

全天球カメラにより配信される 正距円筒図法動画からの リアルタイム人物検出

首都大学東京

林田 和磨 横山 昌平

目次

- 研究背景、目的
- YOLOv3について
- 評価実験
- 考察、まとめ
- 今後の課題

背景

- 全天球カメラが安価に普及

Ex) RICOH THETA Z1 , GoPro HERO7,
NCTech iris360, Insta360 ONE X

→ 360度動画像対応のSNSが増加

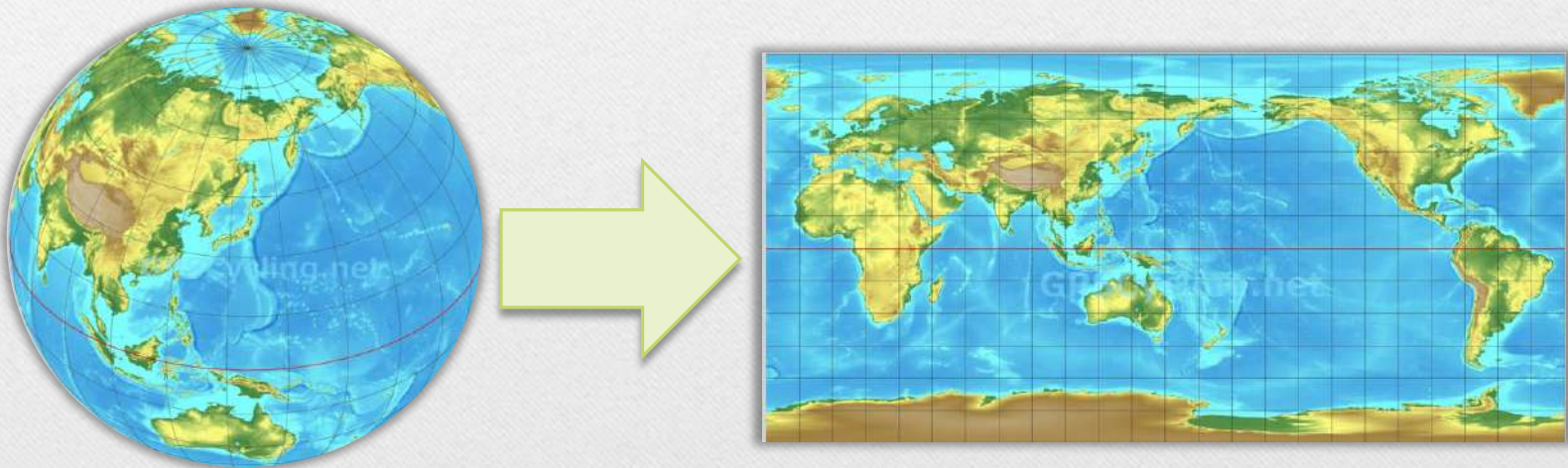
- 全天球カメラを用いる研究事例が少ない
- 取得できる情報量の多さ



正距円筒図法

- 全天球カメラで撮影された動画像は、正距円筒図法という投影法により全景を平面で見ることができる**全天球パノラマ動画像**に変換される

Ex) 地球(球面) →→ 世界地図(平面)



- 緯度が高くなるにつれて(全天球パノラマ動画像上の上部及び下部)、幾何学的歪みが発生し被写体が変形してしまう

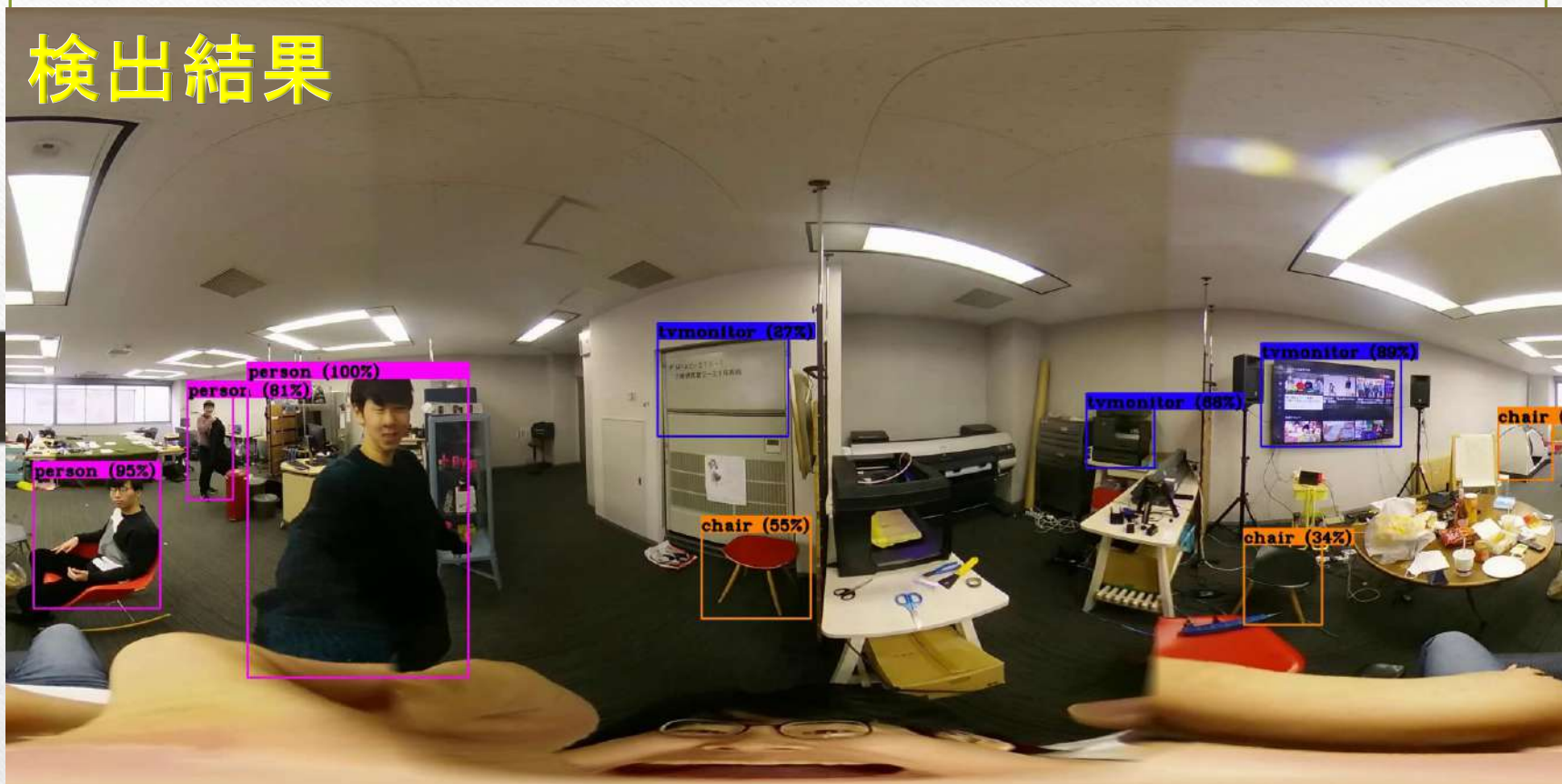
Ex) 日本を中心とした世界地図における南極大陸



YOLOv3

- single shot系の物体検出手法の一つ
 - 本研究では検出・学習において用いる
 - coco datasetを用いて学習したモデル
- 高緯度領域での人物検出精度が著しく低下

検出結果



高緯度領域において人物検出が困難に

関連研究

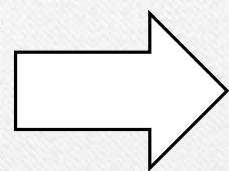
- 球面を回転してから正距円筒図法で変換することで被写体を低緯度領域(全天球パノラマ動画像上の中部)に移動させてから特徴量を抽出

- Robust feature matching for distorted projection by spherical cameras
 - Hajime Taira, Yuki Inoue, Akihiko Torii, and Masatoshi Okutomi
- 全天球カメラ内蔵ボールにおける視点固定手法
 - 中澤 正和 小池 英樹

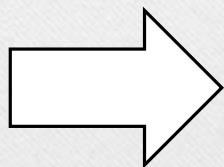
→処理速度の遅さが問題に

→アルゴリズムが複雑

本研究：**学習のみで解決**



全天球パノラマ動画像上で歪んで変形した
被写体を**そのまま学習**させる



通常の画像を用いて、**全天球パノラマ動画像上の
被写体の歪みを再現したもの**をデータセットとして
学習させる

学習方法

転移学習:

学習済みネットワークを固定し、新しいクラスに対し行う学習手法
→高緯度領域で歪んでいる人を別のクラス(ラベル)として学習

ファインチューニング:

学習済みネットワークを固定し、既存のクラスに対して一部の重みを再度学習する手法
→高緯度領域で歪んでいる人を既存の人クラス(ラベル)として学習

本研究で作成した人物検出モデル4つ

検出器1: 全天球パノラマ動画像上の歪んだ被写体を転移学習を用いて学習

検出器2: 全天球パノラマ動画像上の歪んだ被写体をファインチューニングを用いて学習

検出器3: 通常の動画像をファインチューニングを用いて学習

検出器4: 通常の動画像を変形した画像をデータセットとし、ファインチューニングを用いて学習

- 全天球パノラマ動画像: 2679枚 (1名対象)
- 通常の動画像: 9144枚 (4名対象)

ラベル付け(アノテーション)にはpython上で実装されるLabelImgを用いた

検出器4

- 全天球パノラマ動画像のオープンデータセットが非常に少ない。
→通常動画像を用いて全天球パノラマ動画像上の歪み部分を再現
- 球面の一部に通常画像をマッピングし、極に近い位置で正距円筒図法を用いて変換する。
- 球面マッピングには、様々な入出力形式と仰角方位角が操作可能なCube2DMというアプリケーションを使用した。
- 学習に与える画像は、撮影した通常画像のうち1952枚採用した。

変換後の画像



評価実験

RICOH THETA Z1を用いて、入力する全天球パノラマ動画を3つ用意した。

入力動画1:全天球パノラマ画像下部(高緯度領域中央)に人が映っている動画
入力動画2:全天球パノラマ画像下部(高緯度領域両端)に人が映っている動画
入力動画3:全天球パノラマ画像下部, 中央部, 上部に人が映っている動画

作成した4つの検出器モデルとYOLOv3の学習済みモデル、合計5つのモデルそれぞれに動画を入力し、認識精度を記録する。

入力動画1

高緯度領域・画像下部中央に人が映る動画



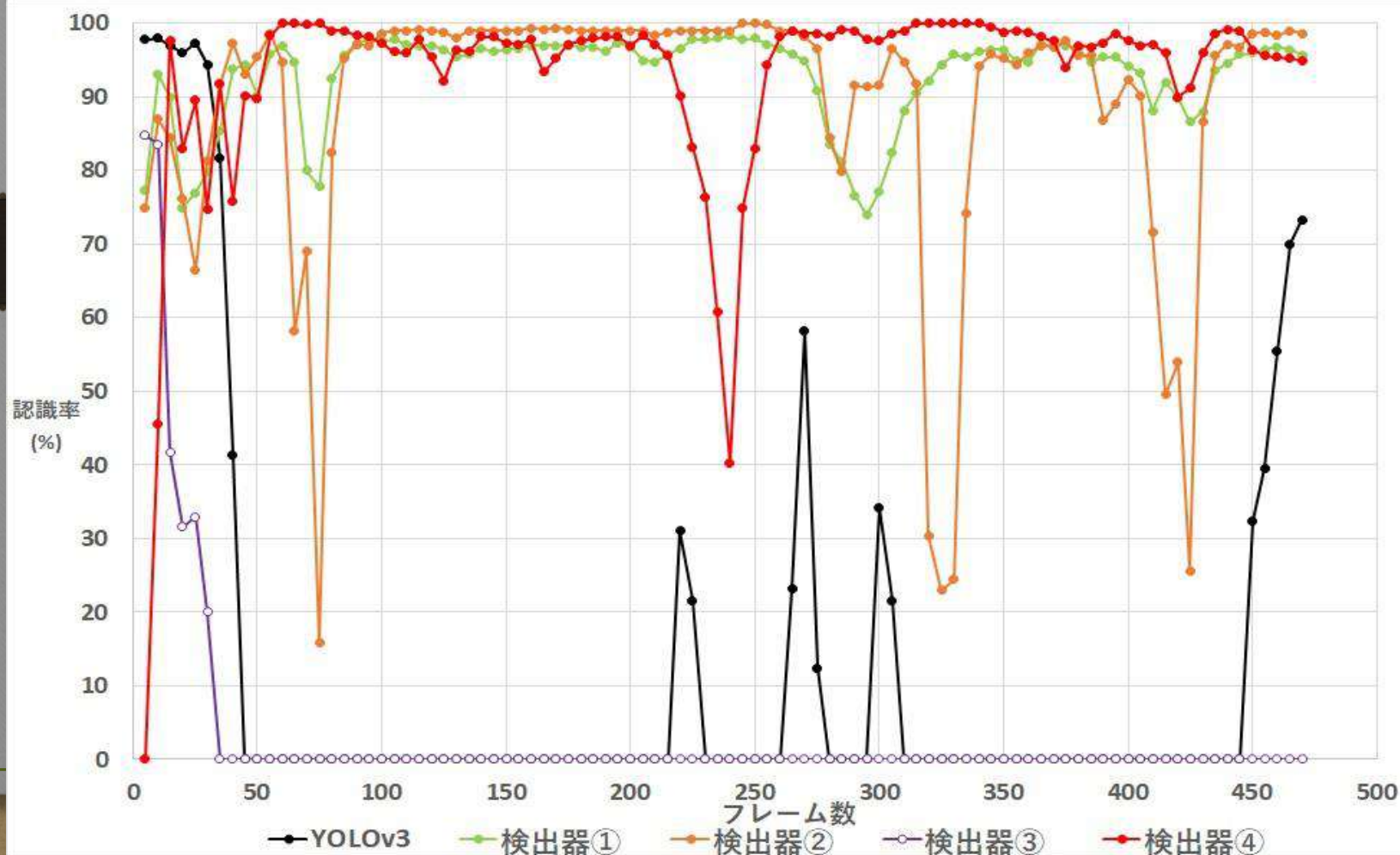
YOLOv3 : coco dataset 既存の学習済みモデル

検出器1:全天球パノラマ動画像 転移学習

検出器2:全天球パノラマ動画像 ファインチューニング

検出器3:通常画像 ファインチューニング

検出器4:通常画像で歪み再現 ファインチューニング



入力動画2

高緯度領域・画像下部両端に人が映る動画



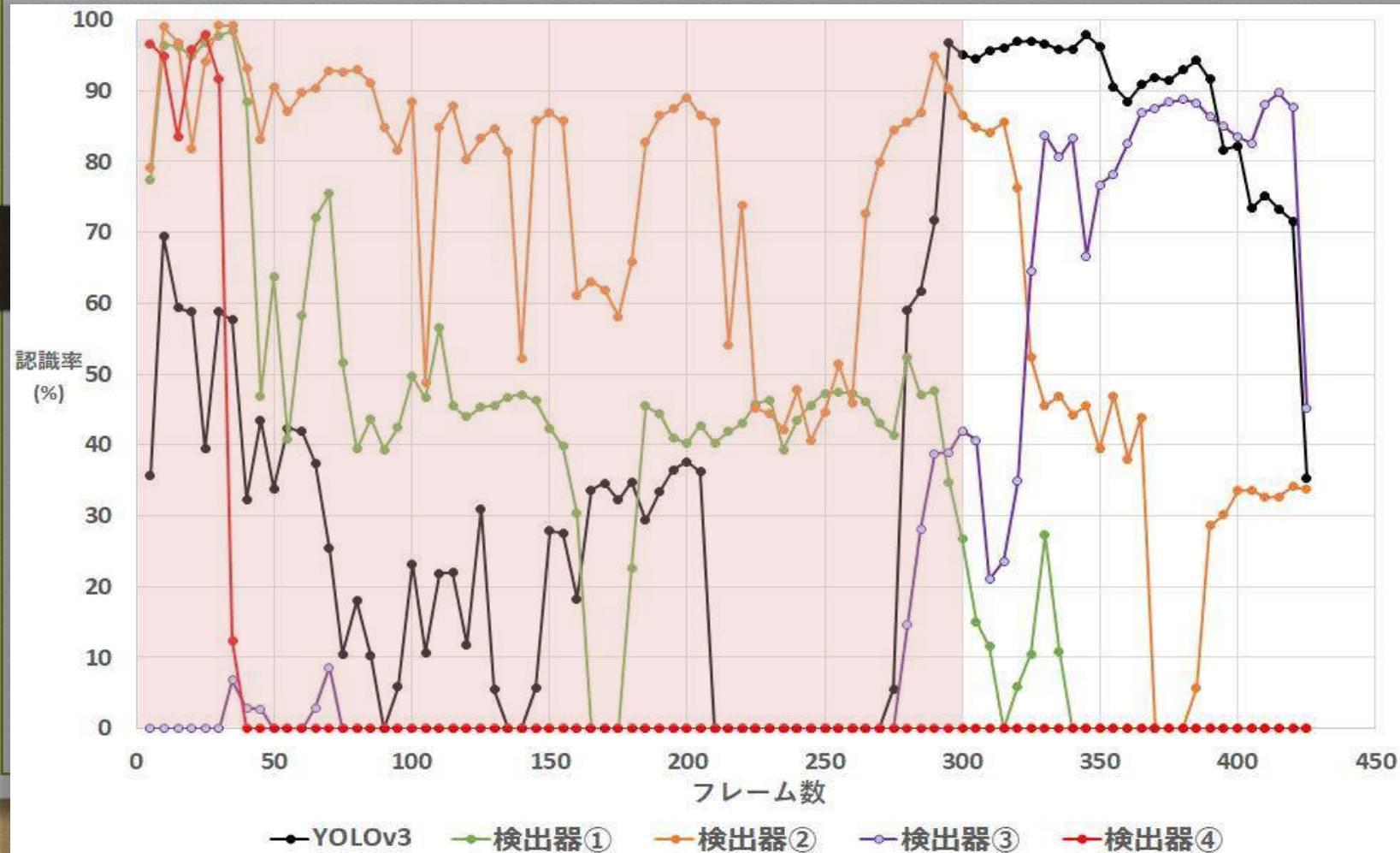
YOLOv3 : coco dataset 既存の学習済みモデル

検出器1:全天球パノラマ動画像 転移学習

検出器2:全天球パノラマ動画像 ファインチューニング

検出器3:通常画像 ファインチューニング

検出器4:通常画像で歪み再現 ファインチューニング



入力動画3



高緯度領域・画像上部に上半身が映る動画

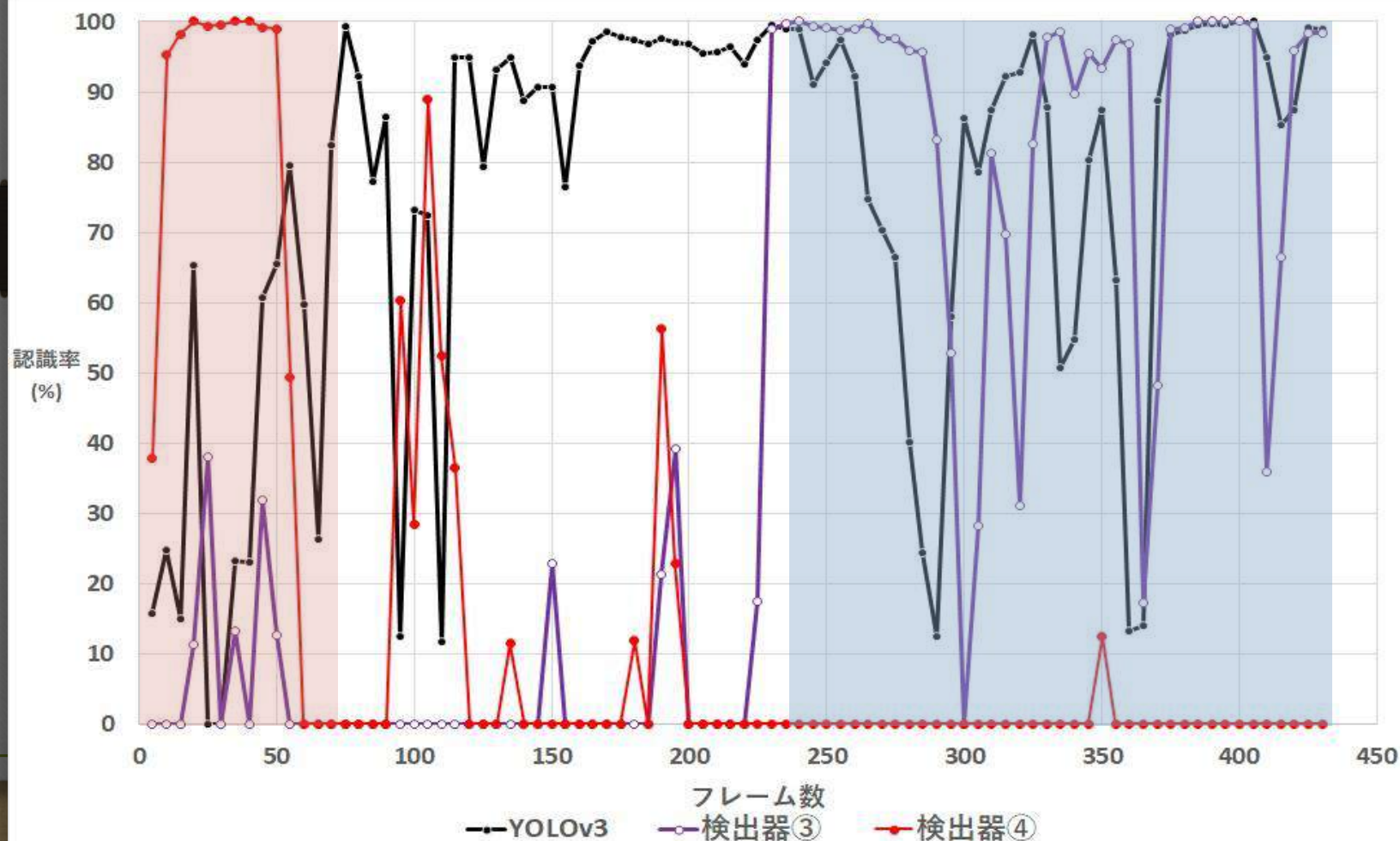
YOLOv3 : coco dataset 既存の学習済みモデル

検出器1:全天球パノラマ動画像 転移学習

検出器2:全天球パノラマ動画像 ファインチューニング

検出器3:通常画像 ファインチューニング

検出器4:通常画像で歪み再現 ファインチューニング



考察

- 全天球パノラマ動画像をそのまま学習させた検出器1、2、また通常画像で歪みを再現した画像を学習させた検出器4は、**高緯度領域・画像下部において検出可能**になった。
- 全天球パノラマ動画像を**ファインチューニングで学習させる**やり方が学習効率が良い。
- 動画像上の位置や被写体とカメラの距離、歪み方が変わると検出が難しくなる。
- **体の一部など特徴的な部分**が低緯度領域にあれば既存の手法でも十分検出できる。

まとめ

- 全天球パノラマ動画像・高緯度領域で歪んで変形した人物に対する検出を行った。
- 歪んで変形した被写体をそのままデータセットとして学習させることで解決。
- 全天球パノラマ動画像をファインチューニングで学習させるやり方が最も効率が良い。
- どの検出器を学習させる場合でも、映り方が様々なパターンあるデータセットで学習できれば、十分に検出できる可能性がある。

今後の課題

- データセットの増強、追加学習
- YOLOv3の既存モデルが学習したcoco datasetと全天球パノラマ動画像を合わせた学習
 - 高緯度領域でも低緯度領域でも十分検出できるモデル
- この手法を用いて監視カメラシステムなどへ応用
- 無線接続状態下でのリアルタイム物体検出を行うシステム構築