

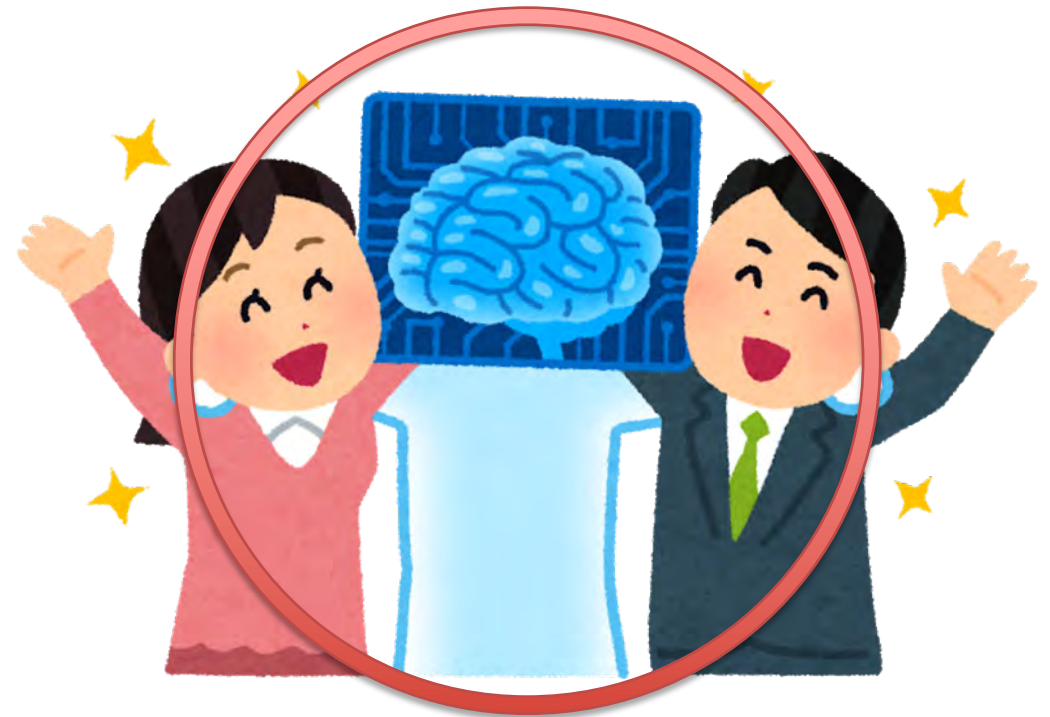
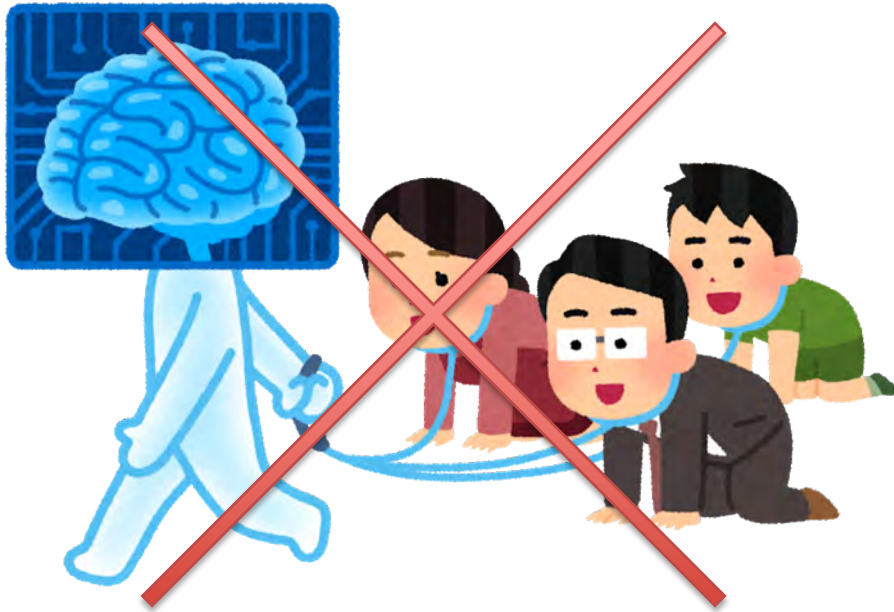
6. データ科学の動向について

プラズマ・核融合学会年会 インフォーマルミーティング
計算科学研究部会

量子科学技術研究開発機構

本多充

データ科学とは



データ科学とは



- データ科学、データ駆動科学、データサイエンス etc.
- データを用いて新たな科学的知見を引き出すアプローチ
~帰納的 ≠演繹的 ~ビッグデータ
- 超領域性
- 数学、統計学、計算機科学、情報工学…

キーワード

- 機械学習
- 深層学習
- ベイズ理論
- データ同化

etc.



近年の論文動向（超私見）



LETTER

Nature 19'

<https://doi.org/10.1038/s41586-019-1116-4>

Predicting disruptive instabilities in controlled fusion plasmas through deep learning

Julian Kates-Harbeck^{1,2,3*}, Alexey Svyatkovskiy^{4,5} & William Tang^{3,4}

Harvardグループ：LSTMによる、JET&DIII-Dのディスラプション予知

Nucl. Fusion **60** (2020) 036022 (16pp)

NF 20'

<https://doi.org/10.1088/1741-4326/ab6c7a>

Classification of tokamak plasma confinement states with convolutional recurrent neural networks

CNN/LSTMによる、TCVのL-D-H遷移の自動検知

F. Matos¹, V. Menkovski², F. Felici³, A. Pau³, F. Jenko¹, the TCV Team^{3,a} and the EUROfusion MST1 Team^b

¹ Max Planck Institute for Plasma Physics, Boltzmannstraße 2, 85748 Garching, Germany

プラズマ・核融合分野への応用(Bayesian)

Plasma Phys. Control. Fusion 61 (2019) 125012 (13pp)

PPCF 19'

<https://doi.org/10.1088/1361-6587/ab4e69>

On the importance of model selection when inferring impurity transport coefficient profiles

ベイズ手法を用いた不純物
輸送のDとVの分離と妥当性
検証

M A Chilenski¹, M Greenwald¹, Y Marzouk², J E Rice¹ and A E White¹

¹Plasma Science and Fusion Center, Massachusetts Institute of Technology, Cambridge, MA 02139, United States of America

Plasma Phys. Control. Fusion 61 (2019) 075012 (14pp)

PPCF 20'

<https://doi.org/10.1088/1361-6587/ab1d26>

Neural network approximation of Bayesian models for the inference of ion and electron temperature profiles at W7-X

ベイズ推定を模擬した
NNを用いた分布推定

A Pavone¹, J Svensson¹, A Langenberg¹, U Höfel¹, S Kwak¹,
N Pablant², R C Wolf¹ and the Wendelstein 7-X Team¹

¹Max-Planck-Institut für Plasmaphysik, Teilinstitut Greifswald, D-17491 Greifswald, DE, Germany

プラズマ・核融合分野への応用(surrogate)

PoP 19'

Machine-learning assisted steady-state profile predictions using global optimization techniques

Cite as: Phys. Plasmas **26**, 102307 (2019); doi: [10.1063/1.5117846](https://doi.org/10.1063/1.5117846)
Submitted: 1 July 2019 · Accepted: 2 October 2019 ·
Published Online: 8 September 2020



M. Honda^{a)} and E. Narita

AFFILIATIONS

Naka Fusion Institute, National Institutes for Quantum and Radiological Science and Technology, Naka, Ibaraki 311-0193, Japan

TGLFの代理モデル構築のためのデータ生成から、モデルを使用したシミュレーションまで手がけた論文

PoP 20'

Fast modeling of turbulent transport in fusion plasmas using neural networks

Cite as: Phys. Plasmas **27**, 022310 (2020); doi: [10.1063/1.5134126](https://doi.org/10.1063/1.5134126)
Submitted: 30 October 2019 · Accepted: 12 January 2020 ·
Published Online: 11 February 2020



K. L. van de Plassche,^{1,a)} J. Citrin,¹ C. Bourdelle,² Y. Camenen,³ F. J. Casson,⁴ V. I. Dagnelie,^{1,5} F. Felici,⁶
A. Ho,¹ S. Van Mulders,⁵ and JET Contributors^{b)}

AFFILIATIONS

¹DIFFER, PO Box 6336, 5600 HH Eindhoven, The Netherlands

罰則つき学習によるQualiKizの代理モデルの開発と応用

プラズマ・核融合分野への応用(データ同化)

Nucl. Fusion **60** (2020) 056001 (8pp)

NF 20'

<https://doi.org/10.1088/1741-4326/ab7596>

Data assimilation system based on integrated transport simulation of Large Helical Device plasma

Yuya Morishita¹, Sadayoshi Murakami¹, Masayuki Yokoyama^{2,4} and Genta Ueno^{3,4,5}

¹ Department of Nuclear Engineering, Kyoto University, Nishikyo, Kyoto, 615-8540, Japan

データ同化手法を用いた、TASK3DによるLHDプラズマシミュレーションの再現性精度の向上

Fusion Engineering and Design **155** (2020) 111566

FED 20'

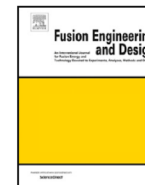


ELSEVIER

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Fusion Engineering and Design

journal homepage: www.elsevier.com/locate/fusengdes



Application of the best-estimate model calibration and prediction through experimental data assimilation methodology to the tests performed on a helium cooled First Wall Mock-up

Bruno Gonfiotti*, B. Ghidersa, M. Ionescu-Bujor, X.Z. Jin, R. Stieglitz

Karlsruhe Institute of Technology, Karlsruhe, Germany



原型炉ヘリウム冷却第一壁の流動性喪失事故の模擬精度向上のためのデータ同化

現在の取り組み



富岳とデータ科学



スーパーコンピュータ「富岳」

<https://fugaku100kei.jp/fugaku/>



「富岳」は、幅広い分野のアプリケーションソフトウェアを利用でき、「京」の最大100倍の実効性能をもちあわせている世界最高水準のスーパーコンピュータです。2020年6月、11月のスパコン性能ランキングでは2期連続4冠を達成しました。

兵庫県神戸市にある理化学研究所計算科学研究センターに設置され、2021年度より本格的に利用され始めます。

成果創出加速プログラム課題の紹介

2021年度からの本格的な利用の前に、試験的に「富岳」を使って社会的・科学的に重要な研究を行う課題です。文部科学省により、シミュレーションとAIやデータサイエンス、災害対策等を組み合わせた19件の課題が選ばれました。優先的に「富岳」を使うことで、早期の研究成果が期待されます。

計7件 →

領域① 人類の普遍的課題への挑戦と未来開拓

宇宙の構造形成と進化から惑星表層環境変動までの統一的描像の構築

牧野 淳一郎（神戸大学）

核燃焼プラズマ閉じ込め物理の開拓

渡邊 智彦（名古屋大学）

シミュレーションで捉える基礎科学・素粒子の根本法則から元素の生成まで

プロジェクトの詳細は次のご講演をご参照ください

5次元空間($k_x, k_y, z, v_{\parallel}, v_{\perp}$)数値データの画像化

- 画像による状態の分類
 - 線形、非線形、飽和
- 飽和までの時刻を予測する
- 研究者の経験に基づく調査箇所決定
 - 網羅的な探索による異常検知
- 人間には気付かない（気づきにくい）差異の検出
- 飽和レベルを予測する
- 時系列データによる解析
- 保存容量の圧縮

⋮

CNNと転移学習

Convolutional Neural Network



CNNとは

- 畳み込みニューラルネットワーク
Convolutional Neural Network $\neq$ Cable News Network
- 深層学習の一分野
- 画像認識技術：画像・動画認識、レコメンド、自然言語処理

認識: 2012年以降

	Error
Before ディープ ラーニング	Imagenet 2011 winner (not CNN)
	25.7%
	Imagenet 2012 winner
	16.4% (Krizhevsky et al.)
	Imagenet 2013 winner
	11.7% (Zeiler/Clarifai)
	Imagenet 2014 winner
	6.7% (GoogLeNet)
After ディープ ラーニング	Baidu Arxiv paper: 2015/1/3
	6.0%
	Human: Andrej Karpathy
	5.1%
	Microsoft Research Arxiv paper: 2015/2/6
	4.9%
	Google Arxiv paper: 2015/3/2
	4.8%
	Microsoft Research CVPR paper: 2015/12/10
	3.6%

2015年2月には人間の精度を超えた

画像認識で人間の精度を超えることは数十年間、実現されていなかった





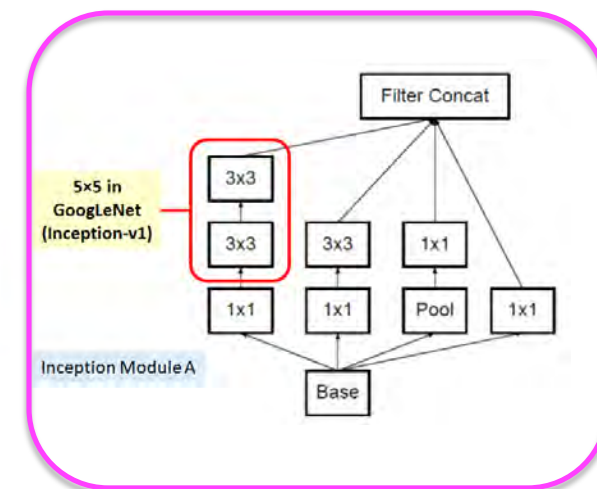
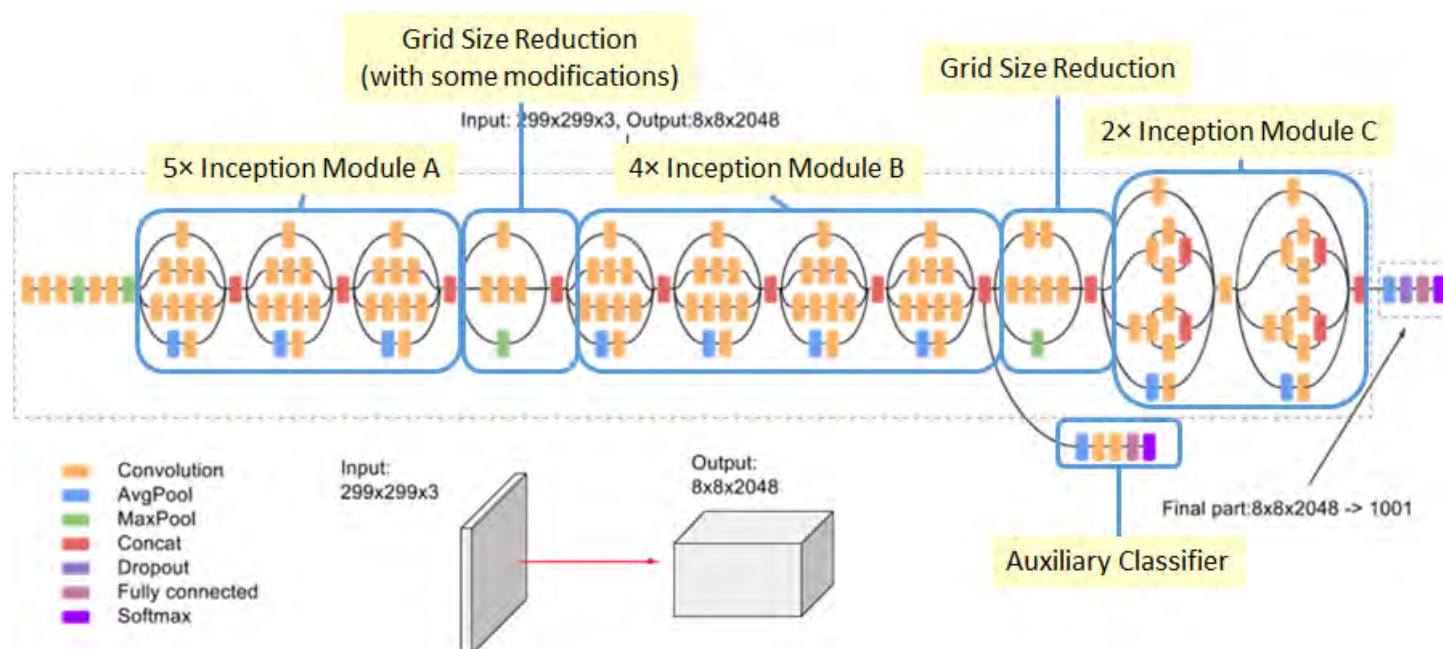
- 画像認識研究のための研究用標準データセット
- 1400万枚を超える自然画像を2万種以上のラベルで分類
- ILSVRCというコンペでは、訓練データ120万枚、1000ラベルで性能を競う

<http://image-net.org/index>

<https://kotobank.jp/word/ImageNet-2119310>

Inception v3

- 代表的なCNN
- パラメータ数は約2,400万
- ILSVRC2014で優勝したモデルGoogLeNetをInception-v1として改良を重ねられたモデルで、**転移学習**に広く応用が進んでいる
- 複数の畳み込み層やプーリング層から構成される**Inceptionモジュール**と呼ばれるマイクロネットワークを重ねている



Inceptionモジュールの概念

転移学習モデルによる犬猫分類

- ImageNetで学習済みのInception v3を用いた転移学習による犬猫分類
 - TensorFlow 2.3 (最新) + Keras on Jupyter Notebook
 - ~10 mins on NVIDIA GeForce GTX 1080 Ti
 - **Fine-tune:** 元の分類器を削除し、GlobalAveragePooling2D+1024ユニットの全結合層+分類器を結合。249層までを凍結し、250層以降を解凍。Data Augmentation
 - Kaggleの“Dogs vs. Cats”データ3千(内、テストデータ千)で転移学習
 - **テストデータ (学習に使用していないデータ) に対する正答率99.2%**



evaluate acc: 0.9919999837875366



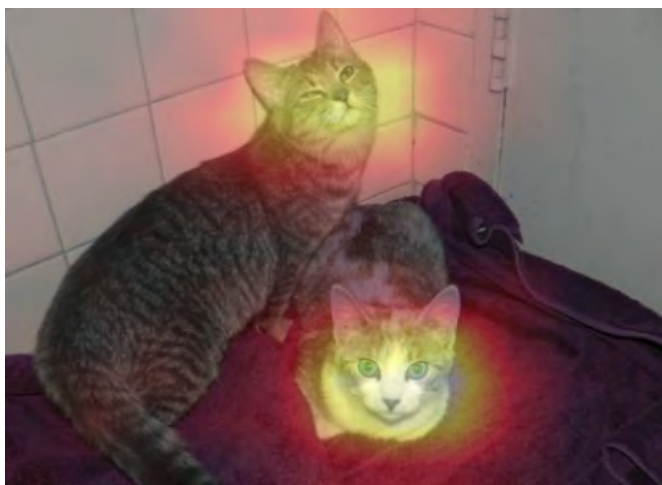
最新の高性能CNN

EfficientNet-B4 vs Inception v3

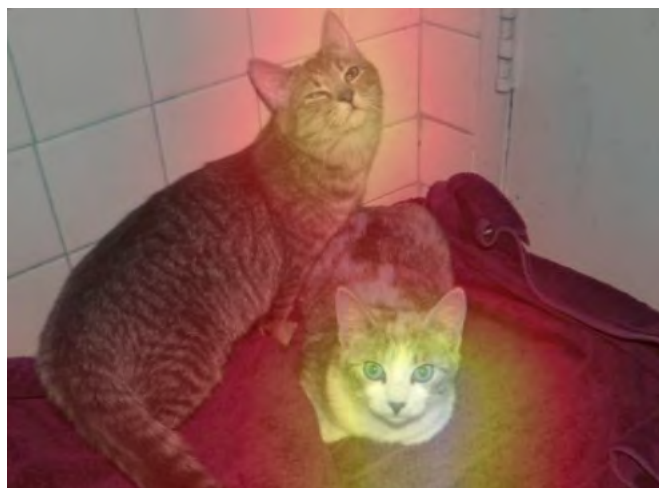
- モデルパラメータ数がInception v3とほぼ同等のEfficientNet-B4を以下では用いる

テストデータに対する検証	EfficientNet-B4	Inception v3
モデルパラメータ数	19,520,098	23,904,035
犬猫分類 (acc)	0.997	0.991
kxky 3クラス分類 (acc)	0.999	0.992
kxky 時刻予測 (mae)	0.0160	0.0213

ENet



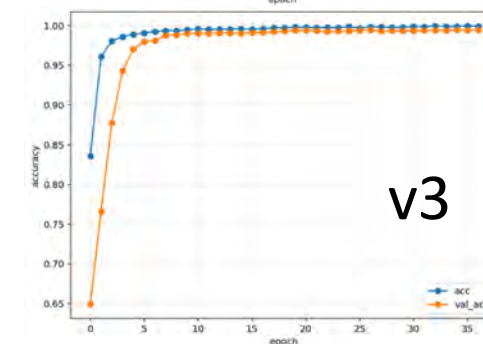
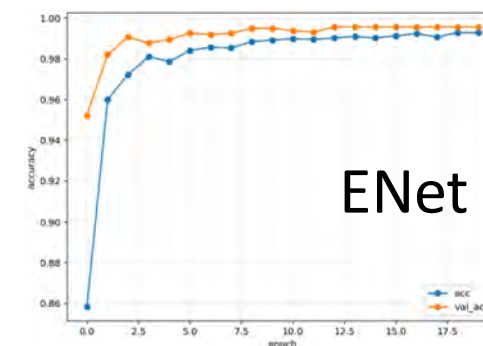
v3



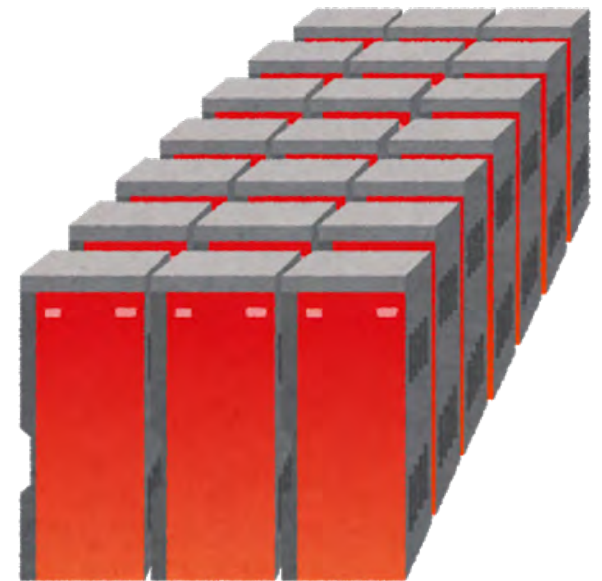
もはや間違えない

注視領域が、より猫の顔にのみ集中するようになる

より速く収束



GKVシミュレーションの 可視化



各時刻の特徴的なパターンの機械学習

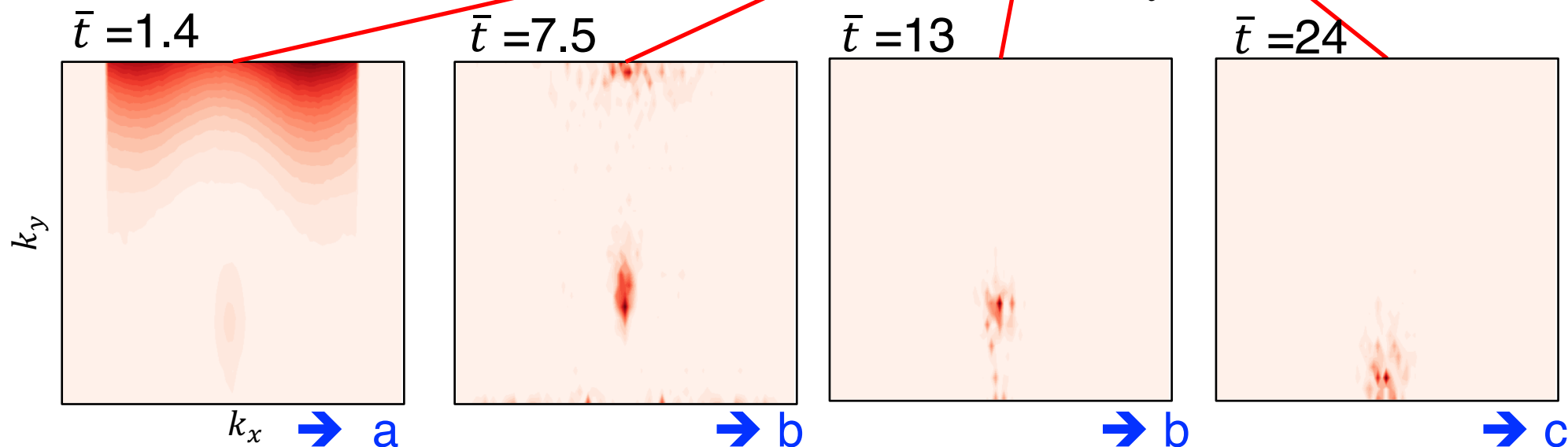
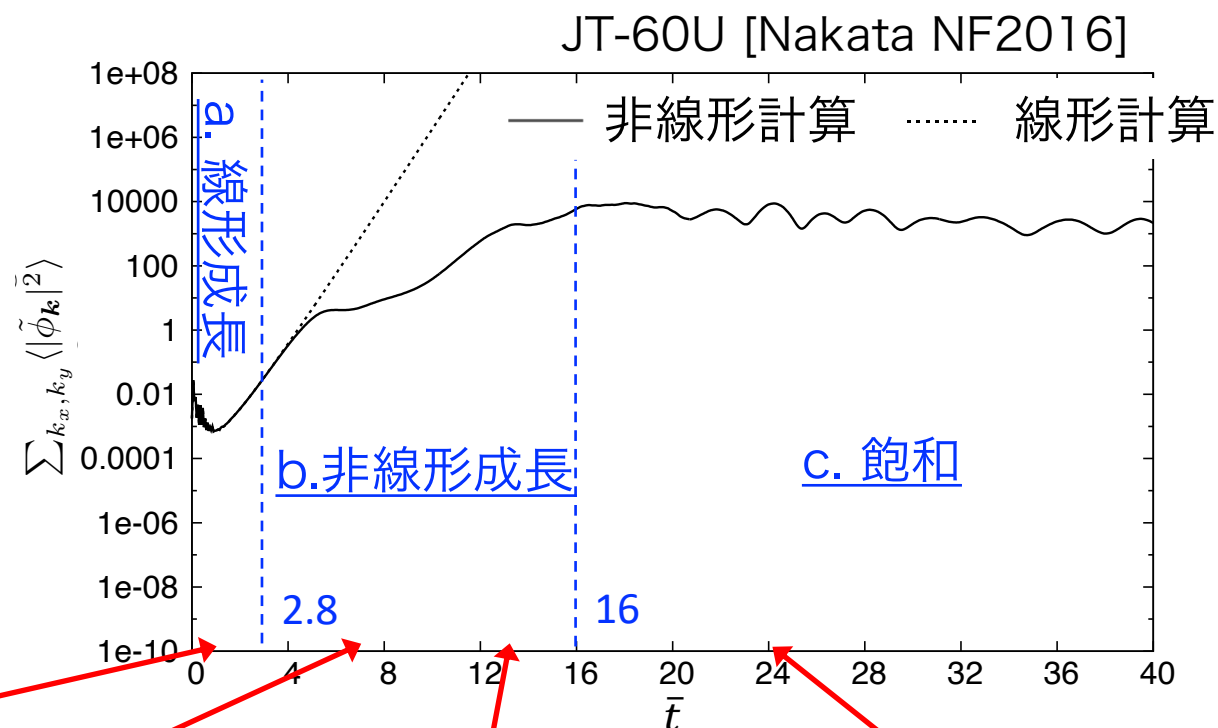
CNNを用いた

- 波数空間パターンに基づく3クラス分類

- 線形成長
- 非線形成長
- 飽和

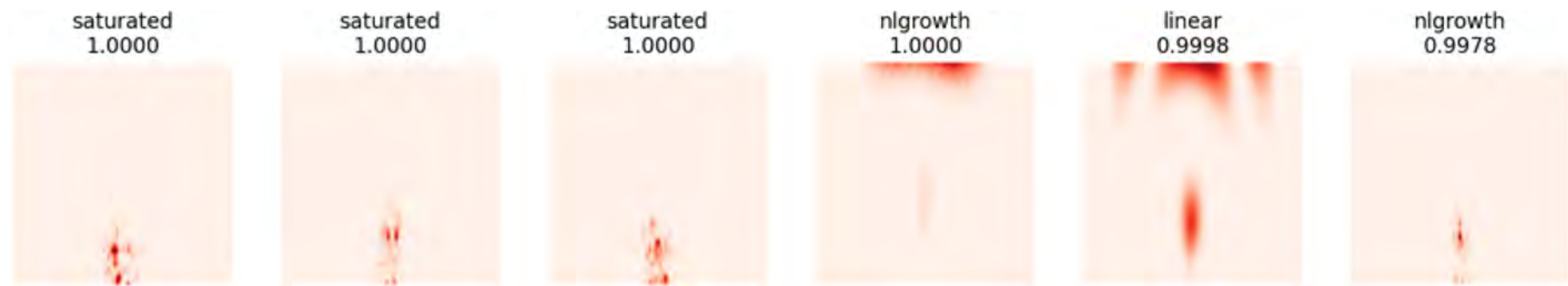
- 時刻予測

[成田 プラ核学会20' 02Ca03]

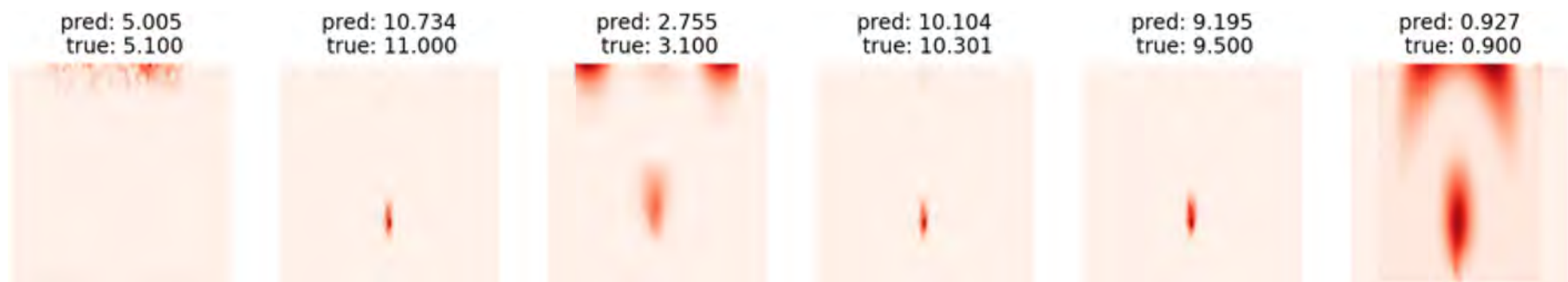


EfficientNet-B4の有効性

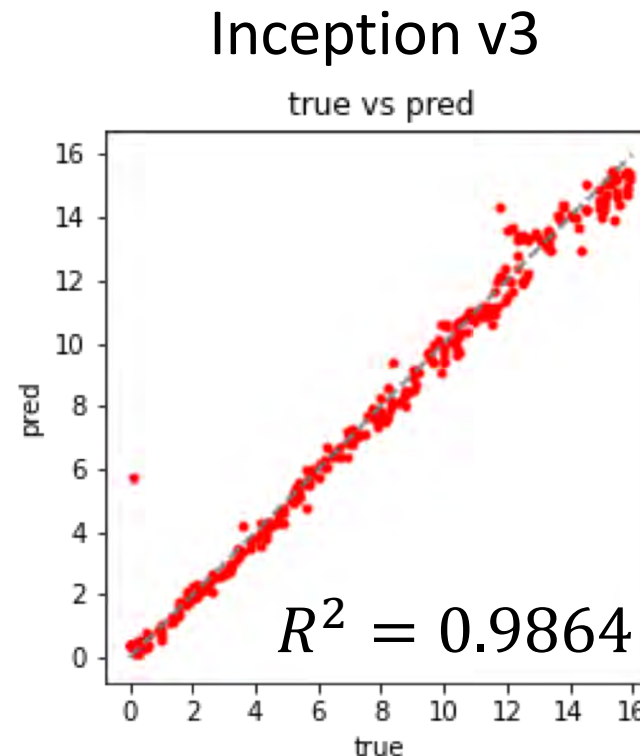
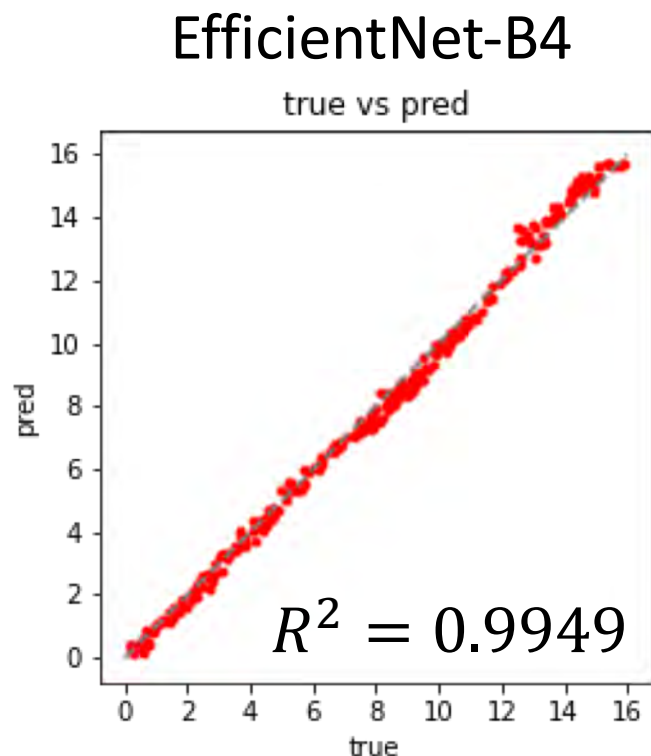
- 3クラス分類



- 時刻予測



EfficientNet-B4の有効性（続き）



- 時刻予測において、EfficientNet-B4は高い決定係数 R^2 を示す
 - Inception v3でも十分高い回帰能力を持っている
 - 線形フェーズから非線形発展フェーズまで、幅広い時間帯でむら無く予測が可能となっている

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

まとめ

- データ科学とは
- 近年の論文動向
- 富岳とデータ科学
- CNNと転移学習
- 最新のCNNモデル：EfficientNet-B4
- GKVデータの画像化による
 - クラス分類
 - 時刻予測（回帰）

