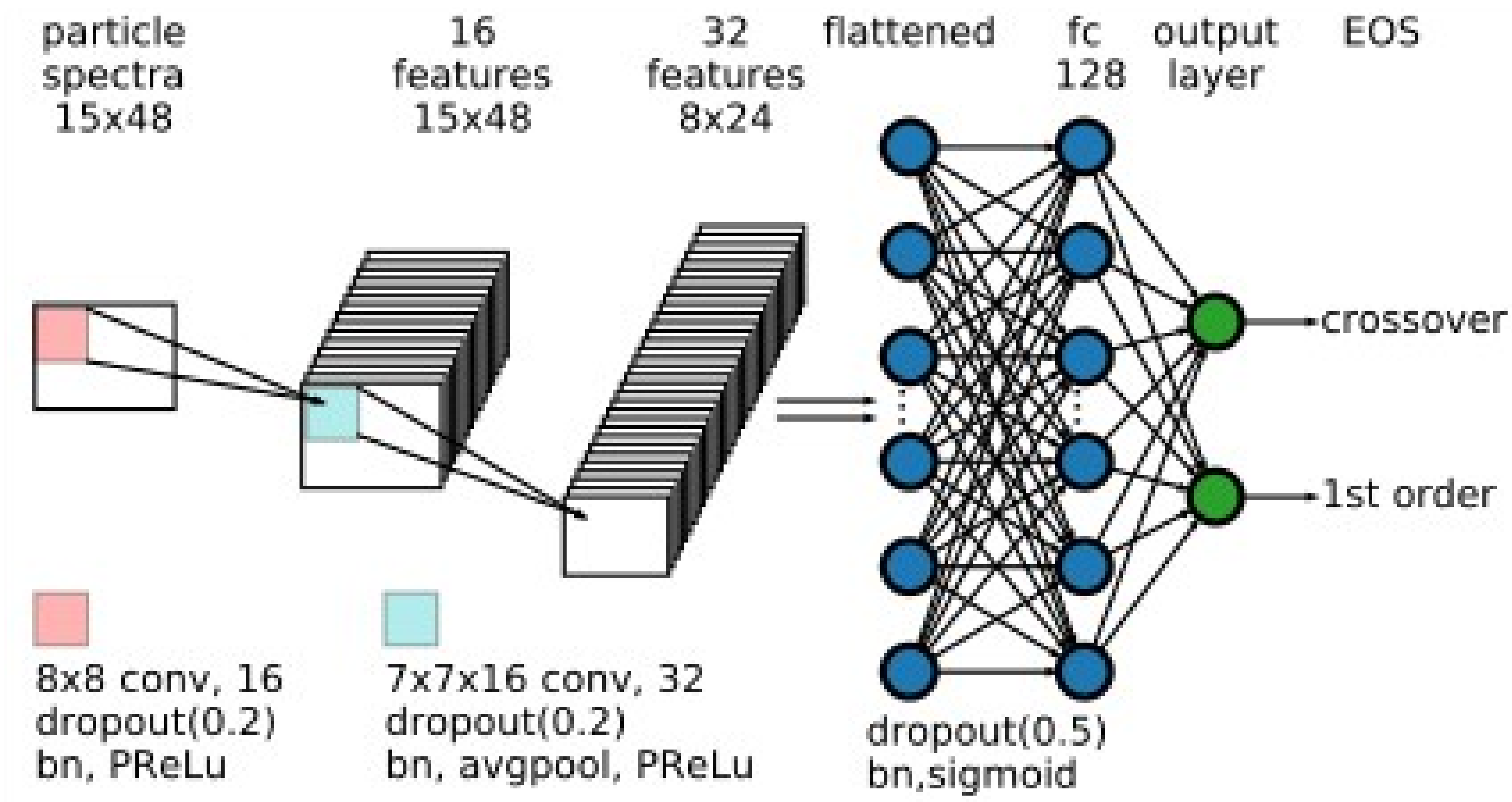
First try with fully connected neural network



Overfit to the training dataset! Does not work for testing dataset.

------ no generalization.

CNN architecture for EoS-meter



Batch normalization, Dropout, L2 Rregularization, PreLU are used to prevent the overfitting.

Train 500 epochs, in mini-batch with size=64
Learning rate: 0.0001, decay with rate: 1.0E-6

Results

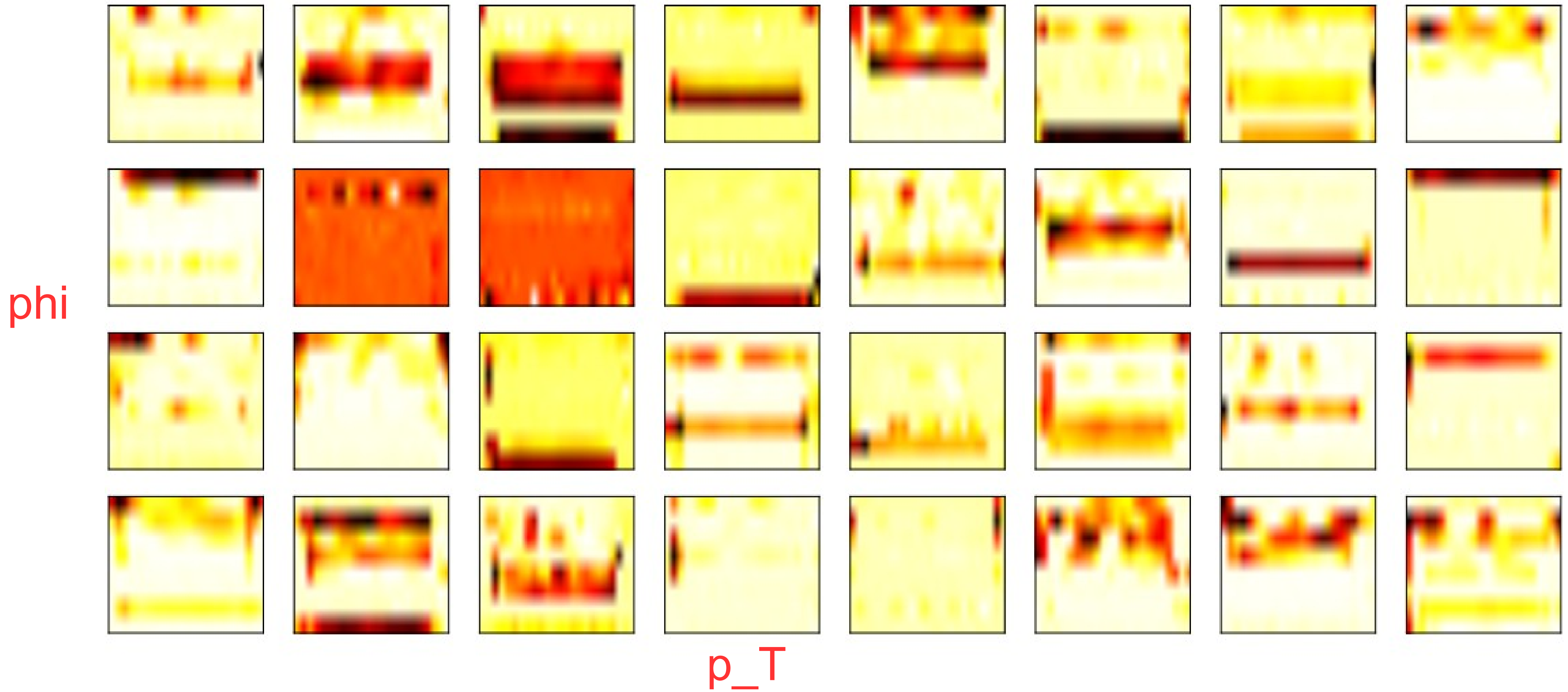
TESTING ACCURACIES	GROUP 1		GROUP 2	
	EOSL	EOSQ	EOSL	EOSQ
Number of events	1800	1895	4164	4752
Accuracy	95.1%	95.8%	96.4%	96.4%

~95% prediction accuracy on avg. in independent testing dataset solely from the raw spectra

The performance is **robust against** : initial conditions/fluctuations, tao_0, eta/s, T_dec

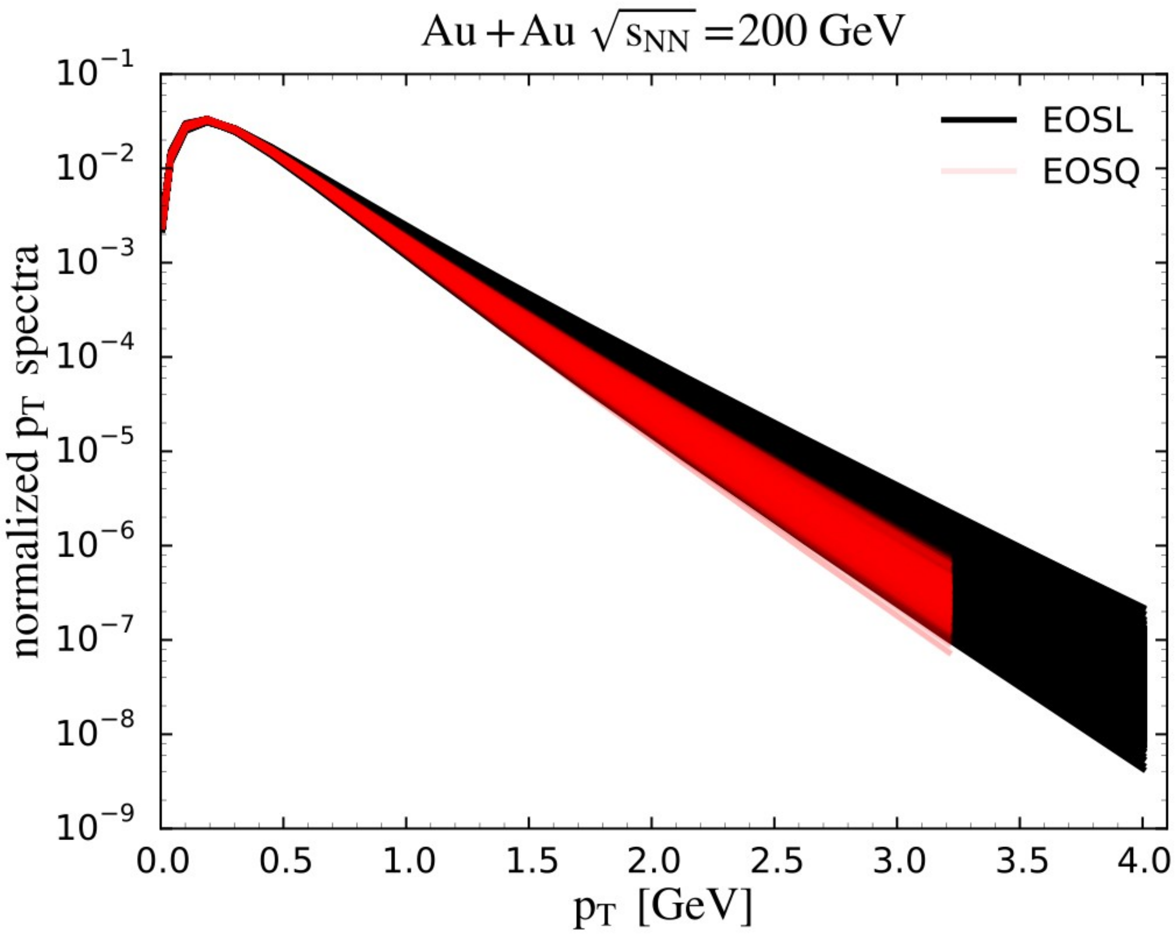
---- model independent (general rule is learned)

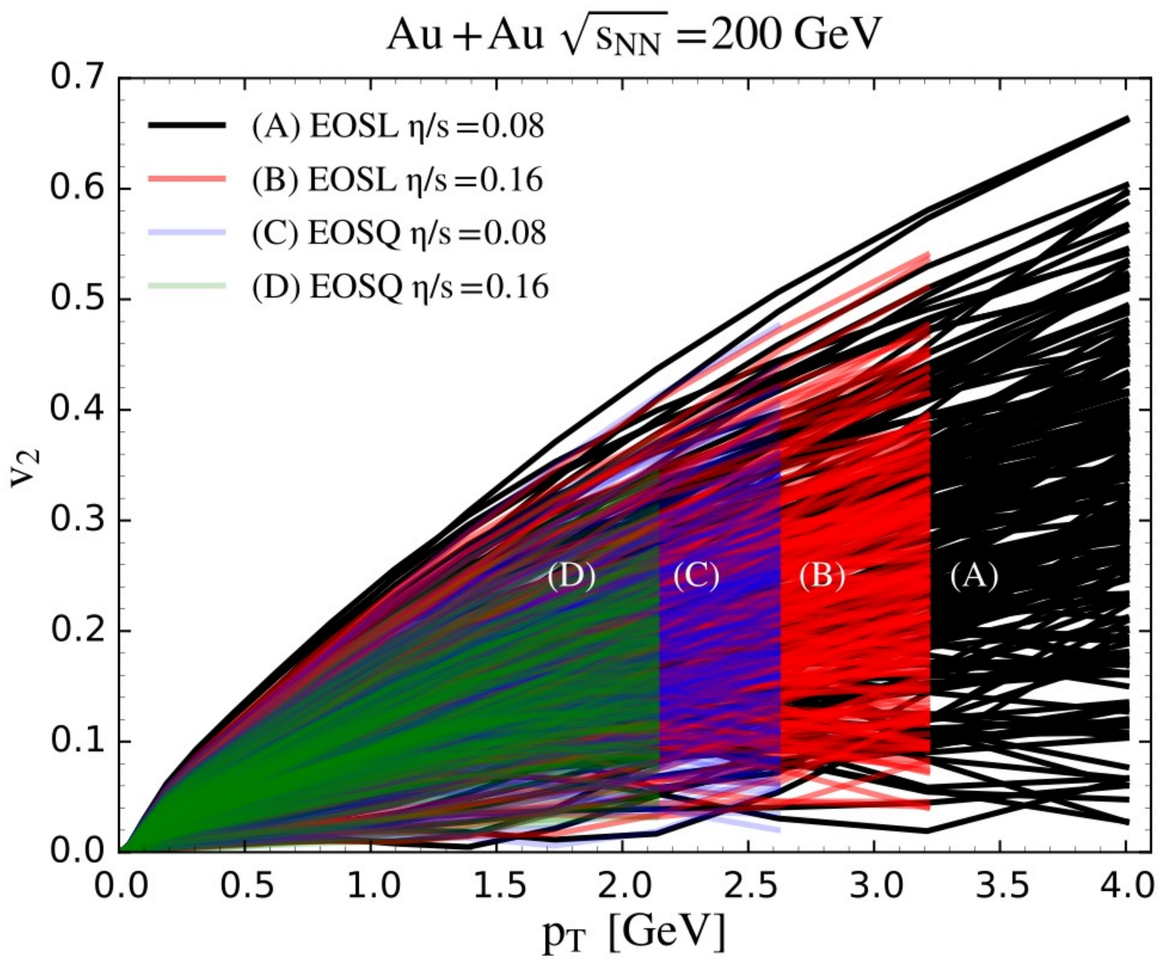
EoS-meter from CNN



- Each hotspot represents correlations in a small phase space (32 features in the 2nd convolution layer)!
- Correlation of these features from the next fully connected layer

Conventional observables





- Strongly depends on initial stage fluctuations and other parameters

CNN provides novel perspective in connecting QCD theory with HIC directly :

- There do exist 'mapping/**Encoders**/projection' from QCD transition onto the final state raw spectra, although they are not intuitive to conventional interpretation yet.

  ---They are clean (robust to other uncertainties and parameters)

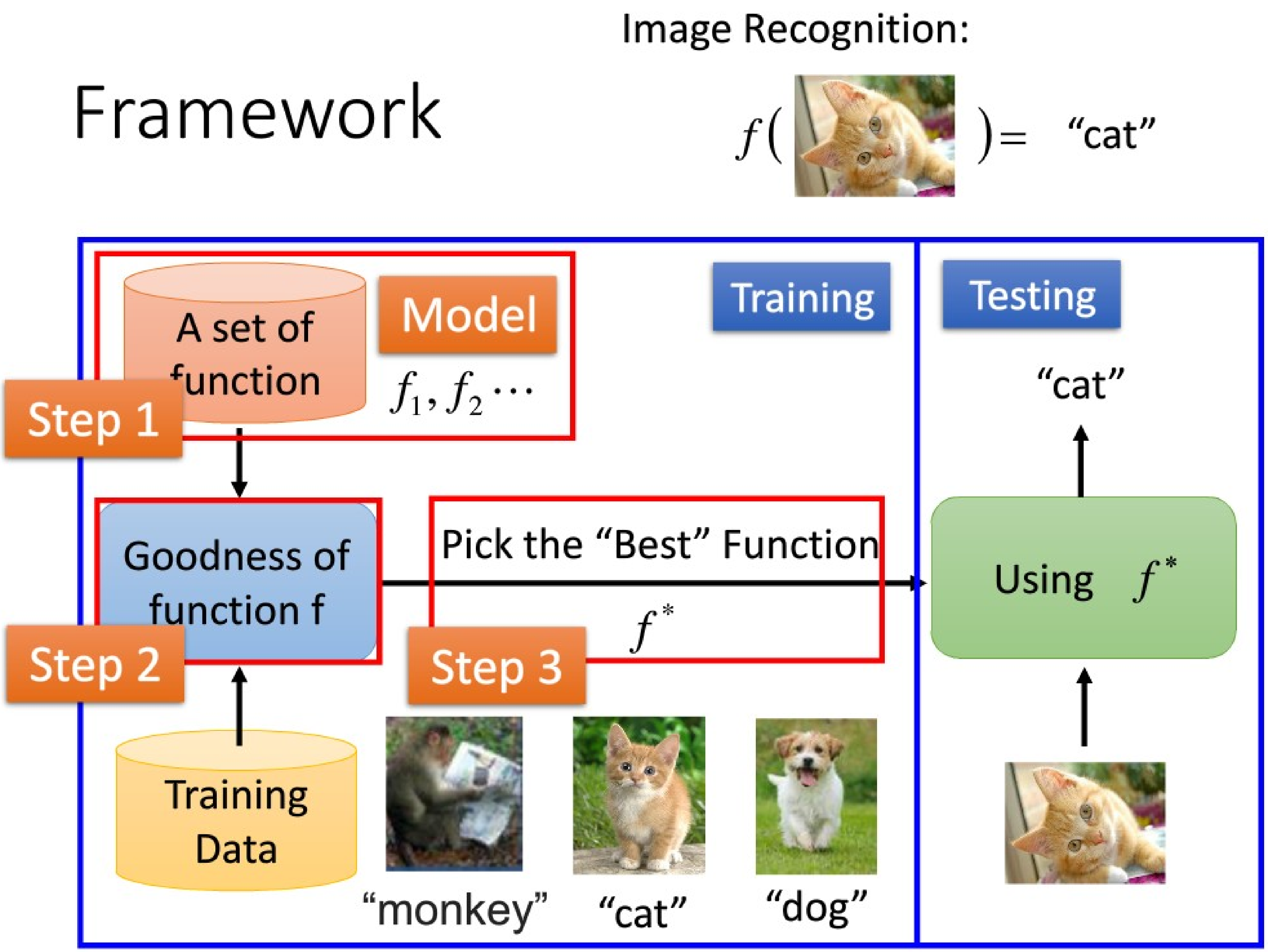
- The deep CNN can provide a powerful and efficient '**Decoder**' for the above Encoder/mapping

  ------the high-level representations act as '**EoS-meter**'

Outlook

- Try to find out the underlying mechanism (guided back-propagation)
- Extend the model to work with real Exp. data
- Extract other dynamical parameters like eta/s.

What DL is doing?



Slides by Long-gang