深層学習によるTomo-e Gozen突発天体 サーベイにおけるReal/Bogus分類改善

Deep-learning real/bogus classification for the Tomo-e Gozen transient survey Publications of the Astronomical Society of Japan, Volume 74, Issue 4, August 2022, Pages 946–960, https://doi.org/10.1093/pasj/psac047

高橋一郎(東京工業大学)

共同研究者:田中雅臣(東北大学),上田修功(NTT コミュニケーション科学 基礎研究所), 冨永望, 大澤亮(国立天文台), 酒向重行, 吉田直紀(東京大学)

^{2/14} Tomo-e Gozen プロジェクト

Kiso schmidt telescope

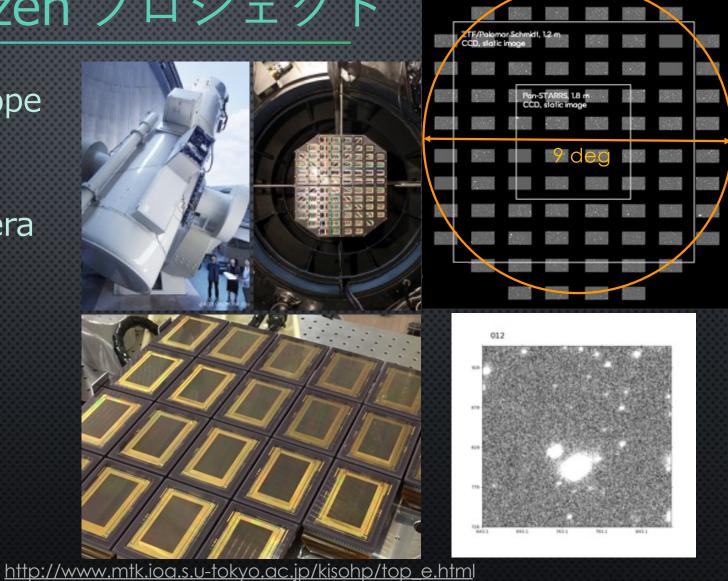
- 1.05 m
- FOV: 9 deg in diameter

Tomo-e gozen camera

- Wide-field CMOS imager
- 84 CMOS sensors
- 20 deg² sky in 2 fps
- 30 TB /night

CMOS image sensor

- 2000 x 1128 pixels
- Non-cooling
- Low read noise of 2 e⁻



3/14 Tomo-e Gozen 突発天体サ-

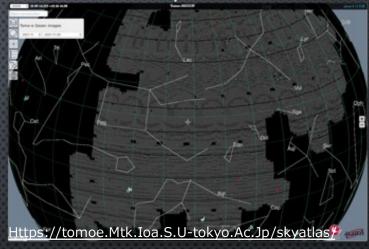
2種類のサーベイ

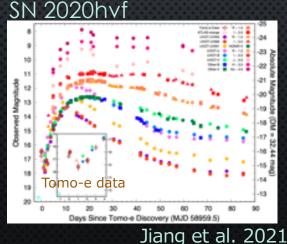
- 1. All-sky survey
 - 露出時間: 9 sec (18 frame)
 - $\sim 12000 \text{ deg}^2/\text{day}$
- 2. High-cadence survey
 - 露出時間: 6 sec (12 frame)
 - $\sim 3000 \text{ deg}^2/30 \text{ min}$
 - 3 visits/day

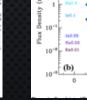
メインターゲット:

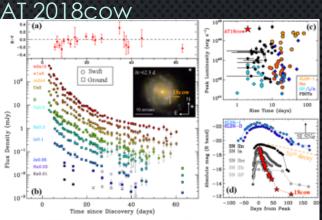
- Early supernovae
- Rapid transient

Tomo-e Gozen Sky Atlas









Margutti et al. 2019

4/14 突発天体の検出

- ・ 画像差分による検出
- 誤検出(Bogus)~10⁶/Night
 - -> 機械学習を使用したReal/Bogus分類

Bogusの例

差分の失敗





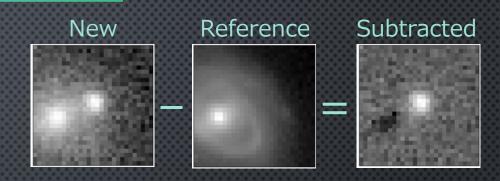
センサー起源



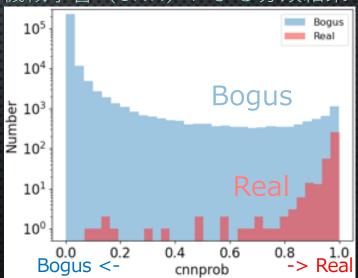








機械学習(CNN)による分類結果



本物が偽物に埋もれる

5/14 データセット

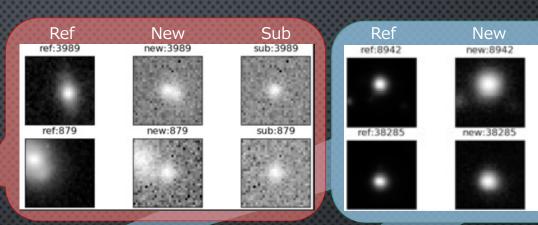
3つの画像で1セット

学習用(従来と同じ)

- Real:擬似的な突発天体・
 - 銀河周辺 (50%)
 - 画像全体にランダム (50%)
- Bogus:実際のもの

検証用

- Real:本物の突発天体
 - Transient Name Server (TNS)
- Bogus:実際のもの



データセットの内訳

Sub

sub:38285

Dataset		Number	Note
Training	Real	1,224,773	Artificial
	Bogus	2,031,193	Actual
Validation	Real	363	TNS
	Bogus	255,777	Actual

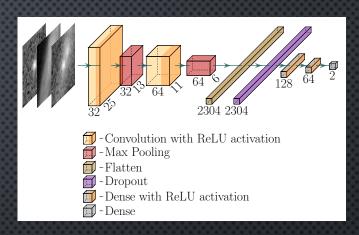
6/14 モデル構造

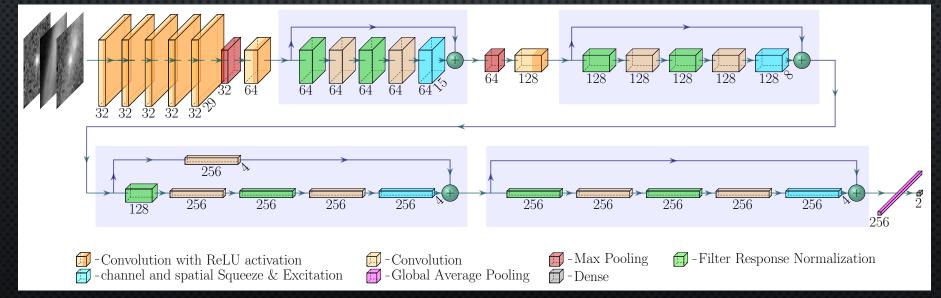
Simple model(従来版)

• 一般的な畳み込みニューラルネットワーク(CNN)

Complex model (改善版)

• より深く複雑な構造





""学習方法

1. 学習データの扱い方

全てのCMOSセンサーのデータを使用

- 2. 複数の目的関数
 - Cross-entropy
 - exp-Cross-hinge loss (AUC maximization)
 - Virtual Adversarial Training (VAT)

CMOS **CMOS CMOS CMOS CMOS CMOS** 84 84 Dataset Dataset Dataset Dataset Dataset aset Classifier Classifier Classifier Classifier

 $L(\{\mathbf{x}_l, y\}, \{\mathbf{x}_u\}; \boldsymbol{\theta}) = \lambda_{ce} L_{ce}(\{\mathbf{x}_l, y\}; \boldsymbol{\theta})$

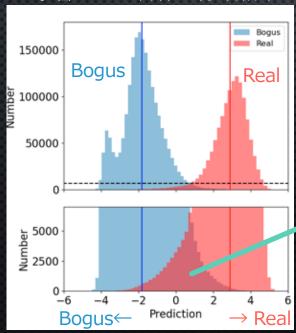
 $+\lambda_{\mathrm{ech}}L_{\mathrm{ech}}(\{\mathbf{x}_{l},y\};\boldsymbol{\theta})$

 $+\lambda_{\mathrm{lds}}L_{\mathrm{lds}}\left(\left\{\mathbf{x}_{l}
ight\}\cup\left\{\mathbf{x}_{u}
ight\};oldsymbol{ heta}
ight)$

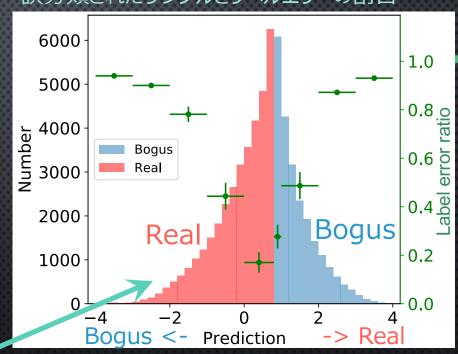
8/14 ラベルエラーへの対処方法

- 学習データにラベルエラーが存在
- CNNで学習データ自体を分類
 - 数が多いので標本調査
 - ラベルエラーの比率を計算

学習データ自体の分類結果

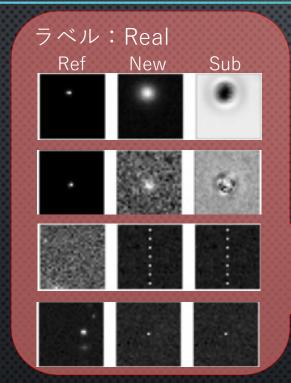


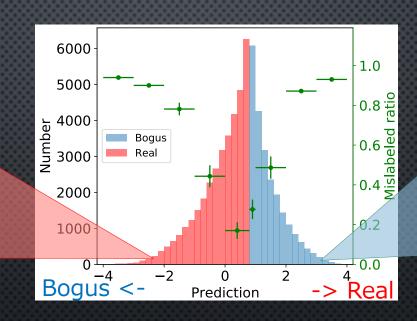
誤分類されたサンプルとラベルエラーの割合

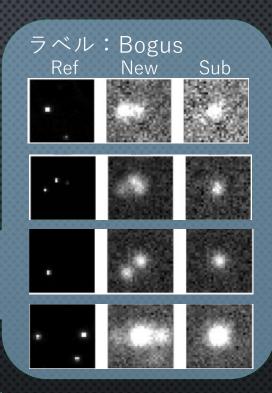


ラベルエラーの比率が裾野ほど増加

9/14 ラベルエラーへの対処方法

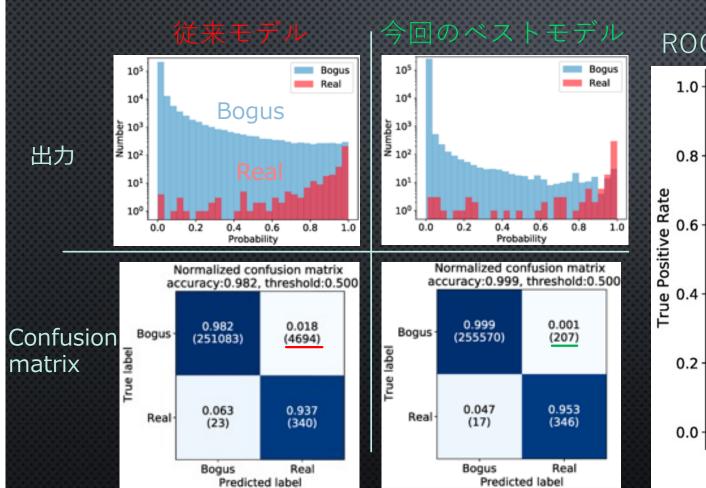




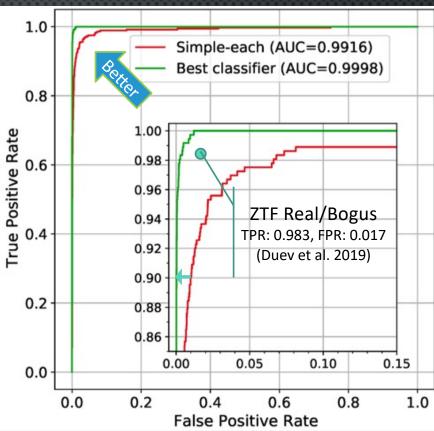


- •ラベルエラーの比率: 0.6% (Bogus), 1.0% (Real)
- •学習を2段階に:
 - (1)学習データ自体を分類し、誤分類しているものを「ラベルなし」に
 - (2)半教師あり学習

10/14 性能評価

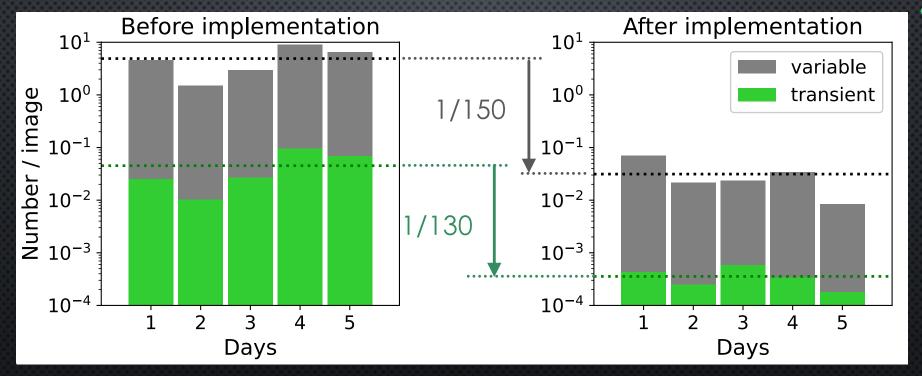


ROCカーブ



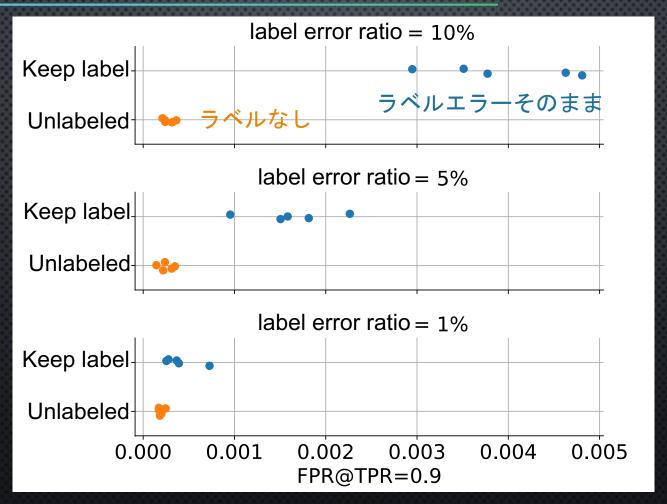
11/14 実運用での成績

- Tomo-e Gozenの解析パイプラインに新しい分類器を実装
- 実装前後での突発天体候補の登録数の変化を調査



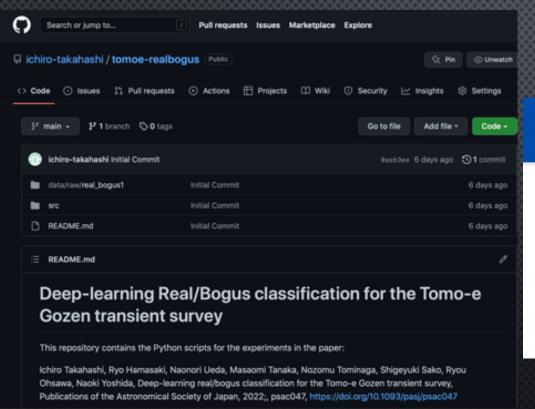
突発天体候補の登録数が1日あたり5000天体から40天体に

12/14 高いラベルエラー率に対する性能



ラベルエラーの割合が大きくても対応可能

13/14 データ公開



ソースコードとデータセットも公開

zenodo

Search Q

Jpload Communities

June 22, 2022

Dataset for the Real/Bogus classifier in the Tomo-e Gozen transient survey

Takahashi, Ichiro; Hamasaki, Ryo; Ueda, Naonori; Tanaka, Masaomi; Tominaga, Nozomu; Sako, Shigeyuki;
Ohsawa, Ryou; Yoshida, Naoki

This is a subset of the dataset used in the following paper:

Ichiro Takahashi, Ryo Hamasaki, Naonori Ueda, Masaomi Tanaka, Nozomu Tominaga, Shigeyuki Sako, Ryou Ohsawa, Naoki Yoshida, Deep-learning real/bogus classification for the Tomo-e Gozen transient survey, Publications of the Astronomical Society of Japan, 2022;, psac047, https://doi.org/10.1093/pasj/psac047

https://doi.org/10.5281/zenodo.6691804

https://github.com/ichiro-takahashi/tomoe-realbogus

14/14 まとめ

Tomo-e Gozen突発天体サーベイ Real/Bogus分類

新しい分類器

より深く複雑なモデル

学習データをまとめて使用

ラベルエラーへの対応

分類性能が従来に比べて100倍以上向上

人によるラベル付け直しの必要なし 高いラベルエラー率にも対応可能

