SU(3)非可換ゲージ理論のトポロジー分類への機械学習の適用

北沢正清, 松本拓也, 河野泰宏

(大阪大学)

T. Matsumoto, M. Kitazawa, Y. Kohno, arXiv:1909.06238

理論物理学と機械学習

機械学習は理論物理学に応用可能か？

- MLはblack box.「理解」にはならない？
- MLにできることは、より洗練した方法で実現可能？
- 100%の正答率は得られない。

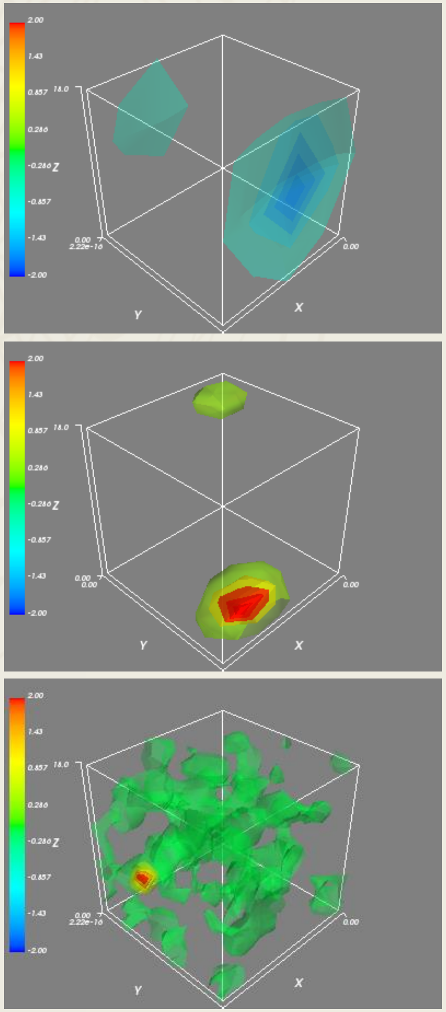
Topological Charge in YM Theory

$$Q = \int d^4x q(x) \quad \text{: integer}$$

$$q(x) = -\frac{1}{32\pi^2}\mathrm{tr}[F_{\mu\nu}\tilde{F}_{\mu\nu}]$$

in continuum theory

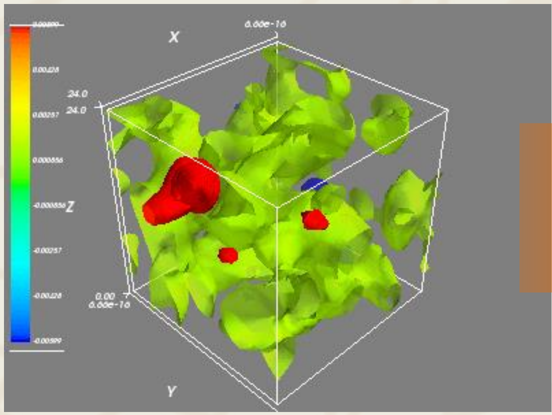
q(x) in SU(3) YM



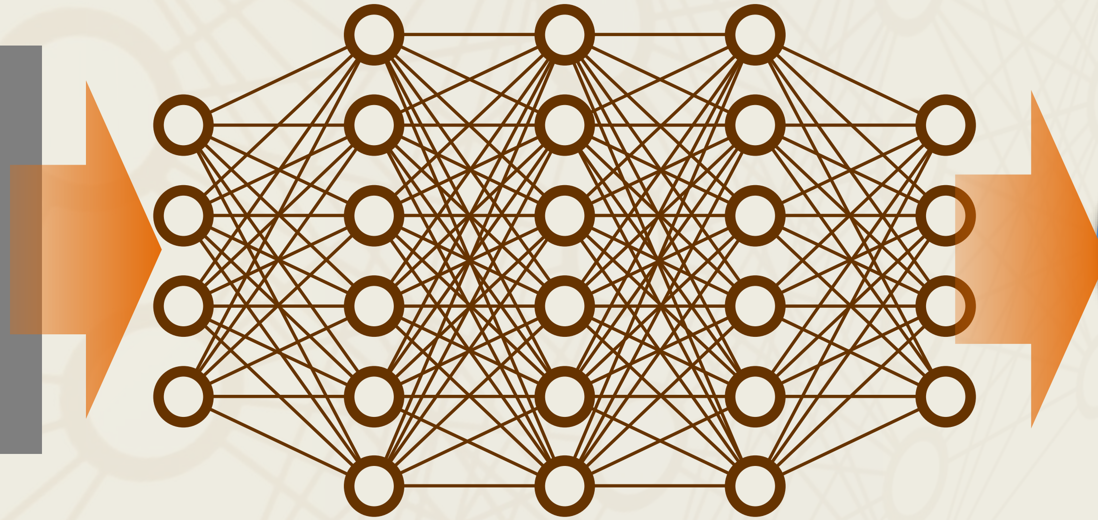
- **Interests / applications**
  - Instantons
  - Axial U(1) anomaly
  - Axion cosmology
  - Topological freezing

Machine Learning

Input: q(x)



4-dimensional field



Output

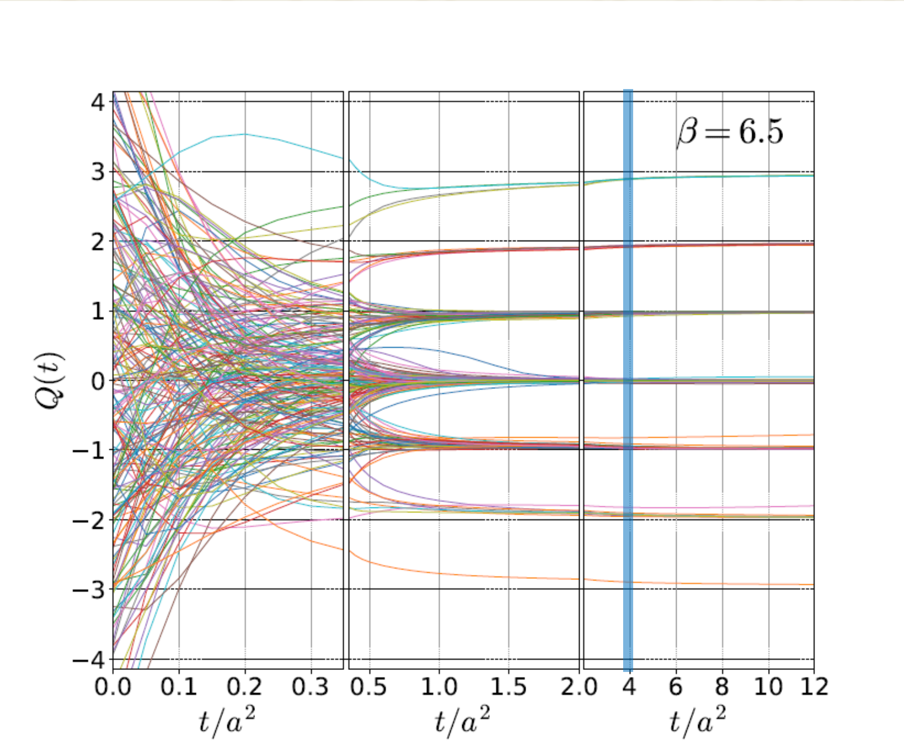
topological charge

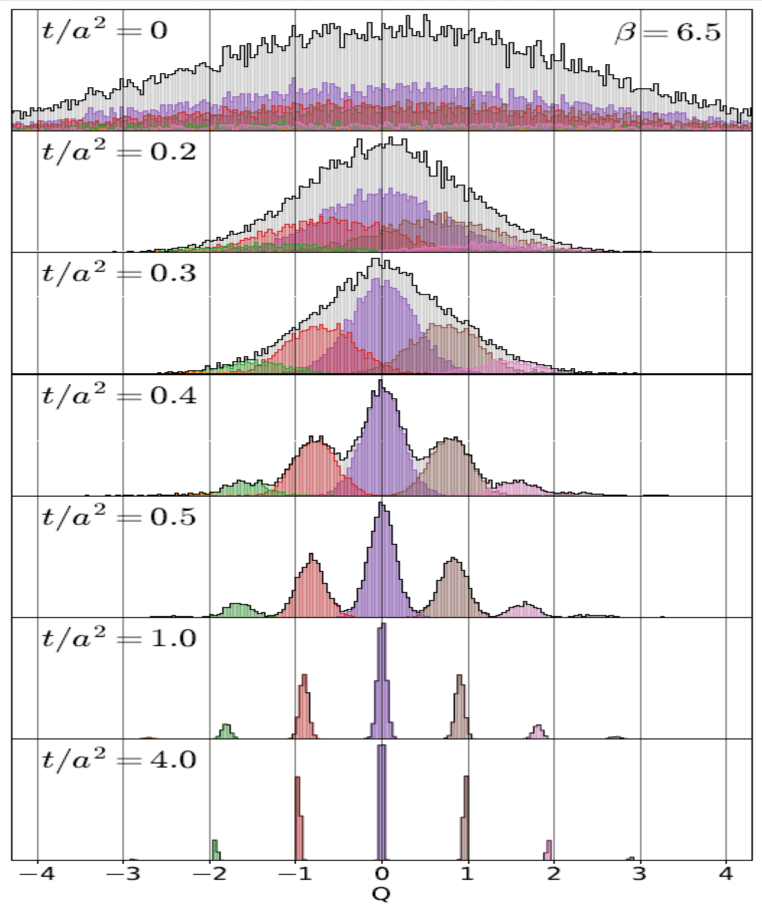
- **Analysis of 4D data by 4D-CNN**
- **Capture “instanton”-like structure?**
- **Acceleration of the analysis of Q?**

Gradient Flow and Topology

gradient flow ≃ Smoothing

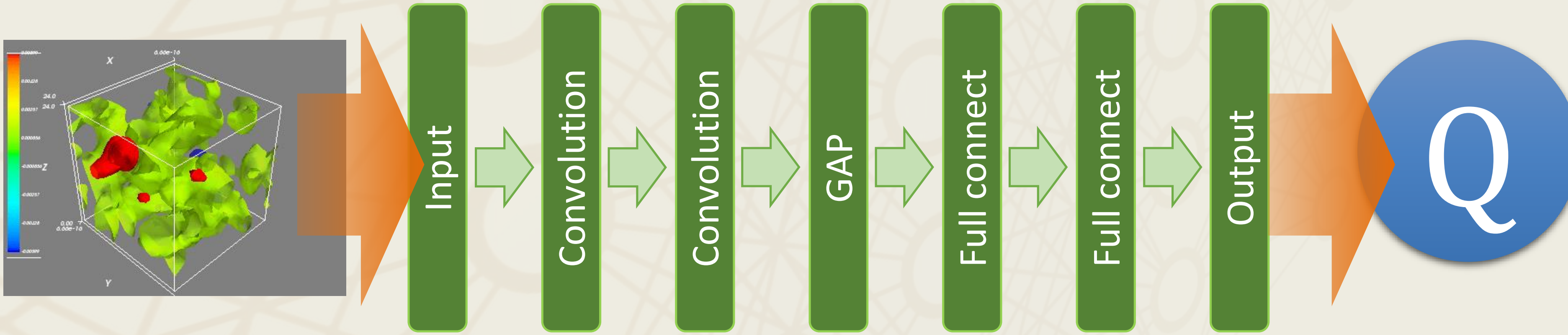
flow time t = smearing radius





Trial 2: Topol. Density @ t>0

- Input: q(x,t) in 4-dim space at nonzero flow time
- Data reduction to 8^4 (average pooling)



Recalls of each topological sector (%)

$N_{\rm ch}$	input t/a^2	P	R_Q								
			-4	-3	-2	-1	0	1	2	3	4
1	0	0.371	0	0	0	0	1.000	0	0	0	0
1	0.1	0.401	0	0	0.002	0.255	0.702	0.341	0.008	0	0
1	0.2	0.552	0	0.043	0.240	0.495	0.687	0.597	0.336	0.111	0
1	0.3	0.776	0	0.391	0.687	0.760	0.821	0.794	0.740	0.569	0
3	0.3,0.2,0.1	0.942	0.200	0.913	0.944	0.950	0.944	0.939	0.937	0.889	0.571

Benchmark

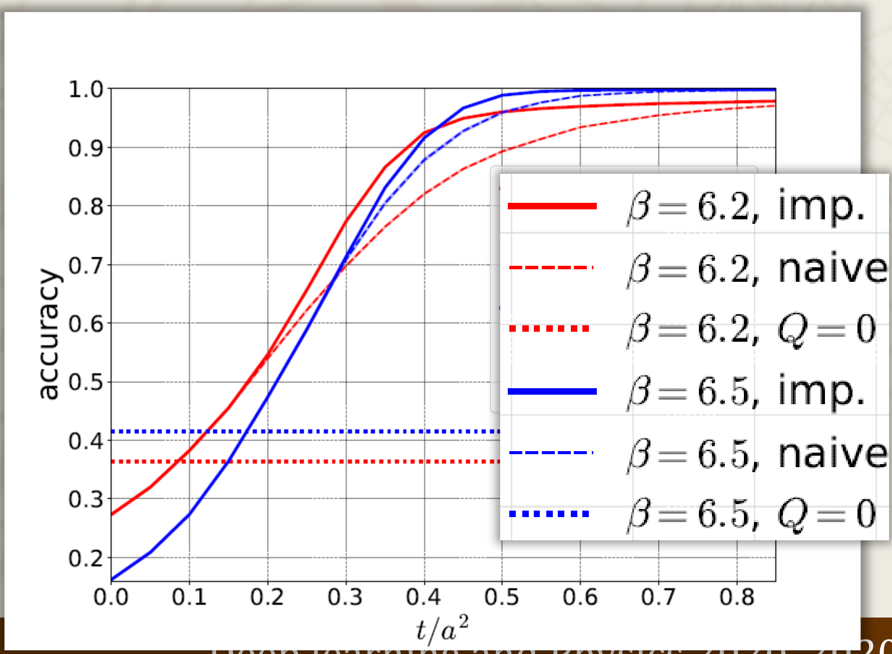
Simple estimator from Q(t)

1) Naïve: $Q = \mathrm{round}[Q(t)]$

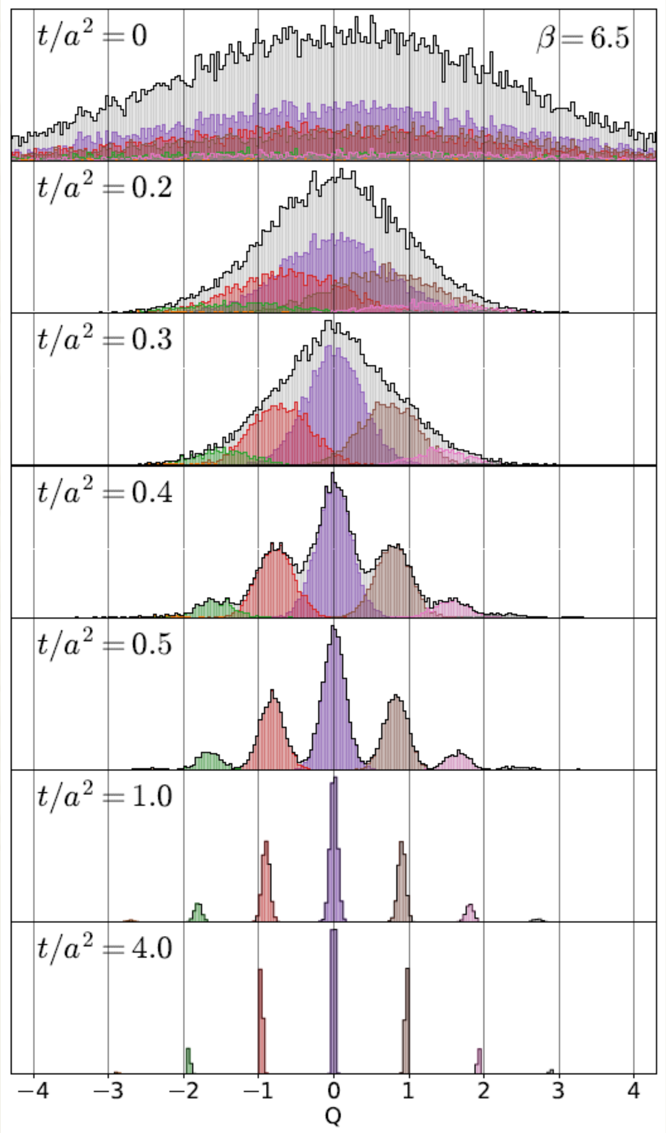
2) Improved: $Q = \mathrm{round}[cQ(t)]$

c>1: optimization param.

3) zero: $Q = 0$



Distribution of Q(t)



Comparison: NN vs Benchmark

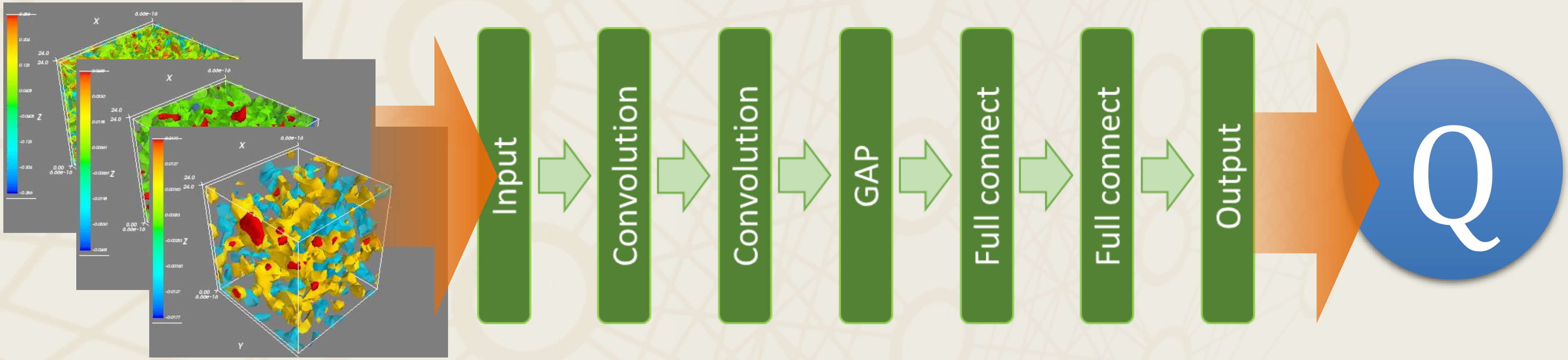
accuracy at β=6.2

	ML (Trial 2)	naïve	improved
$t/a^2=0$	**37.0**	27.3	**27.3**
$t/a^2=0.1$	**40.3**	38.3	**38.3**
$t/a^2=0.2$	**53.7**	54.0	**54.6**
$t/a^2=0.3$	**76.1**	69.8	**77.3**

- Machine learning cannot exceed the benchmark value.
- NN would be trained to answer the "improved" value.
- **No useful local structures found by the NN.**

Trial 3: Multi-Channel Analysis

- Input: q(x,t) in four-dimensional space **at t/a^2=0.1, 0.2, 0.3**

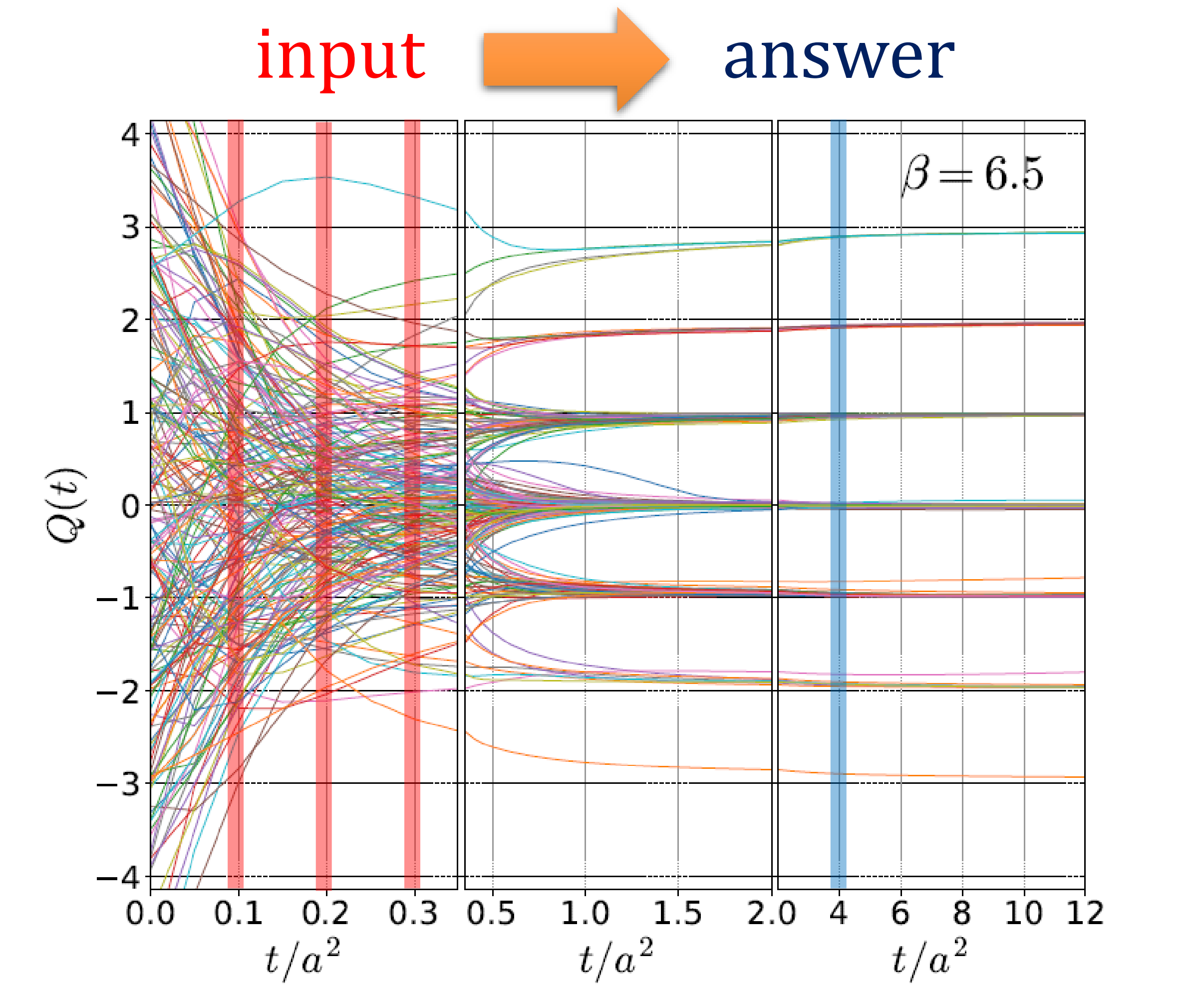


- **Result**

	machine learning	benchmark @ t/a^2=0.3
β=6.2	**93.8**	77.3
β=6.5	**94.1**	71.3

  - **non-trivial improvement from the benchmark!!**

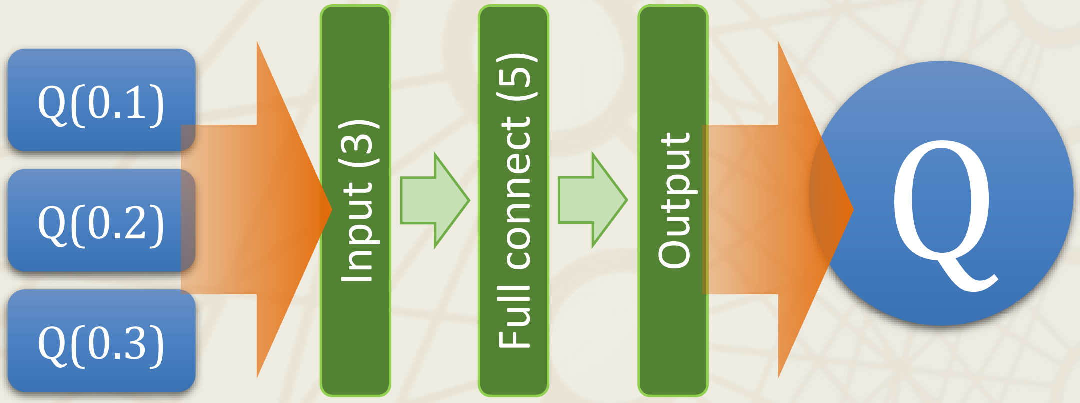
Is this a non-trivial result?

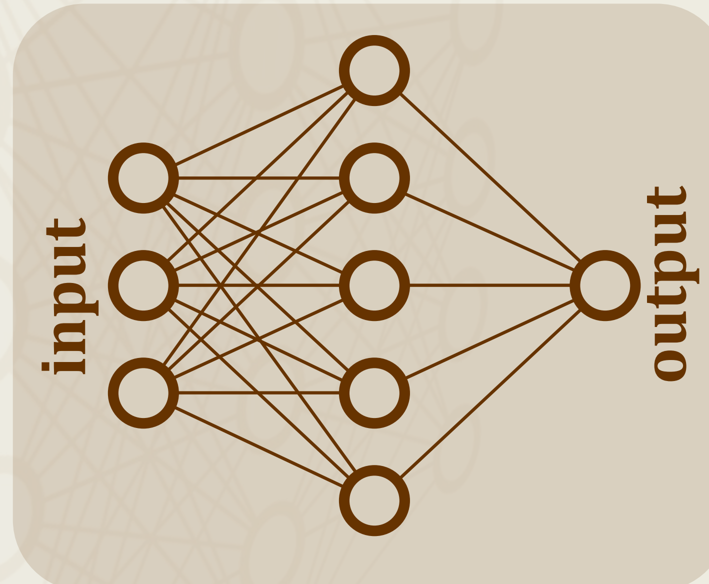


We can estimate the answer from Q(t) by our eyes…

Trial 4: Feed Q(t) [0-dim]

- Input: Q(t) at $t/a^2=0.1, 0.2, 0.3$





- **Result**

	Q(t)	**tuning input *t***	Trial 3 (4dim)	benchmark
β=6.2	**94.1**	**95.9**	93.8	77.3
β=6.5	**95.7**	**99.0**	94.1	71.3

- Good accuracy is obtained only from Q(t)

Reducing the Training Data

- Smaller training data will reduce numerical cost for the training.

Training data	10,000	5,000	1,000	500	100
β=6.2	**95.9(2)**	**95.9(2)**	**95.9(2)**	95.5(3)	90.3(7)
β=6.5	**99.0(2)**	**99.0(2)**	**98.9(2)**	98.9(1)	90.2(8)

- **1000 configurations are enough** to train the NN successfully!
- Numerical cost for the training is small.

Robustness Test

- Analyze configurations with a different parameter set

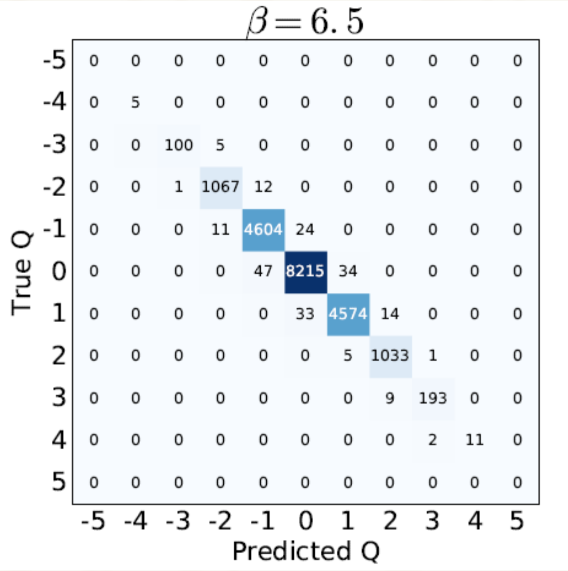
analyzed data

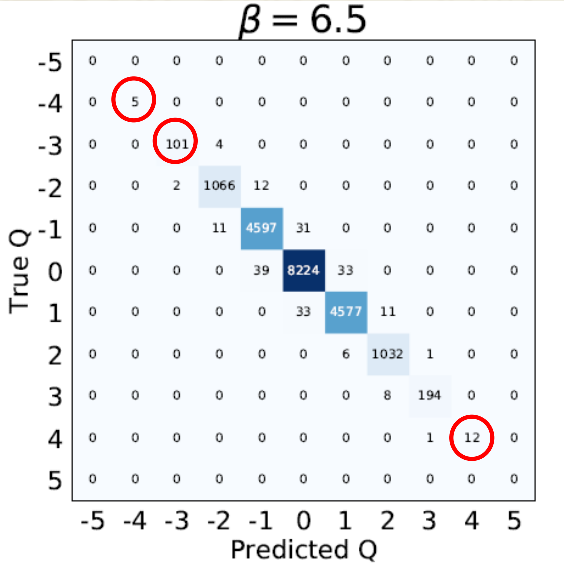
training data	β=6.2	β=6.5
β=6.2	95.9(2)	98.6(2)
β=6.5	95.6(2)	99.0(2)
both	**95.8(1)**	**98.9(2)**

- NNs trained for β=6.2 and 6.5 can be used for another parameter successfully.
- **Universal NN would be developed!**
- Note: same physical volume

Untrained Data

standard training | **training w/o |Q|=4,5**





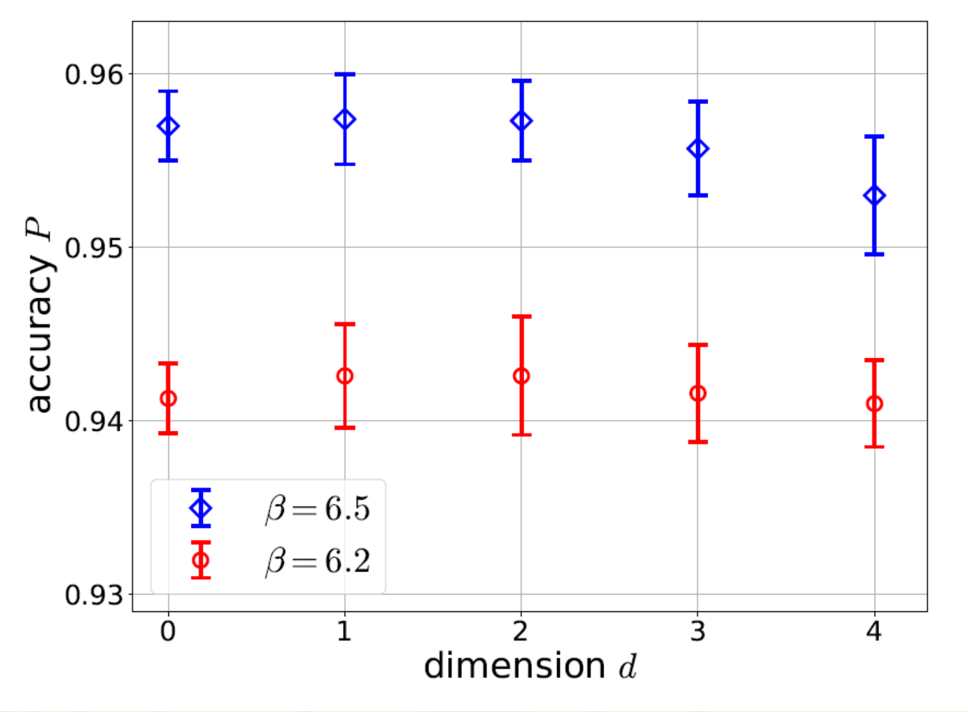
- CNN can make accurate predictions even for untrained values of Q.

Trial 5: Dimensional Reduction

- Optimal dimension between d=0 and 4?
- d-dimensional CNN
- Input: $q_d(x)$ after dimensional reduction
- 3-channel analysis: t/a^2=0.1, 0.2, 0.3

$$q_3(x, y, z) = \int d\tau q(x)$$

$$q_2(x, y) = \int d\tau dz q(x)$$



- No *d* dependence
- Failed in finding features in multi-dim. space.
- No instanton-like local structure in QCD vacuum?

Summary and Outlook

- 4次元データのCNN解析は、配位の局所的構造を特徴抽出するには至っていないようである。
- 次元縮約も精度向上に寄与しない。

- **ゼロ次元FNN解析は、トポロジカル電荷推定の効率化に有効。Gradient flowはt/a^2=0.3までで良い。**
- 高い正答率／少ない学習データ／高い堅牢性／未知のデータへの適用

将来展望

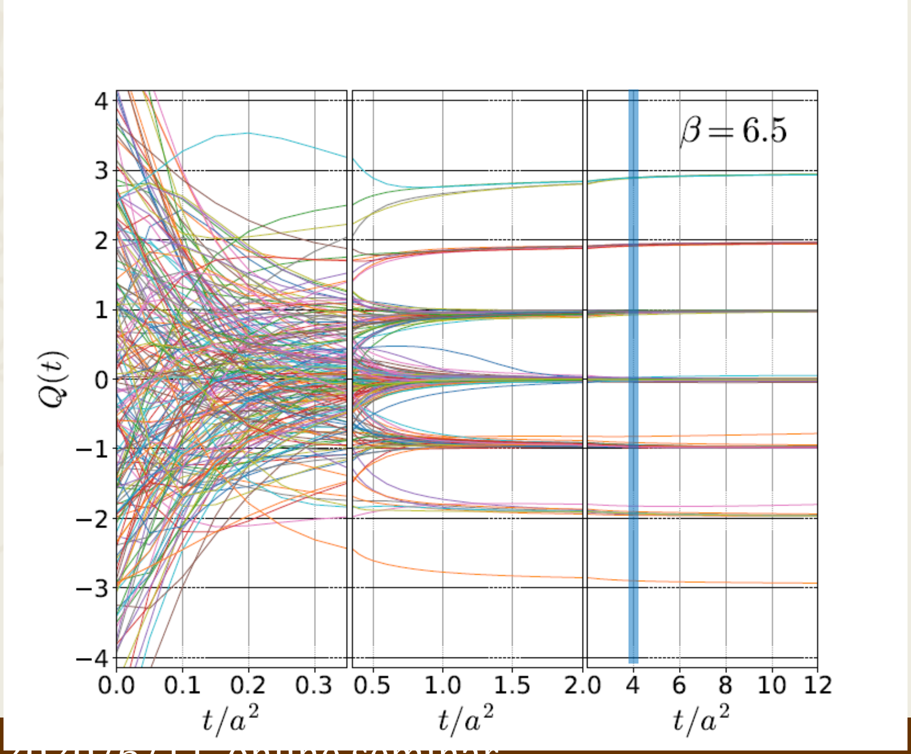
- 連続極限 / 体積依存性
- DIGAが適用可能な高温配位の解析
- 論文の出版

backup

Neural Network Setting

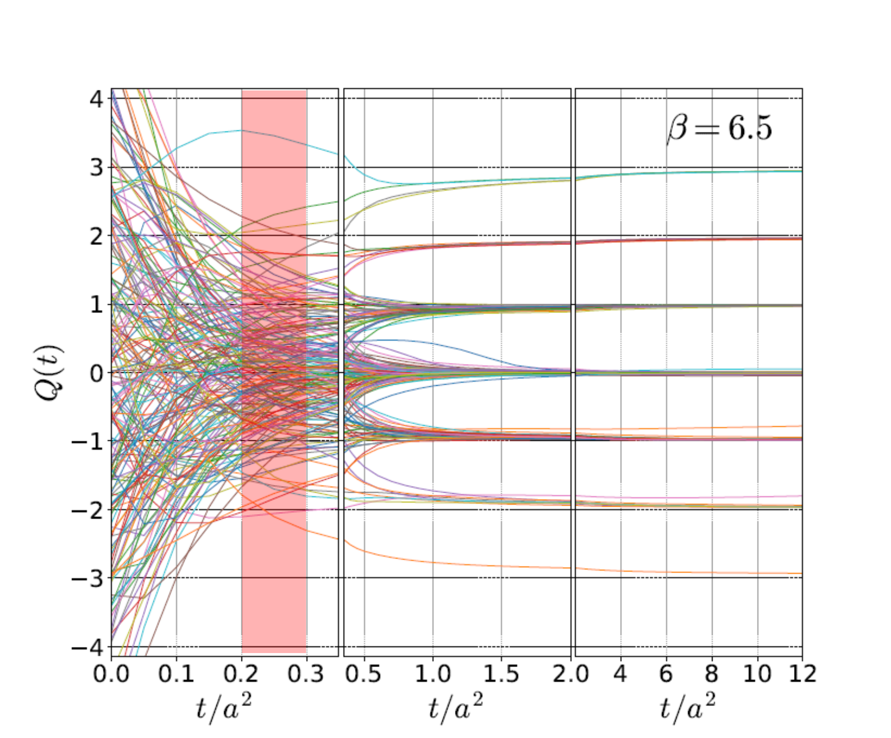
- convolutional neural network by **CHAINER framework**
- supervised learning
- convolutional layer: 4-dim., periodic BC
- regression analysis / round off to obtain integer
- activation: logistic

- answer of Q
  - Q(t) @ t/a^2=4.0
  - round off



Using different flow times

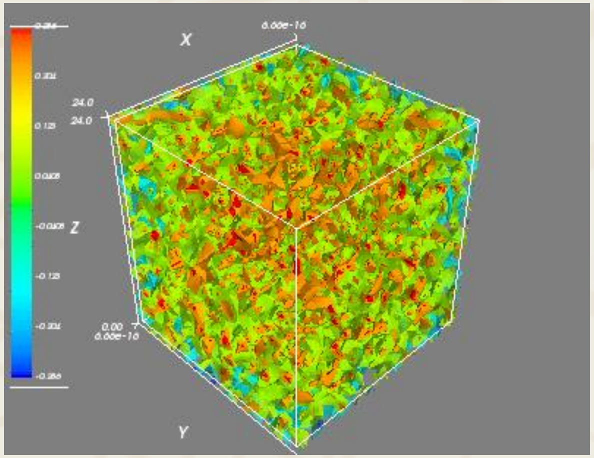
t/a^2	β=6.2	β=6.5
0.3, 0.25, 0.2	**95.9(2)**	**99.0(2)**
0.3, 0.2, 0.1	95.5(2)	95.7(2)
0.25, 0.2, 0.15	95.1(3)	95.0(2)
0.2, 0.15, 0.1	86.9(3)	83.1(4)
0.2, 0.1, 0	75.6(5)	68.2(4)
0.15, 0.1, 0.05	71.8(4)	65.2(4)
0.1, 0.05, 0	54.8(5)	49.9(3)



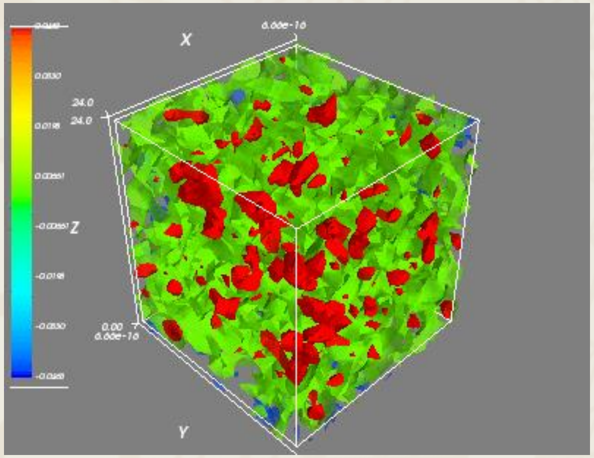
- t/a^2=0.3, 0.25, 0.2 gives the best accuracy.
- Better accuracy on the finer lattice.
- More than three t values do not improve accuracy.
  - error: variance in 10 independent trainings

Topological Charge Density

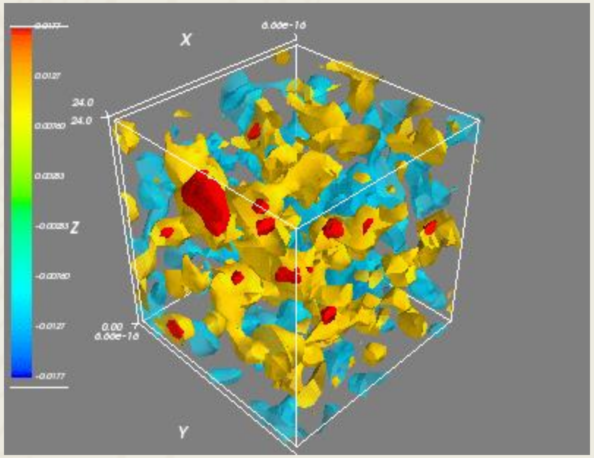
$t/a^2 = 0.1$



$t/a^2 = 0.2$



$t/a^2 = 0.3$



No isolated instanton structure…

Lattice Setting

- **SU(3) Yang-Mills**
- **Wilson gauge action**
- 2 lattice spacings with **same physical volume**
- $LT_c \sim 0.63$
- $\langle Q^2 \rangle \simeq 1.1$

- **Gradient flow** for smoothing

β	N^4	N_{conf}
6.2	16^4	20,000
6.5	24^4	20,000

20,000 confs. in total

Training: 10,000

Validation: 5,000

Test: 5,000

distribution of Q

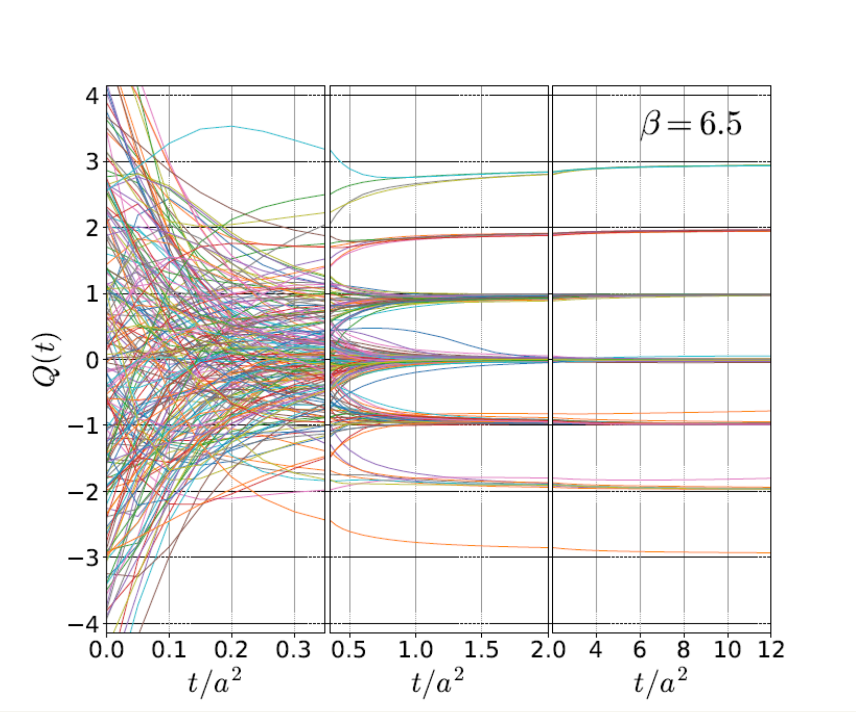
Q	-5	-4	-3	-2	-1	0	1	2	3	4	5
$\beta = 6.2$	2	17	235	1325	4571	7474	4766	1352	240	18	0
$\beta = 6.5$	0	5	105	1080	4639	8296	4621	1039	202	13	0

Topology on the Lattice

- Distinct topological sectors on sufficiently fine lattices

  Luscher, 1981
- Definitions of Q on the lattice:
  - fermionic: Atiyah-Singer index theorem
  - gluonic: q(x) after smoothing
    - cooling, smearing
    - **gradient flow**

    Luscher, Weisz, 2011



- Good agreement b/w various definitions
- **Faster algorithm is desirable!**

Details of NN

Trial 1~3

Layer	Filter size	Output size	Activation
input		$8^4 \times 1$	
convolution	$3^4 \times 5$	$8^4 \times 5$	Logistic
convolution	$3^4 \times 5$	$8^4 \times 5$	Logistic
average pooling	$8^4 \times 1$	5×1	
full connect		5	Logistic
full connect		1	

Trial 4

Layer	Output size	Activation
input	3	
full connect	5	Logistic
full connect	1	

Trial 5

Layer	Filter size	Output size	Activation
input		$12^d \mathrm{or} 8^d \times 3$	
convolution	$3^d \times 5$	$12^d \mathrm{or} 8^d \times 5$	Logistic
convolution	$3^d \times 5$	$12^d \mathrm{or} 8^d \times 5$	Logistic
convolution	$3^d \times 5$	$12^d \mathrm{or} 8^d \times 5$	Logistic
average pooling	$12^d \mathrm{or} 8^d \times 1$	5×1	
full connect		5	Logistic
full connect		1	