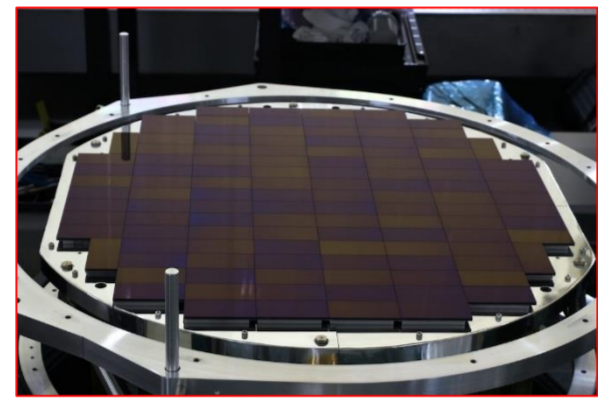
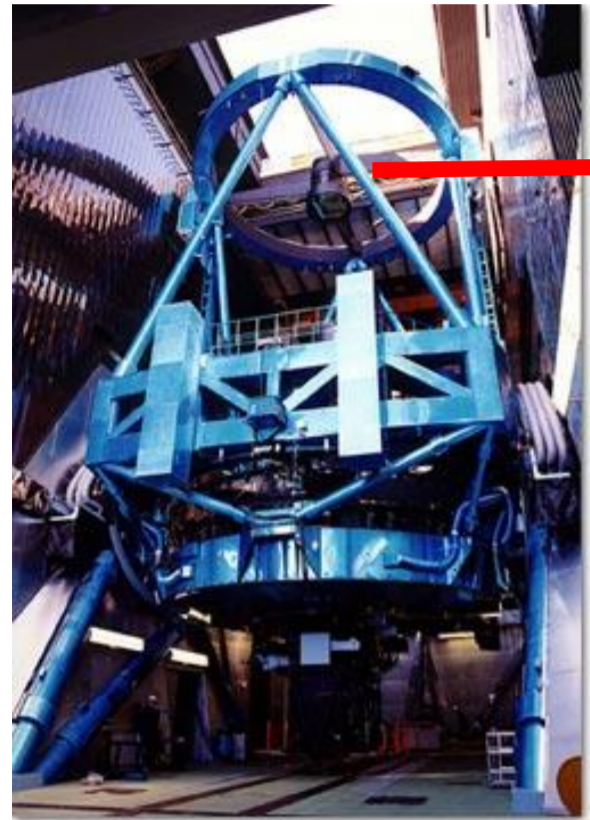


機械学習の手法を用いた突発天体の選別

森井 幹雄 統計的機械学習研究センター 特任研究員

すばる望遠鏡は国立天文台が所有する世界最大級の可視光望遠鏡(口径8.2m)で、主焦点に超広視野を持つカメラ:「Hyper Suprime-Cam (HSC)」が取り付けられている。このたび、遠方銀河の分布を調べて、ダークマター、ダークエネルギーの謎を解明するプロジェクトが始まり、HSCを用いて5年間で300夜の観測が行われる予定である。同じ領域を複数回観測するので、超新星爆発などの突発天体も多数検出されるはずである。しかしながら、偽の突発天体もそれ以上に検出されるため、機械的に突発天体を選び出す必要がある。HSCの2夜の観測で得られた突発天体候補16665天体について、天文学者の目視による判定の結果、61天体が本物と認められた。本研究では、天文学者と同じ判別を行う機械を、AUCを最大化するBoosting法(AUC Boosting; O. Komori 2011)によって作成した。False Negative Rate = 0.1 のとき、False Positive Rate = 0.09 の性能が出ることが確認され、有望な方法であることが確認された。同様の手法は、全天X線監視装置「Monitor of All-sky X-ray Image」で検出された突発天体候補の選別にも応用できると期待される。

すばる望遠鏡、Hyper Suprime-Cam(HSC)

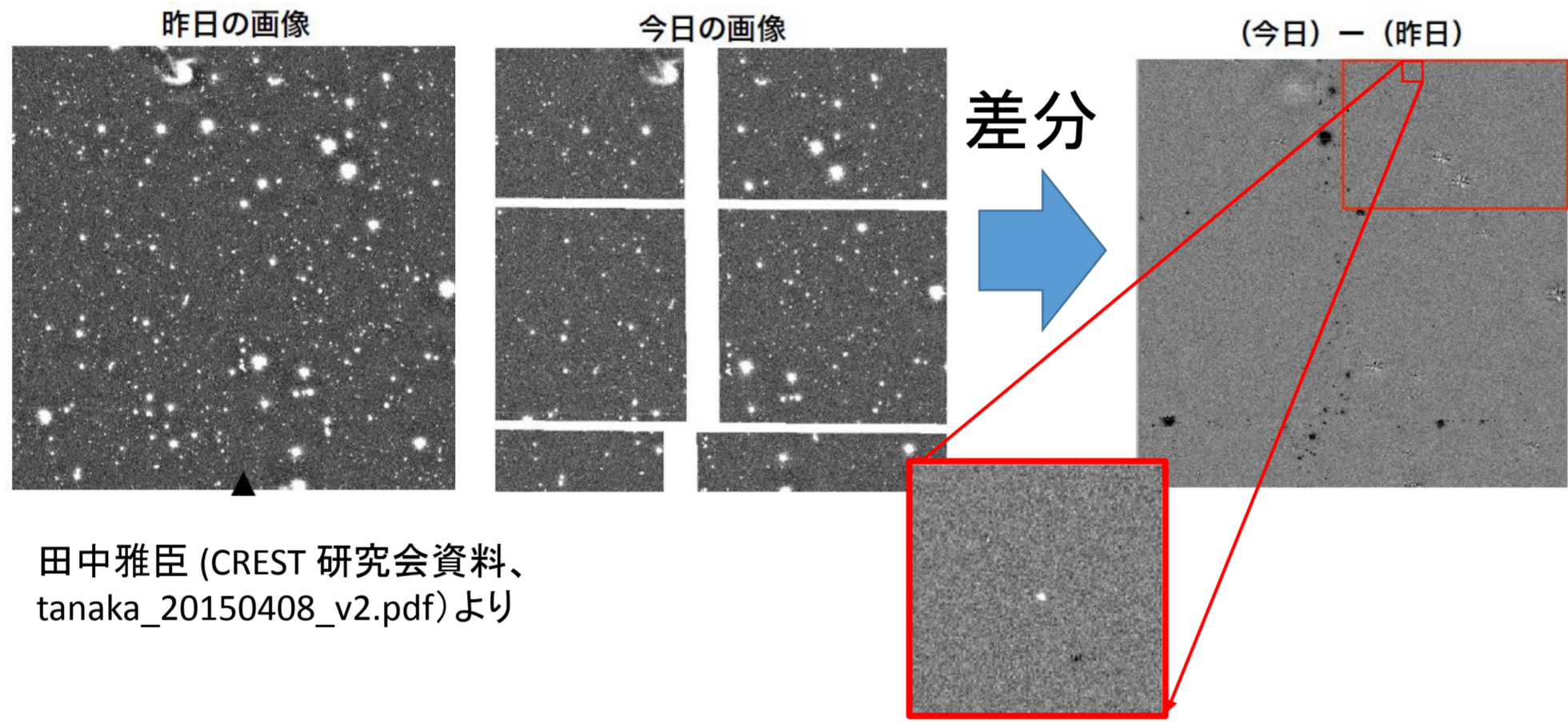


超広視野カメラ
Hyper Suprime-Cam (HSC)
主焦点に取り付けられている。

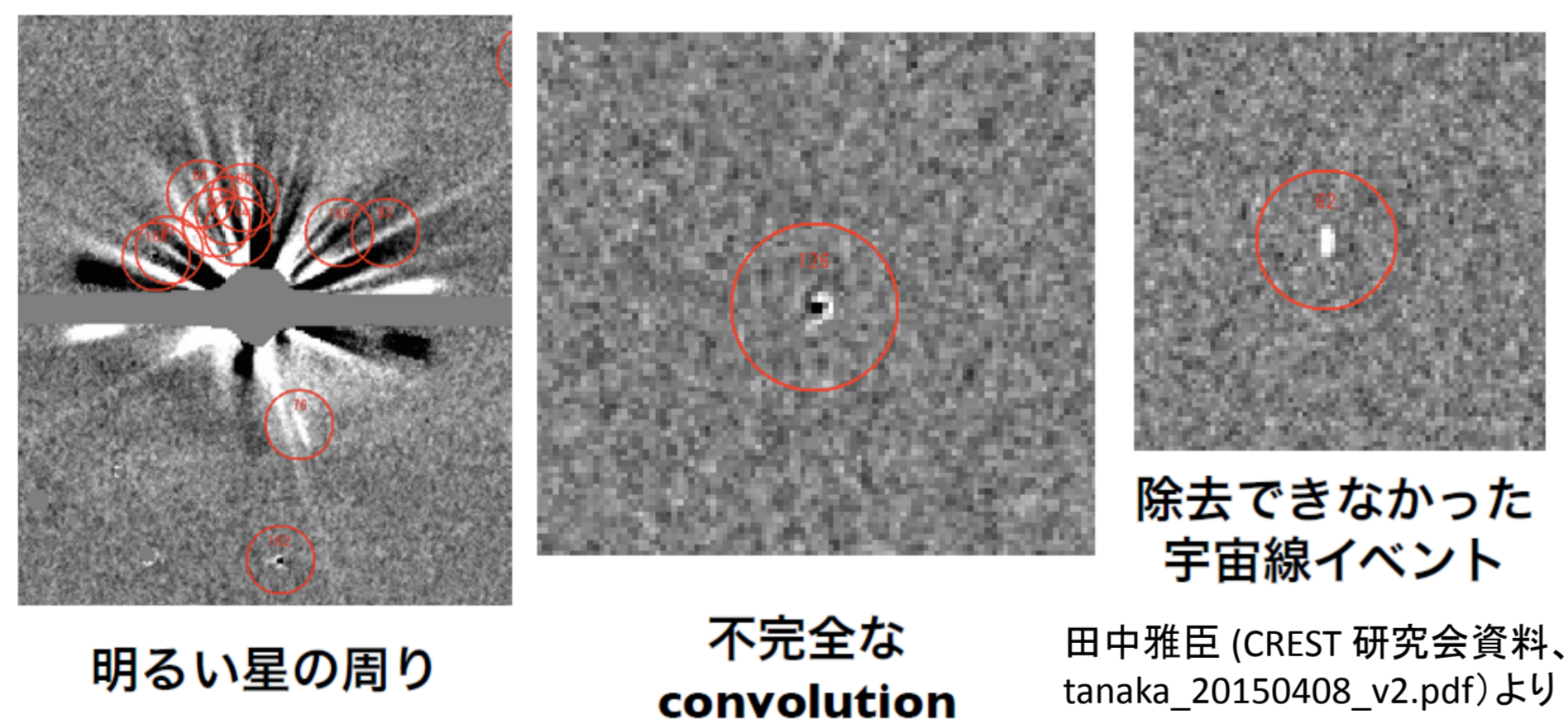
- 国立天文台が所有する世界最大級の可視・近赤外望遠鏡
- ハワイ島、マウナケア山山頂にある。
- 口径8.2メートル。

突発天体の観測例

2夜の画像を比較して、後半だけに検出された天体が突発天体



差分画像に残るニセ突発天体の例



突発天体の選別機械の作成

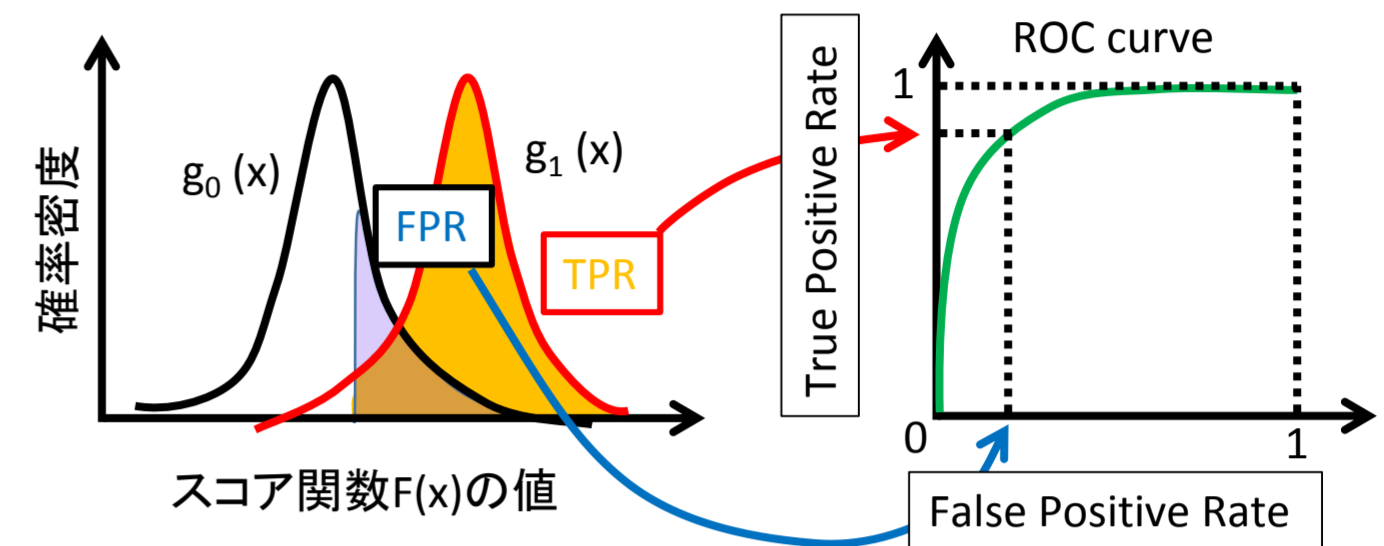
- 実際にHSCで取得したデータ(2夜)に対して突発天体の選別を行う。
- 差分画像に突発天体候補として検出された16665天体を、3人の天文学者(田中さん、富永さん、諸隈さん)が目視でチェックして本物、偽物の判定を行った。
- 本物の突発天体は61個、それ以外は、16339個。
- 全天体について、天体イメージの特徴量のベクトル(プログラムで自動抽出したもの)と、本物・偽物の判定値(天文学者による判定: 1 or 0)のデータセットがある。

天文学者と同じ判別をする機械を作ろう

AUC Boosting

判別機に対する要請: True Positive Rate (> 90%) の条件で、False Positive Rate をできるだけ下げたい。

- 統計数理研究所の小森理氏が開発した手法
 - Komori, O. (2011), Ann. Inst. Stat. Math, 63, 961, "A boosting method for maximization of the area under the ROC curve"
- AUC の値(ROC curve の下側の面積: Area Under the ROC Curve)が最大になるように学習する。

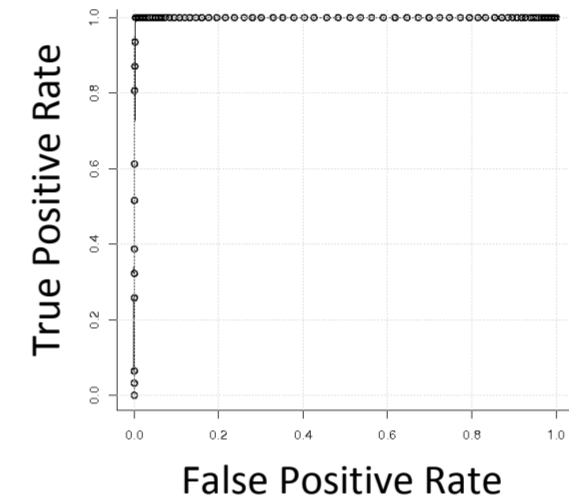


$$\text{AUC}_\lambda(F) = \frac{1}{n_0 n_1} \sum_{i=1}^{n_0} \sum_{j=1}^{n_1} \Phi(F(x_{1j}) - F(x_{0i})) - \lambda \sum_{k=1}^p \sum_{x_k \in \mathcal{B}_k} \{F_k^{(2)}(x_k)\}^2$$

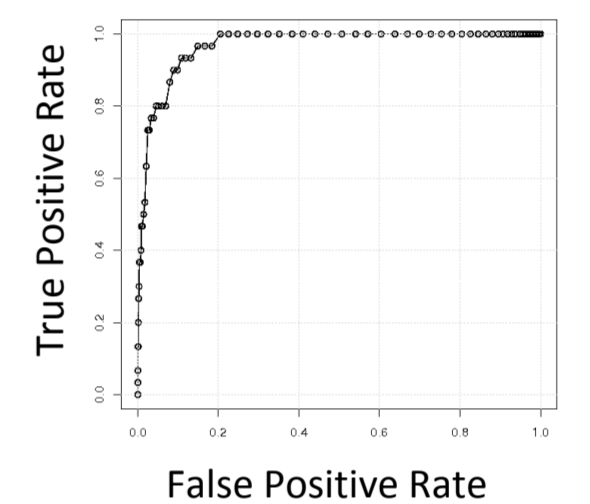
Empirical AUC Penalty Term for smoothness

λ : 未定のパラメータ: Stump関数のstepにおける滑らかさを制御するパラメータ。
 λ が小さいと過学習が起こりやすい。 λ が大きいとfit が悪くなる。
 λ はCross Validation で最適なものを求める。

Training data に対する ROC curve



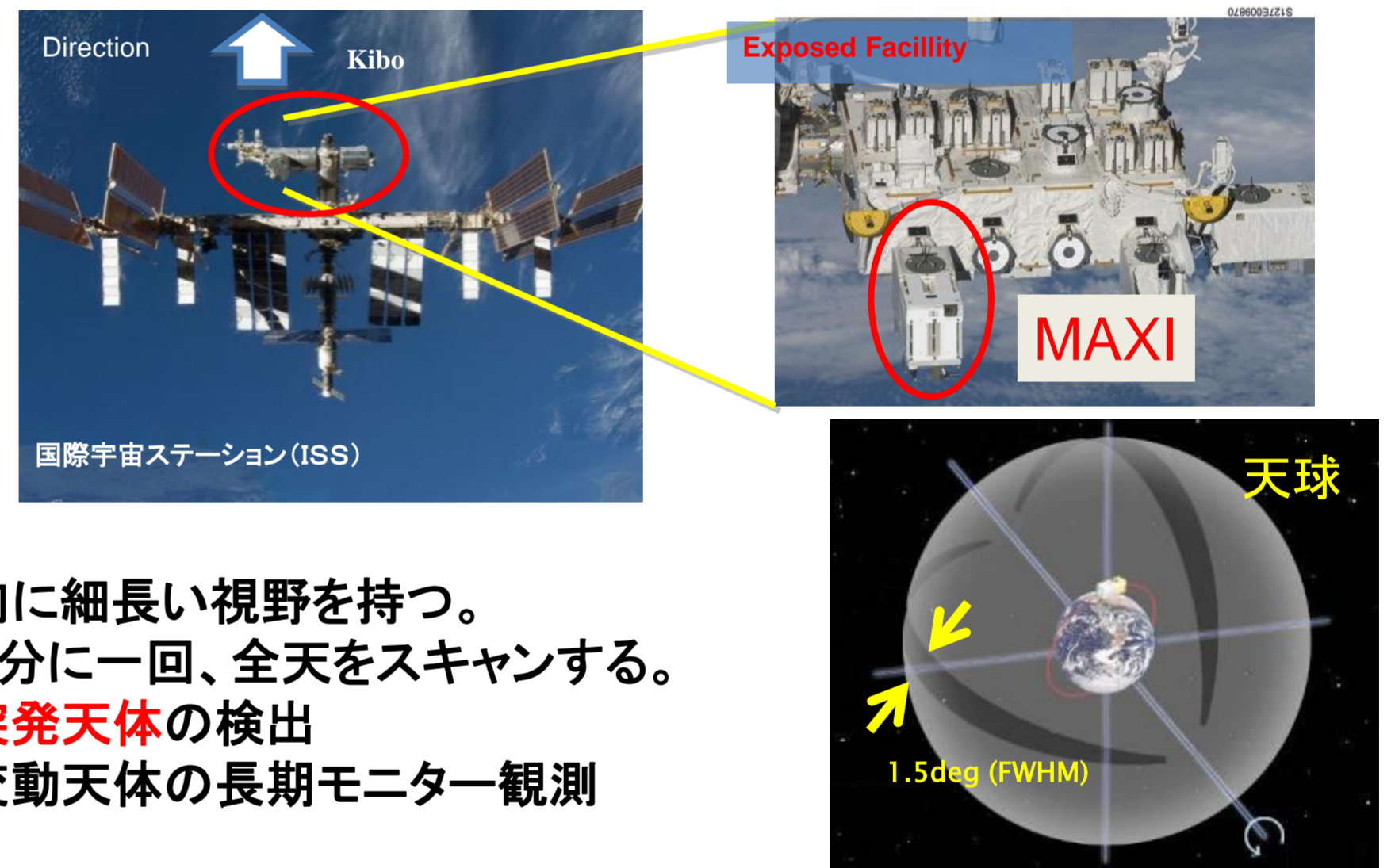
Test data に対する ROC curve



False Negative Rate = 0.1 のとき、
False Positive Rate = 0.09

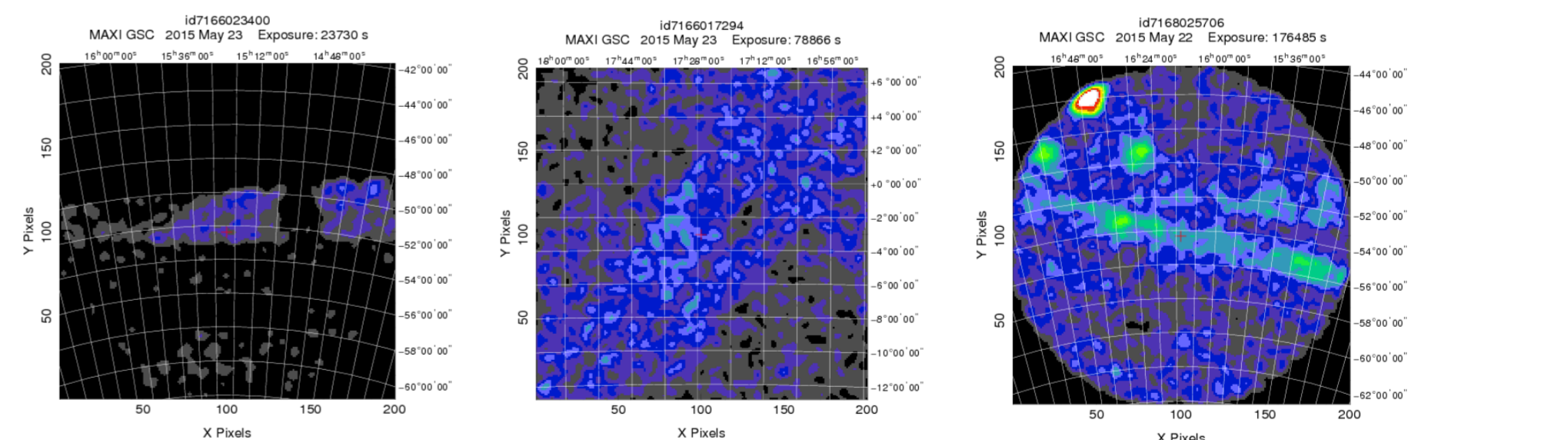
ROC curve が貼り付く。

MAXI (Monitor of All-sky X-ray Image)



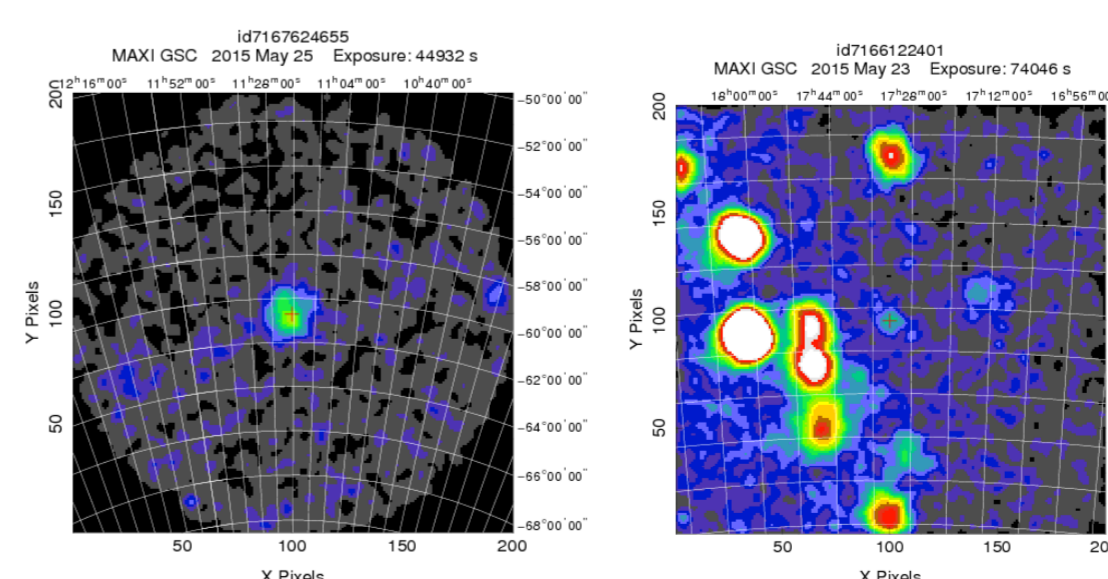
- 2方向に細長い視野を持つ。
- 約90分に一回、全天をスキャンする。
- X線突発天体の検出
- X線変動天体の長期モニター観測

偽物の例



Background が多い時間帯が存在していた。
カウントが多い領域が広がっている

本物の例



- すばる望遠鏡のHyper Suprime-Cam のデータから突発天体を探索する。
 - AUC Boosting の手法で良い結果が得られた。
 - さらに、特徴量の吟味が必要。
- 全天X線監視装置「MAXI」のデータ解析に応用し、偽イベントを除去したい。