

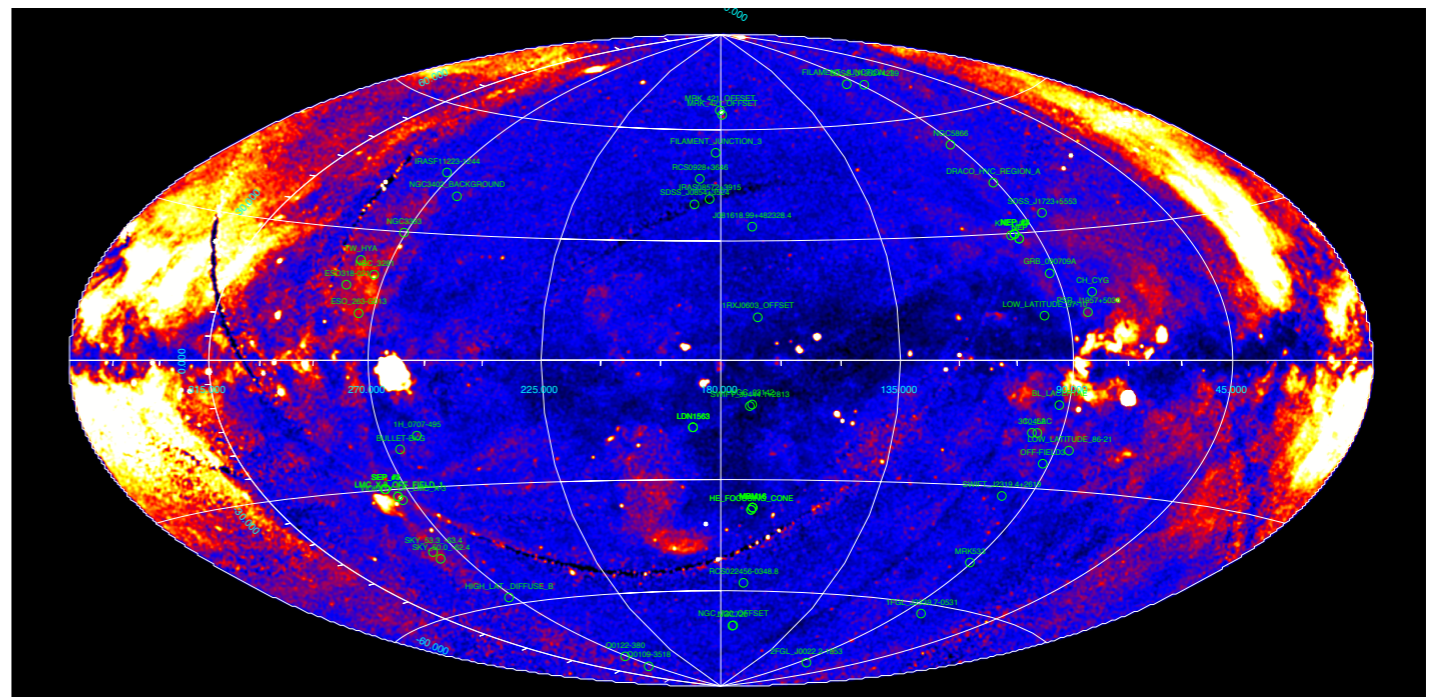
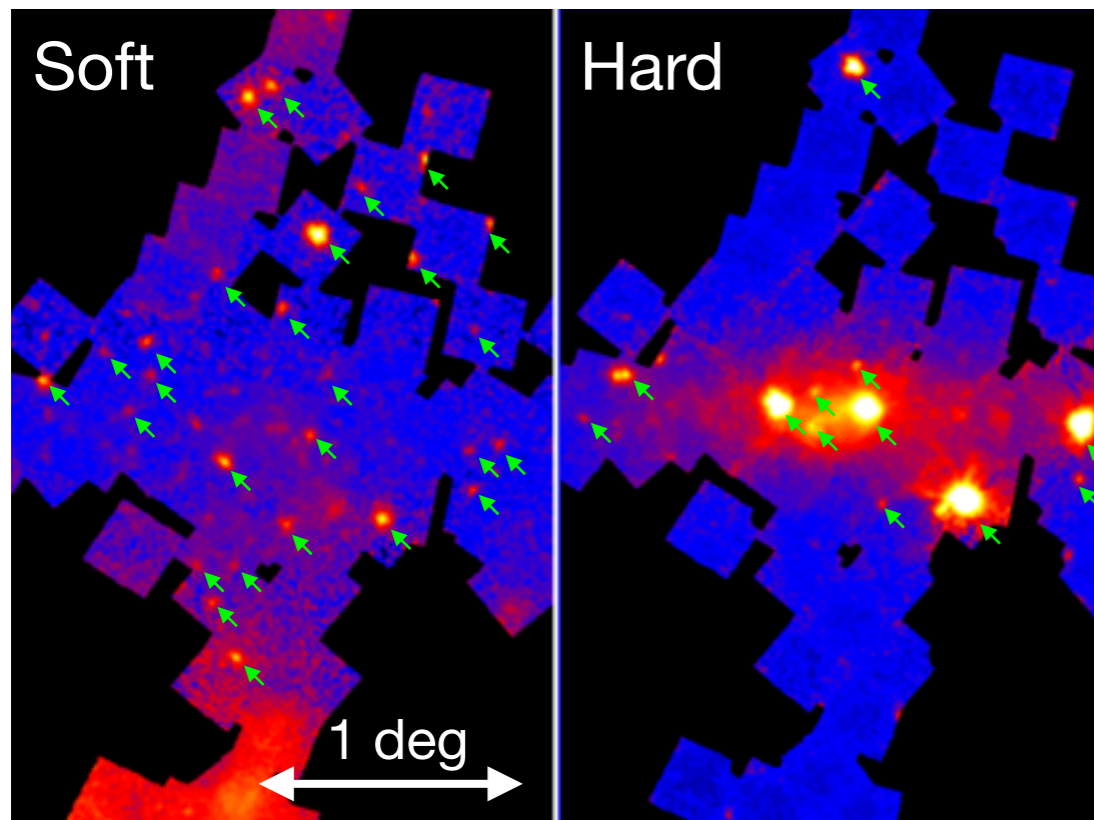
Deep Learning による 点源検出アルゴリズムの開発

中島 真也 (理化学研究所)

2017/09/25 第一回ASTRO-AI研究会@東北大学

モチベーション

- 私の興味は銀河系内の広がった高温プラズマ => 点源は除去してから解析する。
 - ▶ Galactic Center => 解析領域にさまざまなスペクトル・光度を持つ点源が密集
 - ▶ Galactic Hot Halo => 100以上の観測に対して統一的に点源除去作業が必要

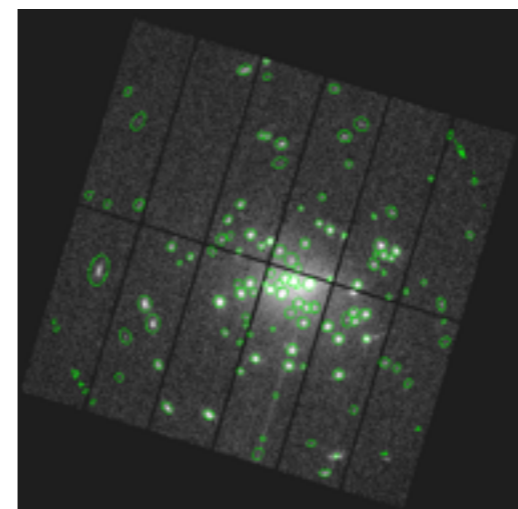
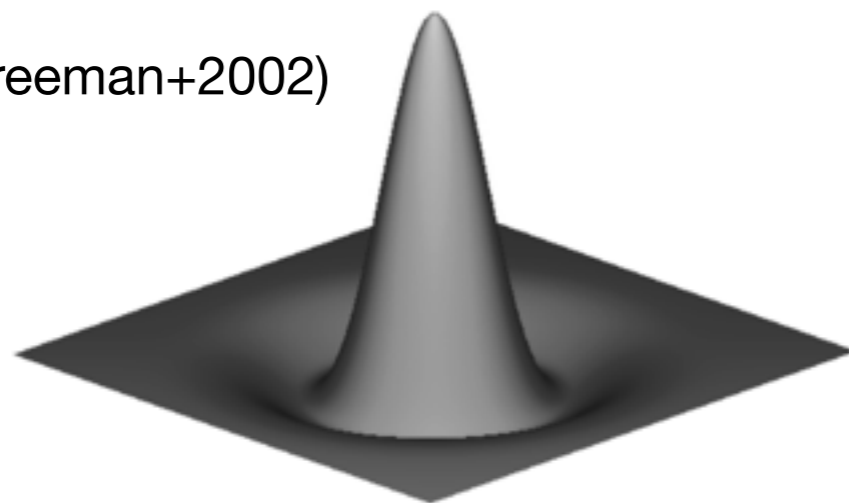


なんとかして自動化したい！

既存の手法

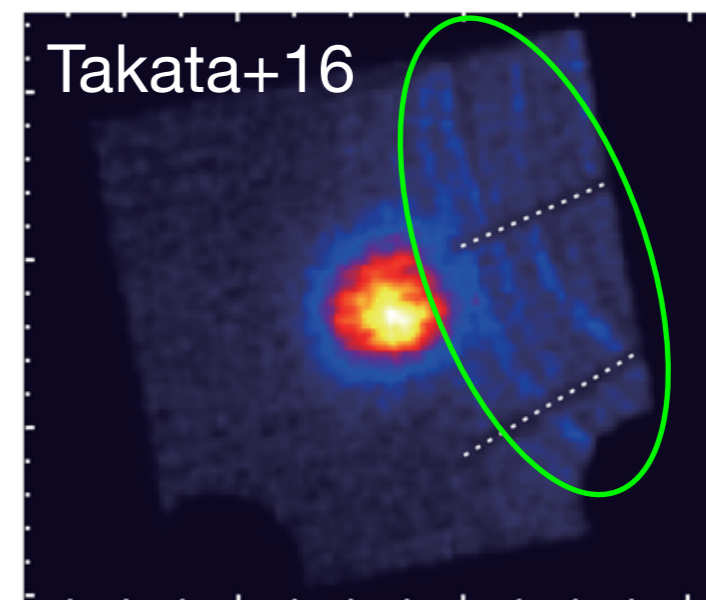
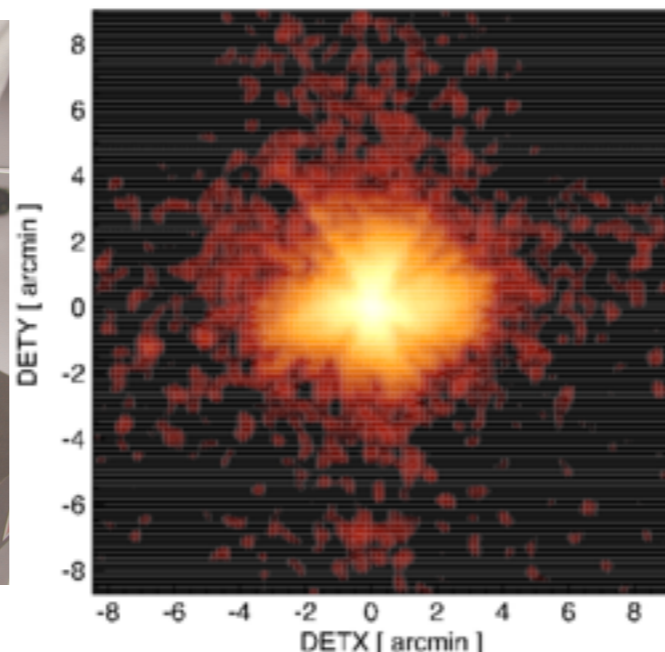
- wavelet function を使ったcorrelation解析 が一般的 (e.g., CIAOのwavdetect)

(Freeman+2002)



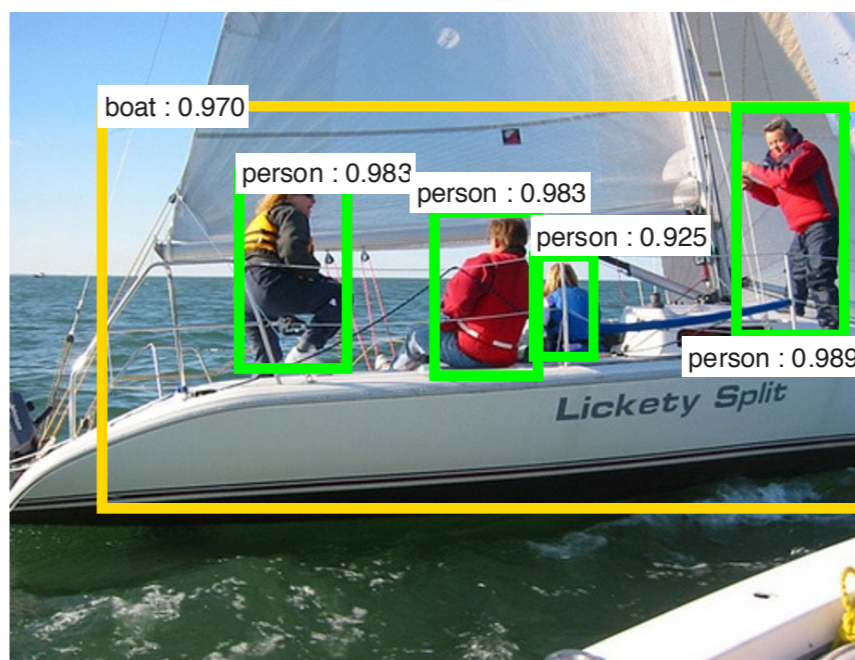
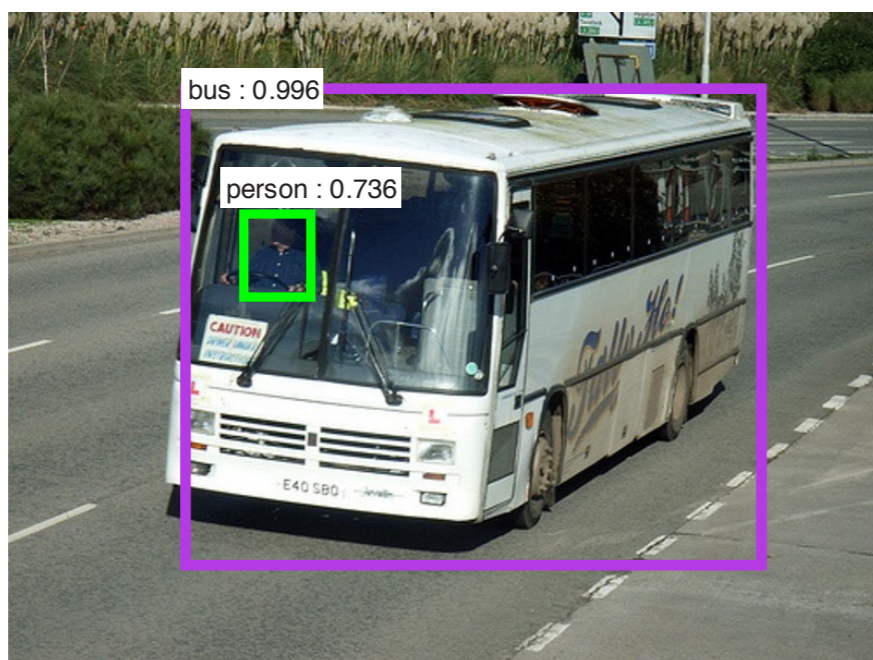
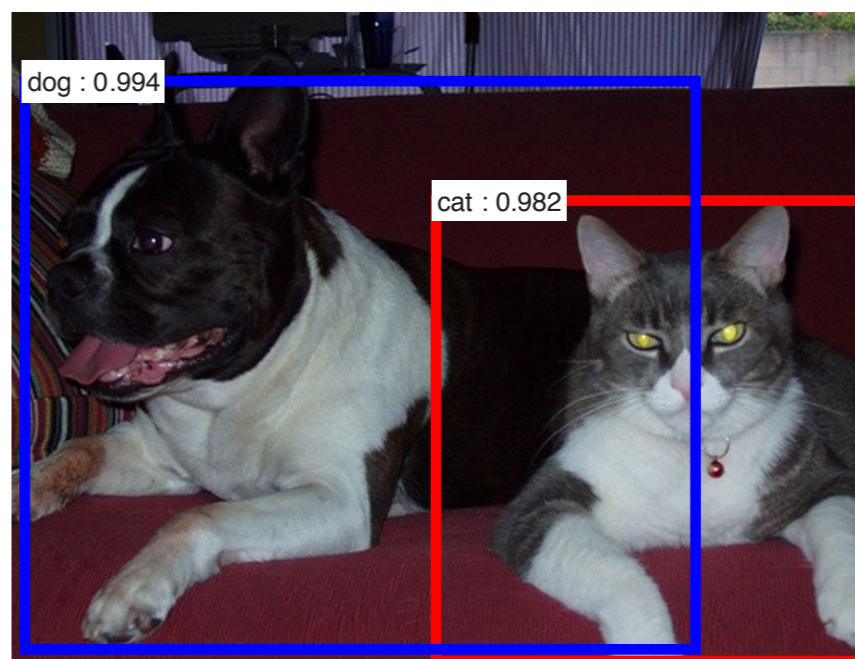
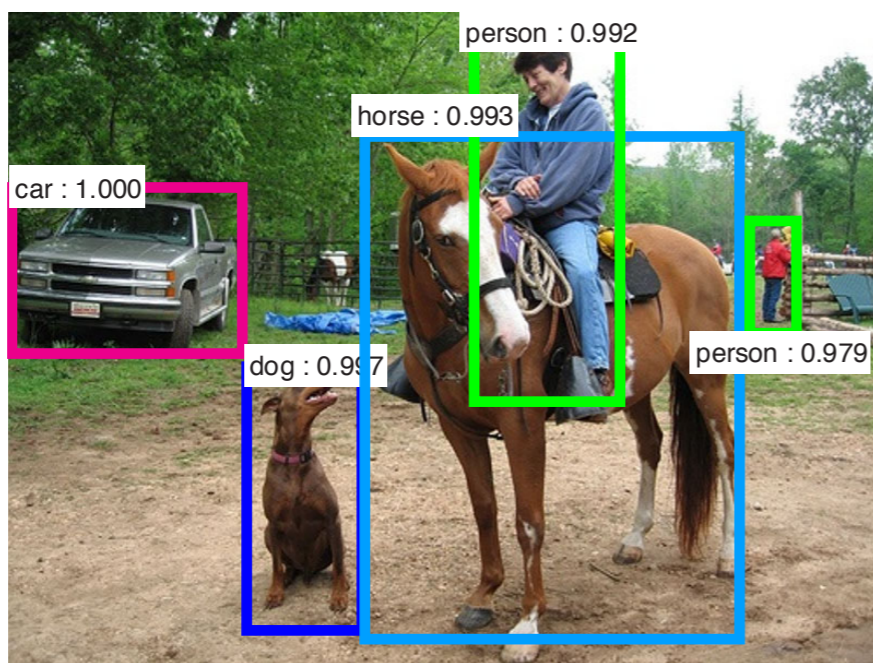
空間分解能の良い Chandra や XMM-Newtonでは実績あり

- すざくやひとみの point spread function は 同心円状ではない
- 視野外からの迷光の検出は難しい



Deep Learning による物体検出

↓ これを応用できないか？



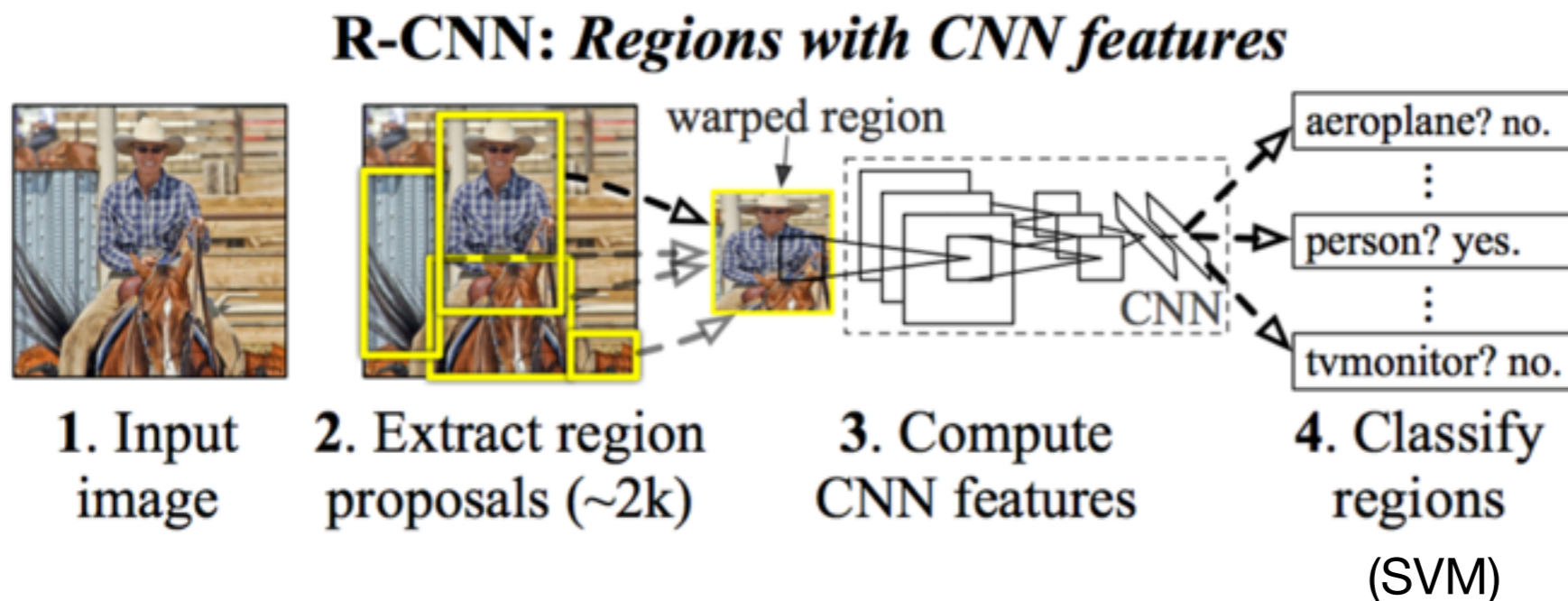
Input = Image

Output = Bounding Box + Label

Deep Learning による物体検出の手法

2013年：R-CNN (Regions with CNN features)

[<https://arxiv.org/abs/1311.2524>]

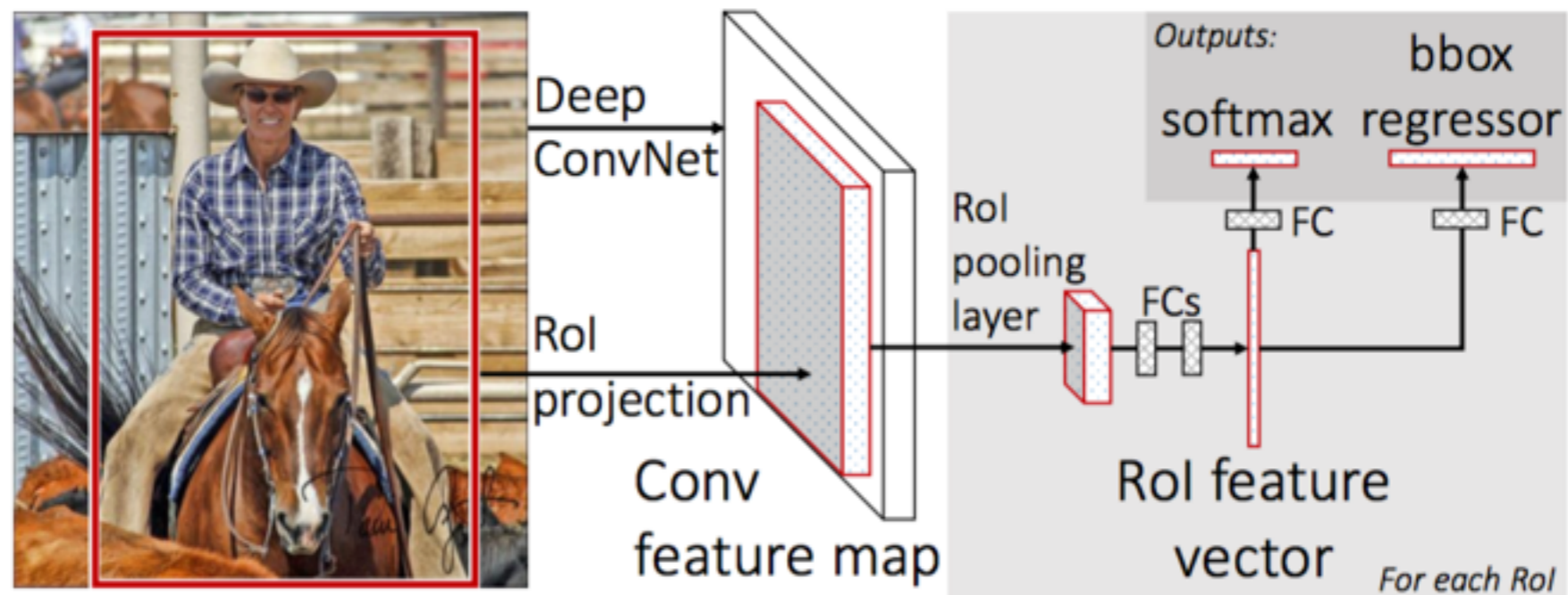


- あらかじめ別アルゴリズム (Selective Search) で領域候補 (region proposal) を多数抽出。
- 領域候補の画像をリサイズしてからCNNにかけて、特徴マップを作成。
- 特徴マップを Support Vector Machine (SVM) でカテゴリライズ。
- 物体であると判断した領域候補に対して、linear regression で bounding box のサイズを改訂。
- 問題点：遅い、各ステップを個別に学習させる必要有り。

Deep Learning による物体検出手法

2015年4月：Fast R-CNN

[<https://arxiv.org/abs/1504.08083>]

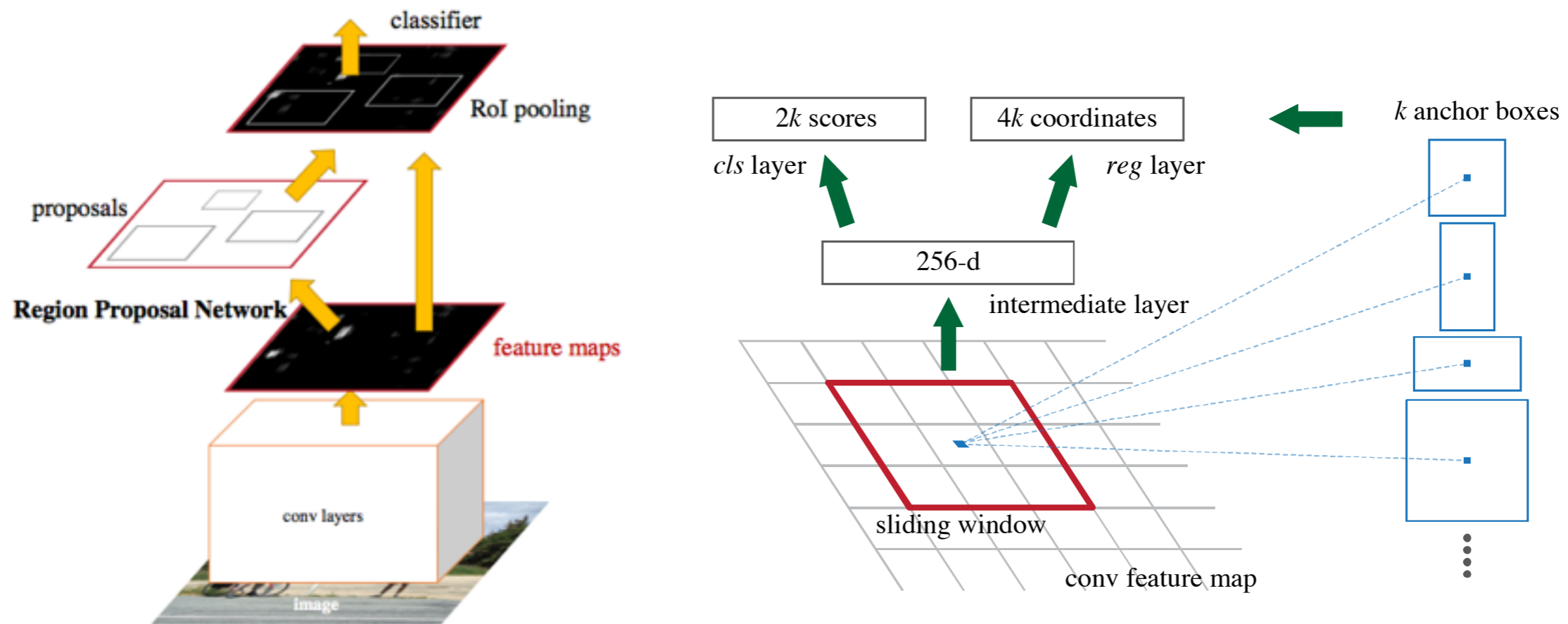


- Region of Interest (Rol) Pooling：領域候補ごとにCNNにかけるのではなく、画像全体をCNNにかけておいて、特徴マップに対して候補領域を射影する。=> 1回のCNN run で済む。
- 特徴抽出、カテゴリイズ、bounding box 推定を一つのモデルに統合。=> 学習が容易に。
- 問題点：領域候補検出は依然として別アルゴリズムで、ここが速度のボトルネック。

Deep Learning による物体検出の手法

2015年6月：Faster R-CNN

[<https://arxiv.org/abs/1506.01497>]

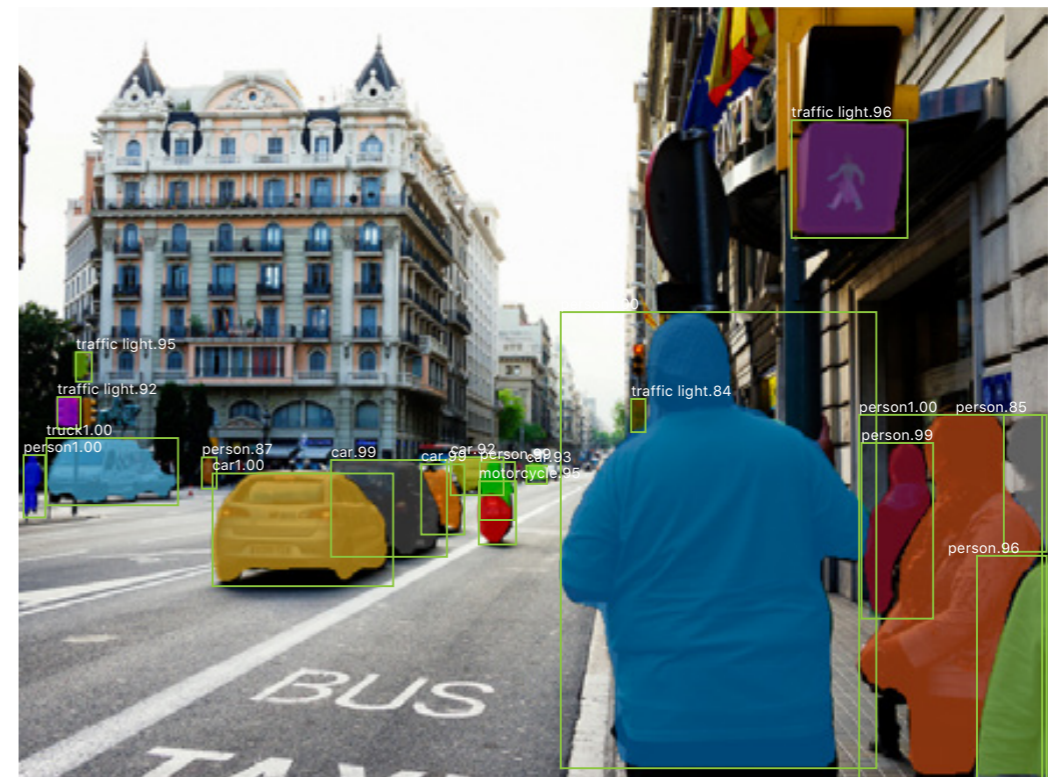
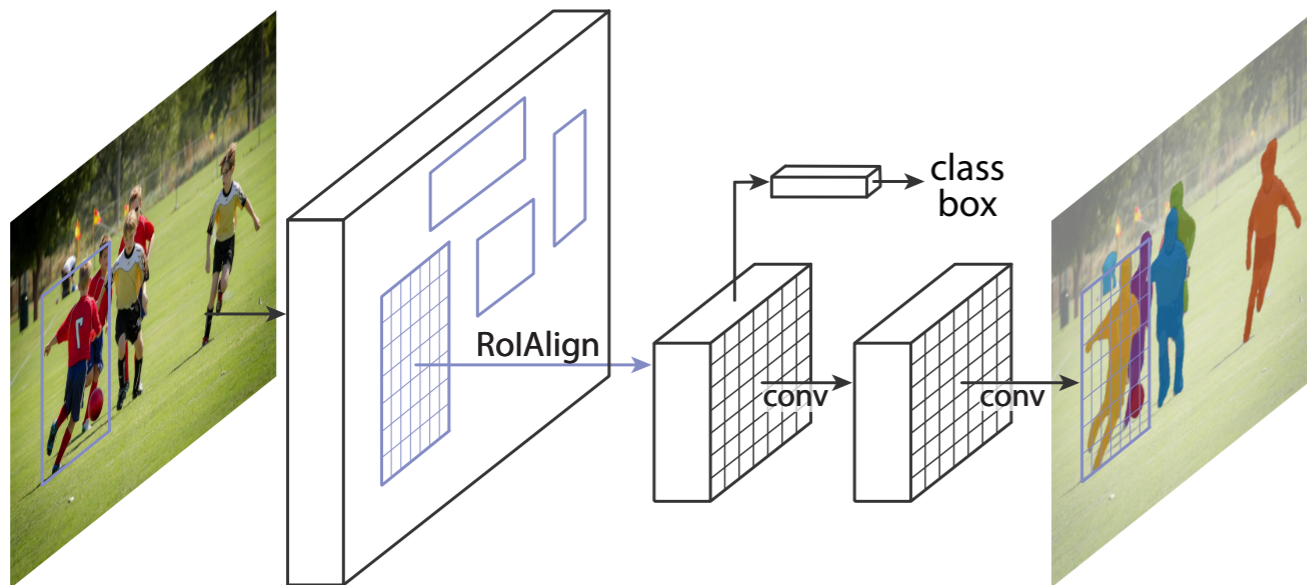


- Region Proposal Network：CNNの出力を領域候補の抽出にも使用することで、全プロセスを一つのニューラルネットに統合。

Deep Learning による物体検出の手法

2017年：Mask R-CNN

[<https://arxiv.org/abs/1703.06870>]



- classify, box regression の機能と並行して masking 機能を加えることで、ピクセルレベルで物体境界の識別ができるように。

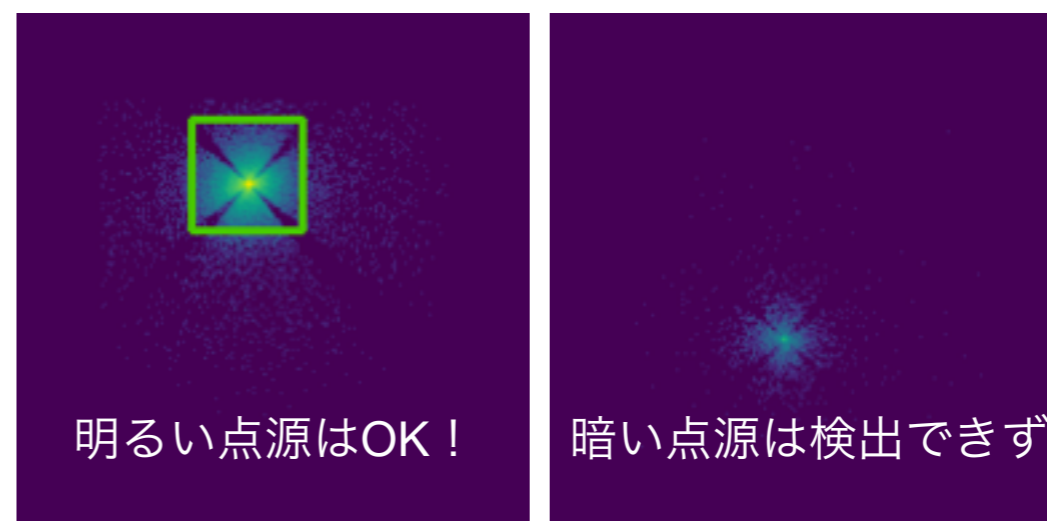
実際に試してみた

- GitHubから、Faster R-CNNの Keras実装をclone [<https://github.com/yhenon/keras-frcnn>]
- さまざまな点源画像をシミュレーションで作成 (1000枚)。
 - ▶ 簡単のため、FoV 中に一個の点源のみ。フラックスと位置を乱数で振る。
 - ▶ FITS image を 192 x 192 の png に変換。フラックスに合わせて bounding boxのサイズをスケール
- 上記画像でFaster R-CNNを訓練。同様に作成したシミュレーション画像でテスト

訓練画像 + bounding box 座標



識別テスト



訓練のパラメータを調整すれば、識別率は上がるはず

今後の課題

- FITSイメージを直接入力できるようにする。
- 領域を四角ではなく、円で定義できるようにする。
- 現実的なシミュレーションを行う (バックグラウンド+複数ソース)
- スペクトル情報も持たせたい。(e.g., Soft · Meidum · Hard の識別など)
- 計算環境を整える (CPUだと前ページの訓練ですら丸1日かかる。要GPU)
- X線代替機でpipeline processに組み込めれば、SXIの広視野を生かせるかも？