革新的画像解析技術を用いた 広域宇宙撮像データ分析

吉田直紀

東京大学 理学系研究科 物理学専攻 カブリ数物連携宇宙研究機構 次世代知能科学研究センター併任

JST CREST「ビッグデータ」 2014-2019, AIP加速課題 2020-2022

AIP加速研究の目標

1.宇宙観測画像・動画解析により時間変動天体や高速移動天体を検出

- すばるHSCデータを解析
- Tomo-e Gozen 動画データ即時解析
- 機械学習による自動天体分類システム構築
- 2. 広域宇宙マップを作成
 - ブレンド画像処理技術の開発
 - 深層学習を用いた密度マップ再構築法の開発
- 3.自律的エミュレーション技術を確立
 - 実データ解析とパラメータ推定
 - シミュレーションデータベースと融合









宇宙にAI? 深宇宙画像 before / after 2014





©国立天文台ハワイ観測所すばる望遠鏡

すばるHSCは1晩300ギガバイトのペースでデータを取得 1スナップショット ~ 1ギガピクセル

広域宇宙サーベイ: 空をくまなく探索する

Vera C. Rubin Observatory 1晩で20テラバイト Google が支援 10年で100ペタバイトとか



https://www.lsst.org

ESA's Euclid 撮像 + 分光



NASA's Nancy Grace Roman Telescope



https://www.Euclid-ec.org

Square Kilometre Array 建設中の電波望遠鏡群



Image credit: SKAO

- ・ 11か国協同で2024年完成を目指す
- ・ 南アフリカとオーストラリアに~1,000,000個のアンテナ
- 毎日生データは 3 Exabyte、保存するのは1Pbyte (Youtube 1 exabyte 分を見るのに20万年かかる)
 データ処理に世界最大級の専用スパコンが複数必要

宇宙の理解

■ 我々の住む宇宙の成り立ちを理解する

■様々な場所での物理現象を理解する

■未知の現象、未知の物理を発見する

画像差分解析による超新星爆発検出 明るさを変える星を探せ: レベル1





機械学習自動検出器により1晩で100個もの超新星がみつかる!



N. Yasuda et al., Pub. Astron. Soc. Jpn, 71 (2019) 74 (2019.9 プレスリリース)

深層学習自動分類器の性能評価



宇宙の距離測定に用いることの できる重要な「Ia型」の抽出は 精度96パーセントで可能。 数の少ないサンプルは今後の観 測により検出例と精度を向上。



I. Takahashi, Pub. Astron. Soc. Jpn, 72 (2020) 89 自動分類器を公開 https://github.com/ichiro-takahashi/snclass

時間領域天文学のフロンティア



従来の天文学では1日おきの観測・検出が限界であった

これまでに発見された短時間突発天体



未解明の謎:これらの現象は可視光でも輝いているのか?



東京大学木曽観測所







数秒だけ光る謎の天体









Single Shot Multibox detector の応用





https://soralab.space-ichikawa.com /2017/01/mxnet-ssd-training/

検出された物体ごとに「星」「ごみ」「突発天体」など信頼度を 計算していく。GPU でリアルタイム処理が可能。

リアルタイム処理



一晩でこの例の10万倍、30テラバイトのデータ!
 動画でみたように大気揺らぎやノイズにさらされている.
 差分解析では天体より「ゴミ」が10倍-100倍みつかる.
 重要な天体ほど逃してはならない
 (=平均的な性能は追求していない)

銀河の集団が作り出す宇宙の「虫めがね」







宇宙の地図づくり



宇宙の地図を作成



1億光年

深層学習を用いたノイズ除去 (2021/7 プレスリリース)

ISM2021-06

2021年7月2日

埋もれた暗黒物質の地図を掘り起こす ―観測・シミュレーション・人工知能のタ ッグで描くクリアな宇宙―



概要

大学共同利用機関法人 情報・システム研究機構 統計数理研究所(所在地:東京都立川市、所長:椿 広計、 以下「統数研」) と 自然科学研究機構 国立天文台(所在地:東京都三鷹市、台長:常田佐久) は、 遠方 銀河観測での暗黒物質探査に有効な深層学習ネットワークを構築しました。

これまでの天文観測により、我々の宇宙を占める物質の80パーセント程度は光を発することのない物質であ ることが示唆されています。この物質は暗黒物質と呼ばれ、その正体はいまだ謎に包まれています。暗黒物質 の正体を解明するには、暗黒物質が宇宙のどこにどれくらいあるかを調べる必要があります。暗黒物質の地図 を作成するために、遠方銀河の重力レンズ効果を利用する手法が近年注目されています。今回、白崎正人 助教 (国立天文台・統数研)、森脇可奈 博士課程3年(東京大学)、大木平 研究員(千葉大学統合情報センター)、 吉田直紀 教授(東京大学)、池田思朗 教授(統数研)、西道啓博 特定准教授(京都大学基礎物理学研究所) **Science News**

from research organizations

Observation, simulation, and AI join forces to reveal a clear universe

Date:	July 2, 2021					
Source:	National Institutes of Natural Sciences					
Summary:	Astronomers have developed a new artificial intelligence (AI) technique to remove noise in astronomical data due to random variations in galaxy shapes. After extensive training and testing on large mock data created by supercomputer simulations, they then ap- plied this new tool to actual data from Japan's Subaru Telescope and found that the mass distribution derived from using this method is consistent with the currently ac- cepted models of the Universe.					
Share:	f	y	P	in		
RELATED TOPICS					FULL STORY	
Space & Time					Japanese astronomers have developed a new artifi-	

Astrophysics
 Galaxies
 Astronomy
 Dark Matter
 cial intelligence (Al) technique to remove noise in astronomical data due to random variations in galaxy shapes. After extensive training and testing on large mock data created by supercomputer simulations, they then applied this new tool to actual data from

統計数理研 白崎、池田ら 2021, Mon. Not. Roy. Astron. Soc, 504, 1825

敵対的生成アルゴリズムによるノイズ除去

ノイズの特徴を学習した生成器(G)



膨大な数の理論シミュレーションを遂行→学習データ



宇宙の理解

■ 我々の住む宇宙の成り立ちを理解する

■様々な場所での物理現象を理解する

■未知の現象、未知の物理を発見する



「エミュレータ」統合統計解析システム







推定できるモデルパラメータを9つにまで拡張

Dark Emulator version 1 を公開 (1万ダウンロード!)

Dark Quest Project Webpage

A suite of cosmological N-body simulations and a handy emulator to explore cosmological parameter space

pypi package 1.0.23 Downloads 9616 Anaconda.org 1.0.23 ? downloads 70 total license MIT_License Last updated 29 Oct 2021

What's new?

Dark Emulator is now publicly available! (March 19, 2021)

Overview

Dark Quest is a cosmological structure formation simulation campaig understand the complex parameter dependence of various large-scal inference problems with observational datasets. The first series of sin second phase (DQ2). A Gaussian-Process based emulation tool, **Dark**

Our Team

2021年3月 専用ウェブ(英語) https://darkquestcosmology.github.io/ 最新版は github で随時更新 pip や condaで簡単インストール

pip install dark_emulator

conda install –c nishimichi dark_emula

Downloads 9616

As of 11/11/2022

すばる望遠鏡重カレンズ観測の実データ解析に適用



宇宙の進化を司るフリードマン方程式

ー様等方な宇宙を仮定するロバートソン-ウォーカー計量に対して アインシュタイン方程式は次の式に帰着する:

$$\left(rac{\dot{a}}{a}
ight)^2 = H_0^2 \begin{bmatrix} \Omega_{
m m} \\ rac{\Omega_{
m m}}{a^3} + \Omega_{\Lambda} \end{bmatrix}$$
 $\Omega_{
m m}$:物質密度 Ω_{Λ} :宇宙定数



まとめ

- 宇宙観測データ、特に全天サーベイ観測のデータは 解析そのものが困難になっている (~ エクサバイト級)
- ・リアルタイム処理(時間軸)は困難。現状は10TB/night. データの取捨選択はできるのか
- ・この後 ~2030までこのトレンドは加速していく
- ・画像データ分析により遠くの超新星や新種の天体が
 多数発見され、宇宙の成り立ちを理解できるようになる
- ・基礎科学データ(volume, quality) は現代の統計学, 機械学習, データサイエンスの格好の材料でもある