



DIVE INTO CODE

# 機械学習エンジニアコース 卒業課題

## テーマ：物体追跡 (Object Tracking)

2020年 4月期：井伊 喬稔

## 自己紹介：

名前：井伊 喬稔（イイ タカトシ）

～2019年3月 貨幣計数機メーカー子会社にてPG/SEとして勤務

取引先：流通(デパート、百貨店)

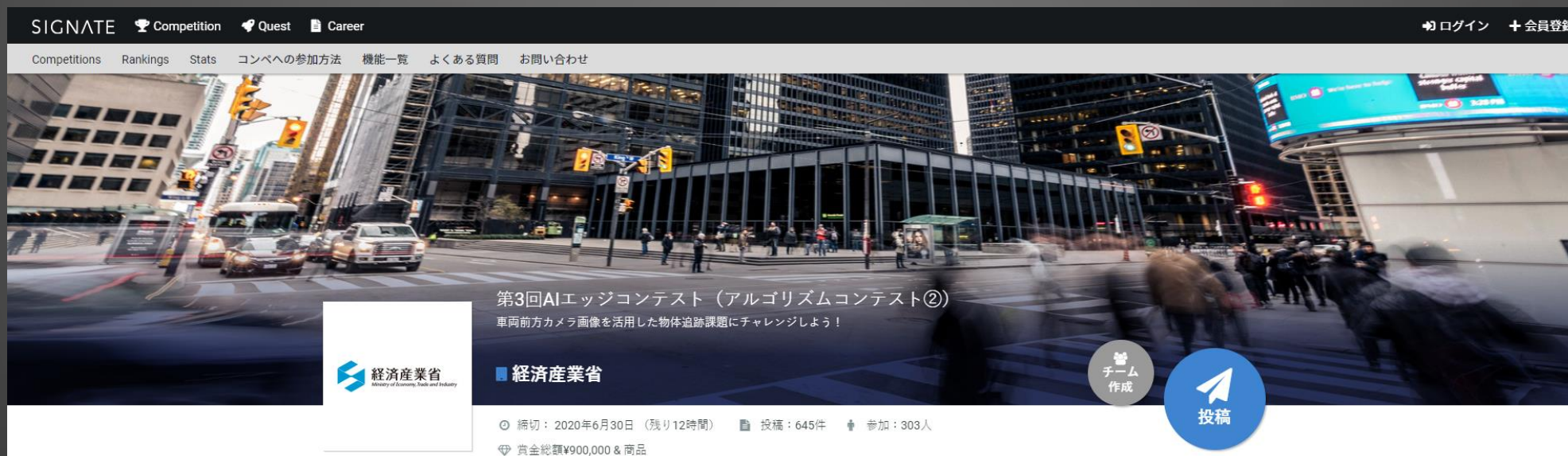
金融(大手銀行及び各地銀)

2018年 JDLA G検定/E資格取得

2019年9月～  
2020年3月 AI関連企業にてSESとして勤務

2020年4月～ 機械学習を極める為、Dive Into Code入校

# <卒業課題>



SIGNATE Competition Quest Career ログイン 会員登録

Competitions Rankings Stats コンペへの参加方法 機能一覧 よくある質問 お問い合わせ

第3回AIエッジコンテスト (アルゴリズムコンテスト②)  
車両前方カメラ画像を活用した物体追跡課題にチャレンジしよう！

経済産業省  
Ministry of Economy, Trade and Industry

■ 経済産業省

チーム作成

投稿

締切: 2020年6月30日 (残り12時間) 投稿: 645件 参加: 303人  
賞金総額¥900,000 & 商品

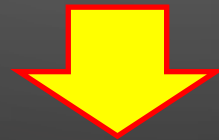
## SIGNATE : 第3回AIエッジコンテスト (アルゴリズムコンテスト②)

6/30 コンペ終了



# <なぜ物体追跡を卒業課題として選択したか>

- 現在、各AIメーカーが挙って物体検知のシステムを開発/販売しており、ブームの真っただ中にある。
- 物体検知には、まだまだ今後の伸びしろが残されている。
- 物体検知よりも、物体追跡の方が難易度は高くなるが、将来性があると考えた。
- 物体追跡によって、自動運転の観点では、前の車に追従して移動するといった事が可能となる。
- AIエッジコンテストは一つのコンペが終了しても、翌日には次回のコンペが開催される為、モチベーションの維持につながる。

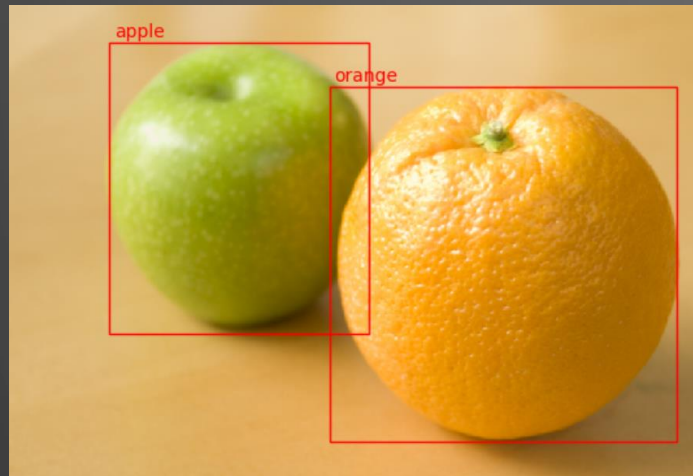


今のタイミングで機械学習の最先端である、物体検知/追跡を理解しておくことによって、今後 機械学習エンジニアとして活躍するための橋渡しになるのではないかと！



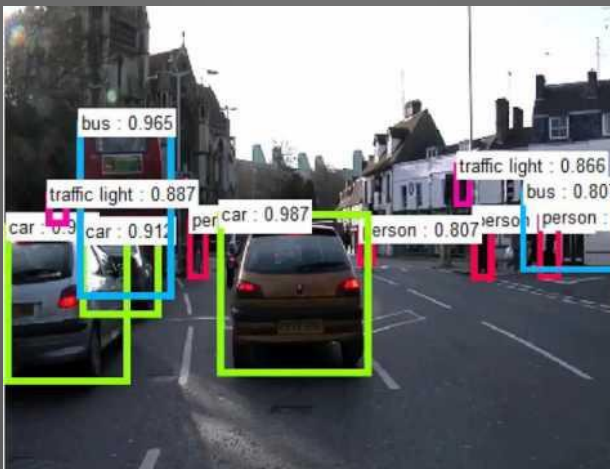
# <物体検知のこれから（想像）>

（過去）



過去に撮影した  
静止画に対して物体検知

（現在）



カメラからの画像に対して  
リアルタイムで推定を行い  
物体検知/追跡

（未来）

## 物体移動推定？



（予想）

物体のこれまでの移動奇跡から未来  
の物体の推定位置/動作を予測する。

（実用例）

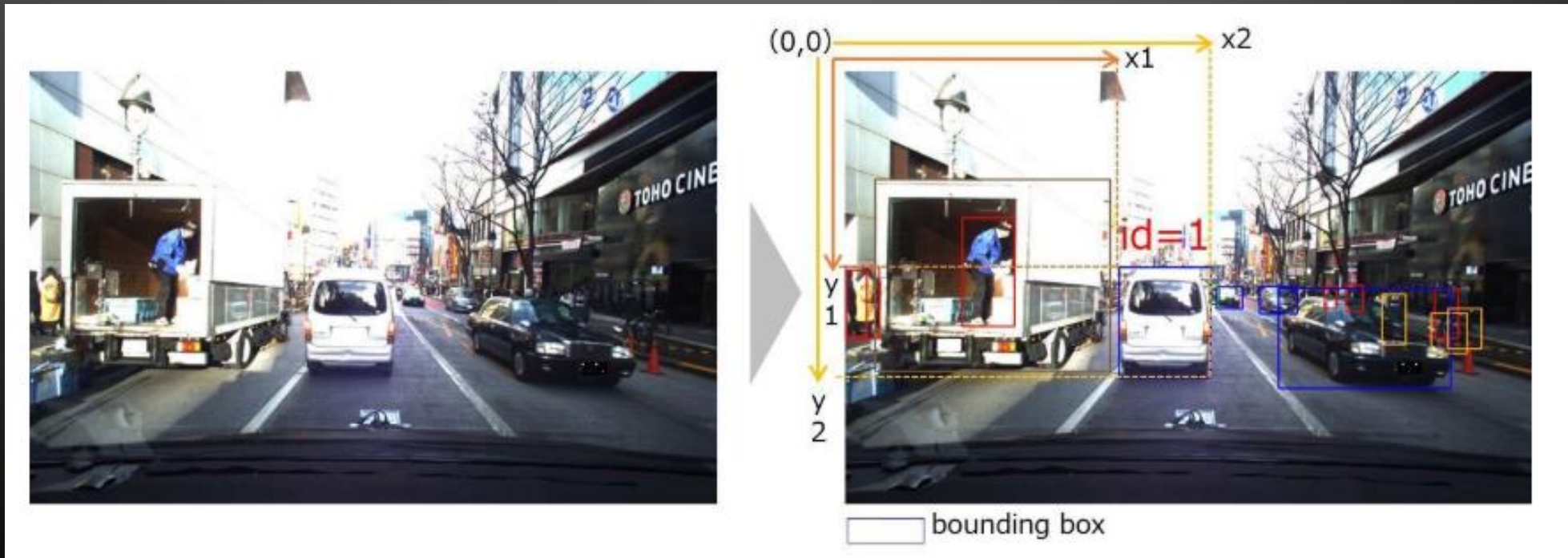
- ・ 万引き事前予測
- ・ 交通事故回避
- ・ マイノリティ・リポート？
- ・ 迎撃ミサイル？

## まだまだ発展していく余地はありそう



コンペ内容：

車両前方カメラで撮影された動画に対して、予測対象の物体を含む矩形領域をbounding box =  $(x1, y1, x2, y2)$  として割り当て、さらにそれぞれの動画内の同一の物体に対して、任意のユニークなオブジェクトIDを付与する。



## <データ>

- 学習用動画 : 25本 (train\_00.mp4~train24.mp4)  
時間 : 2分 (120秒) /本  
フレームレート : 5fps  
フレーム数 : 600フレーム/本
- アノテーションデータ : 25本 (train\_00.json~train\_24.json)

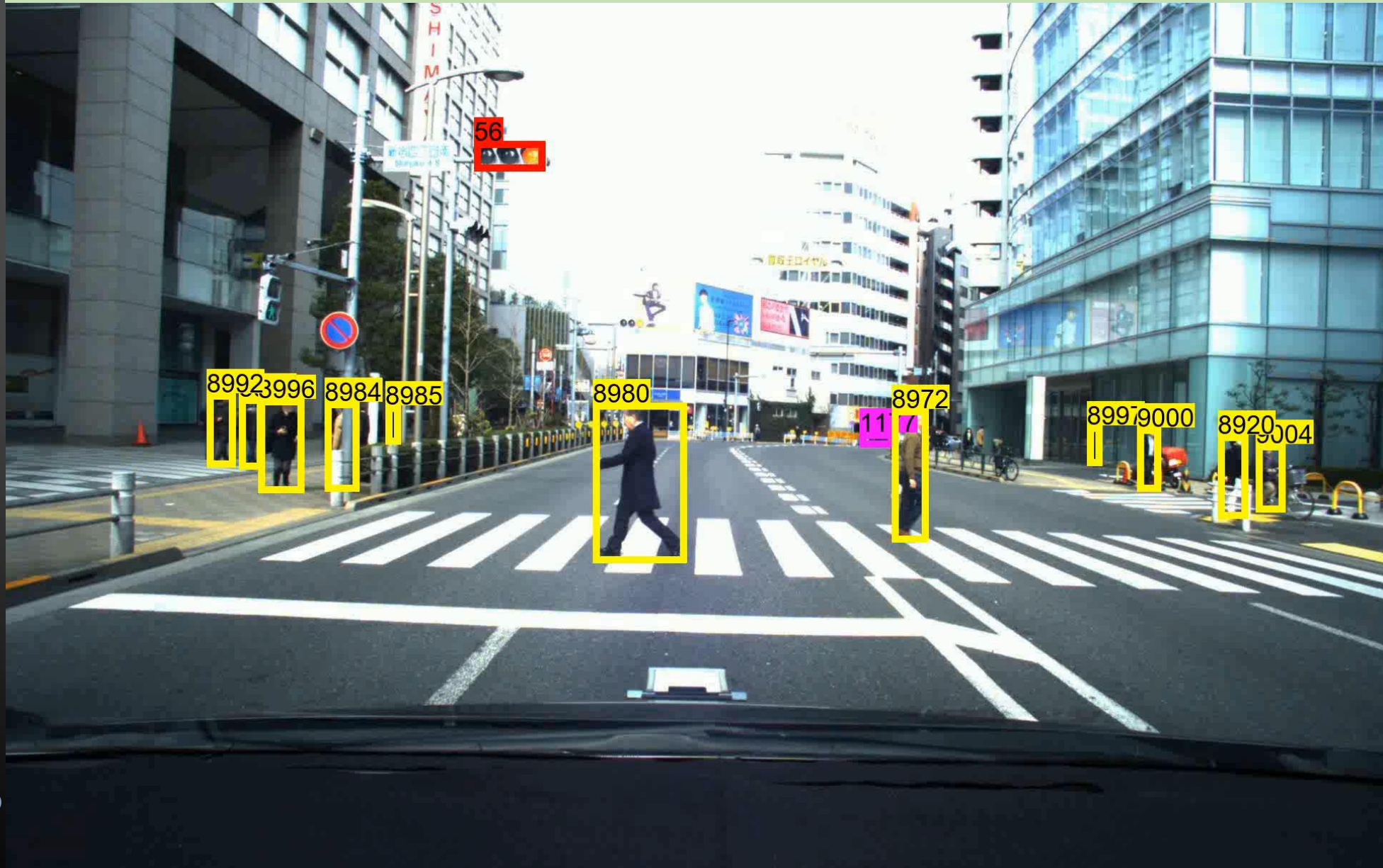
以降、順位2位のIRAFM-AI teamの解法を元に説明する。

(参考)[https://gitlab.com/irafm-ai/signate\\_3rd\\_ai\\_edge\\_competition](https://gitlab.com/irafm-ai/signate_3rd_ai_edge_competition)

(参考) [https://gitlab.com/irafm-ai/signate\\_3rd\\_ai\\_edge\\_competition/-/blob/master/readme.pdf](https://gitlab.com/irafm-ai/signate_3rd_ai_edge_competition/-/blob/master/readme.pdf)

# 【提供データ（目標とする動画）】

<https://drive.google.com/file/d/1OBF-q0WVChIQH6NI50L2VmAyU6gaMBSC/view?usp=sharing>



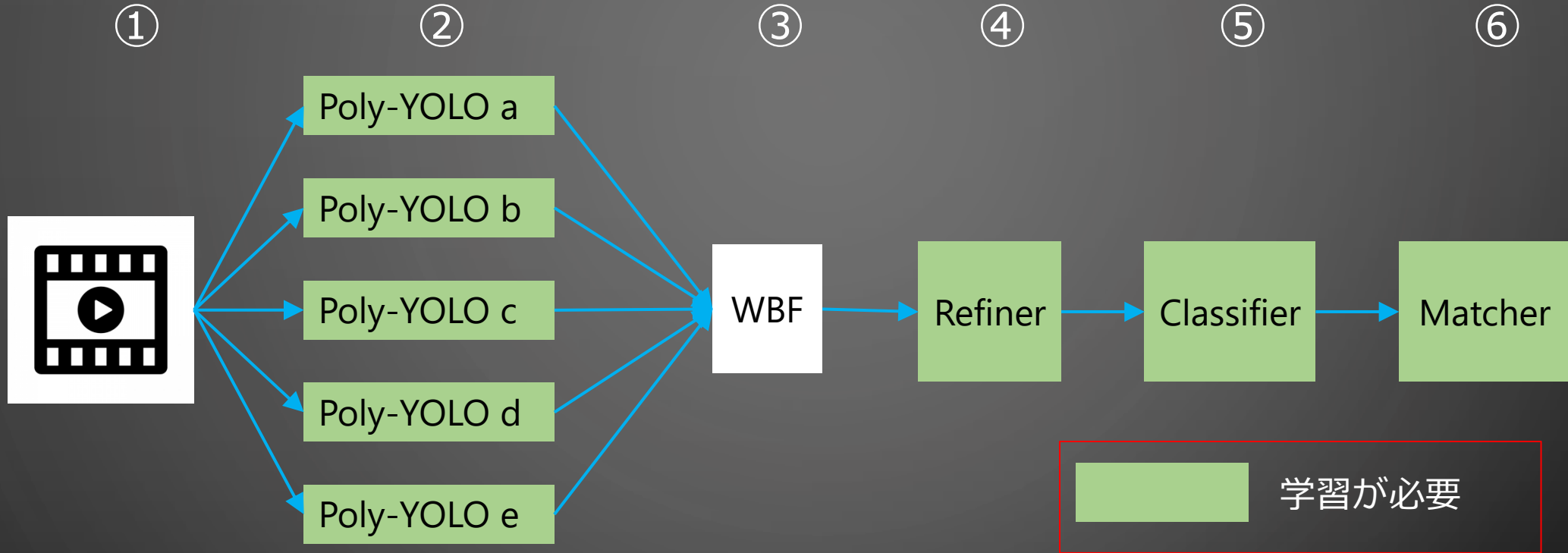


## 検知対象：

- ・ カテゴリは、「乗用車」「歩行者」の2種類
- ・ それぞれの動画で、3フレーム以上写っている物体  
(フレームは連続している必要は無い)
- ・ 矩形の大きさが $1024\text{pix}^2$ 以上の物体

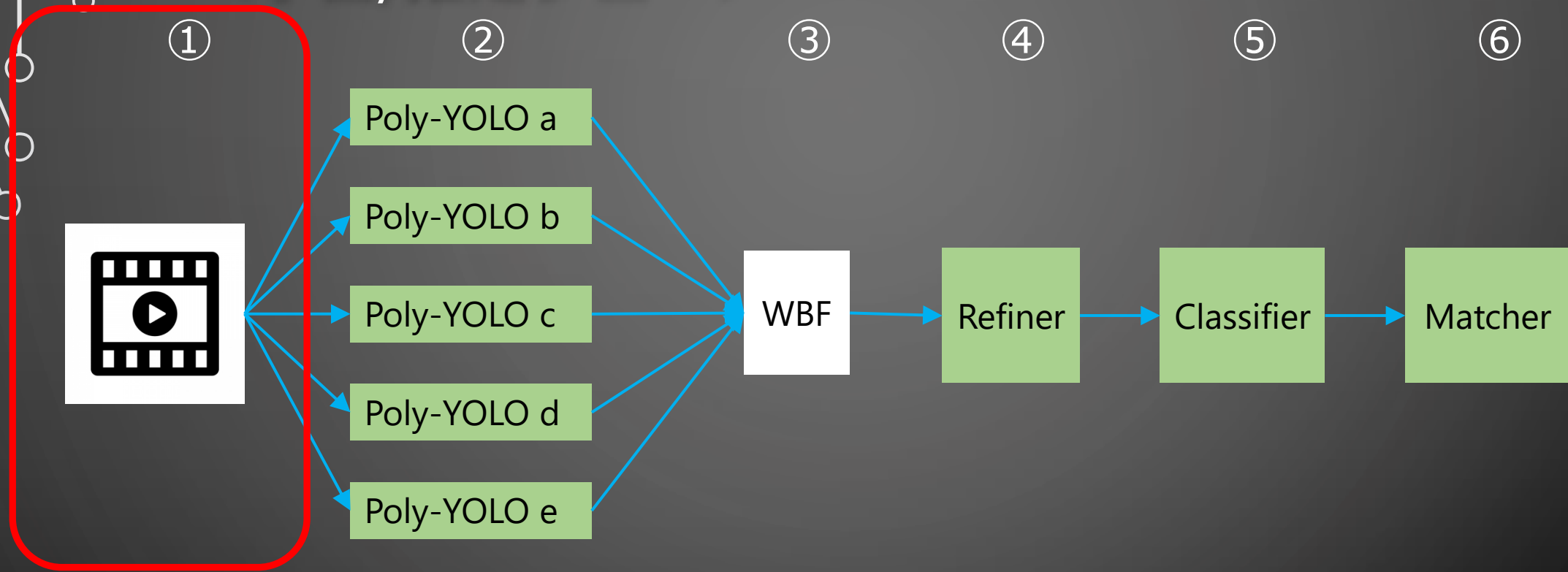


# <学習/推定フロー>



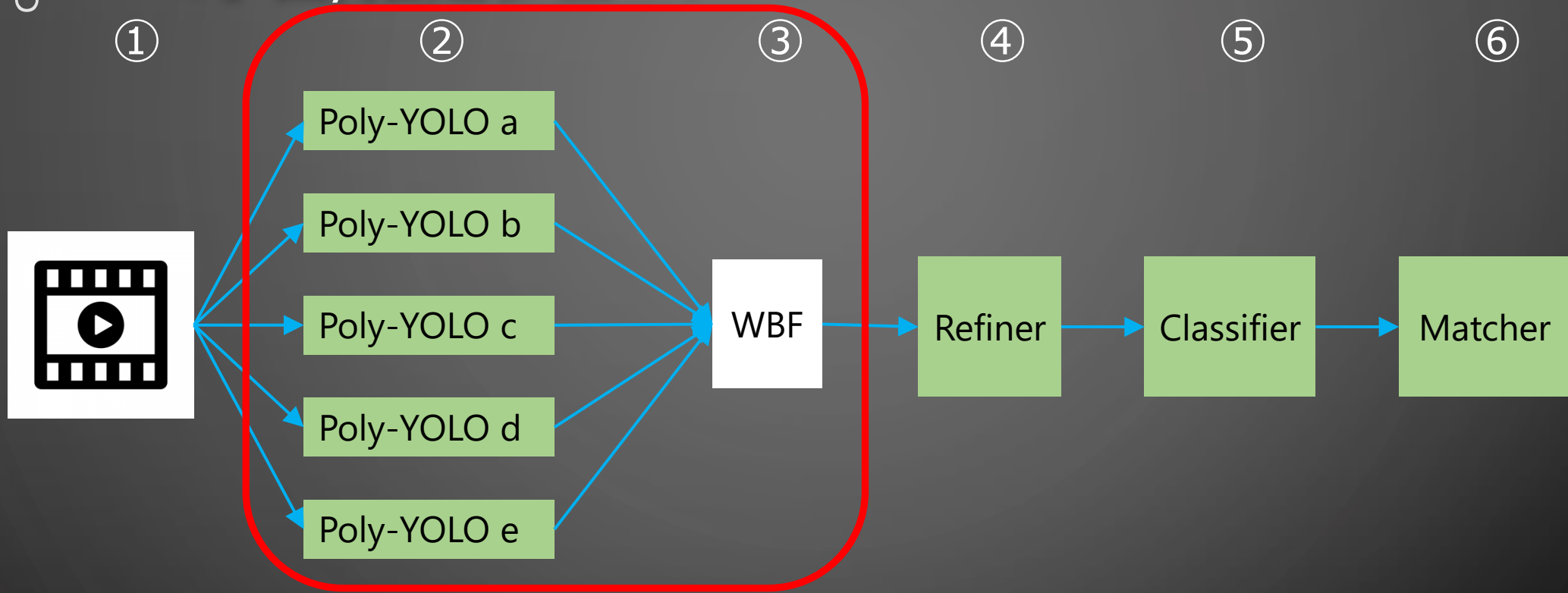
- ①入力動画をフレーム単位で静止画に切り出し
- ②複数のPoly-YOLO検出器で学習
- ③WBF(Weighted Box Fusion)でアンサンブル学習
- ④Refiner:検出境界を高精度化
- ⑤クラス分類(自動車、歩行者 分類)
- ⑥フレーム間オブジェクトの関連付け(追跡処理)

# <学習/推定フロー>



- ①入力動画をフレーム単位で静止画に切り出し
- ②複数のPoly-YOLO検出器で学習
- ③WBF(Weighted Box Fusion)でアンサンブル学習
- ④Refiner:検出境界を高精度化
- ⑤クラス分類(自動車、歩行者 分類)
- ⑥フレーム間オブジェクトの関連付け(追跡処理)

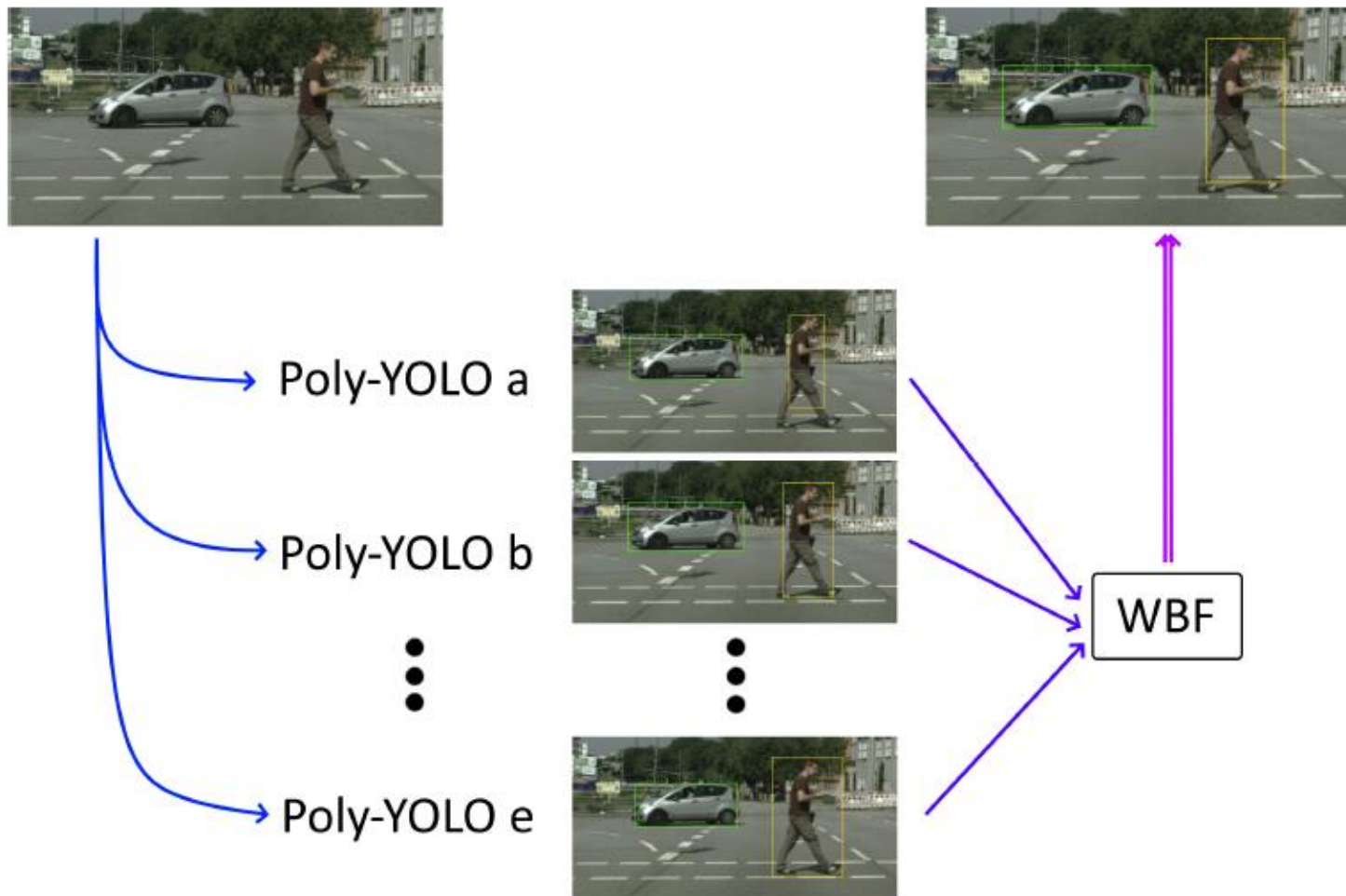
# <学習/推定フロー>



- ①入力動画をフレーム単位で静止画に切り出し
- ②複数のPoly-YOLO検出器で学習
- ③WBF(Weighted Box Fusion)でアンサンブル学習
- ④Refiner:検出境界を高精度化
- ⑤クラス分類(自動車、歩行者 分類)
- ⑥フレーム間オブジェクトの関連付け(追跡処理)



# 採用モデル：Poly-YOLO



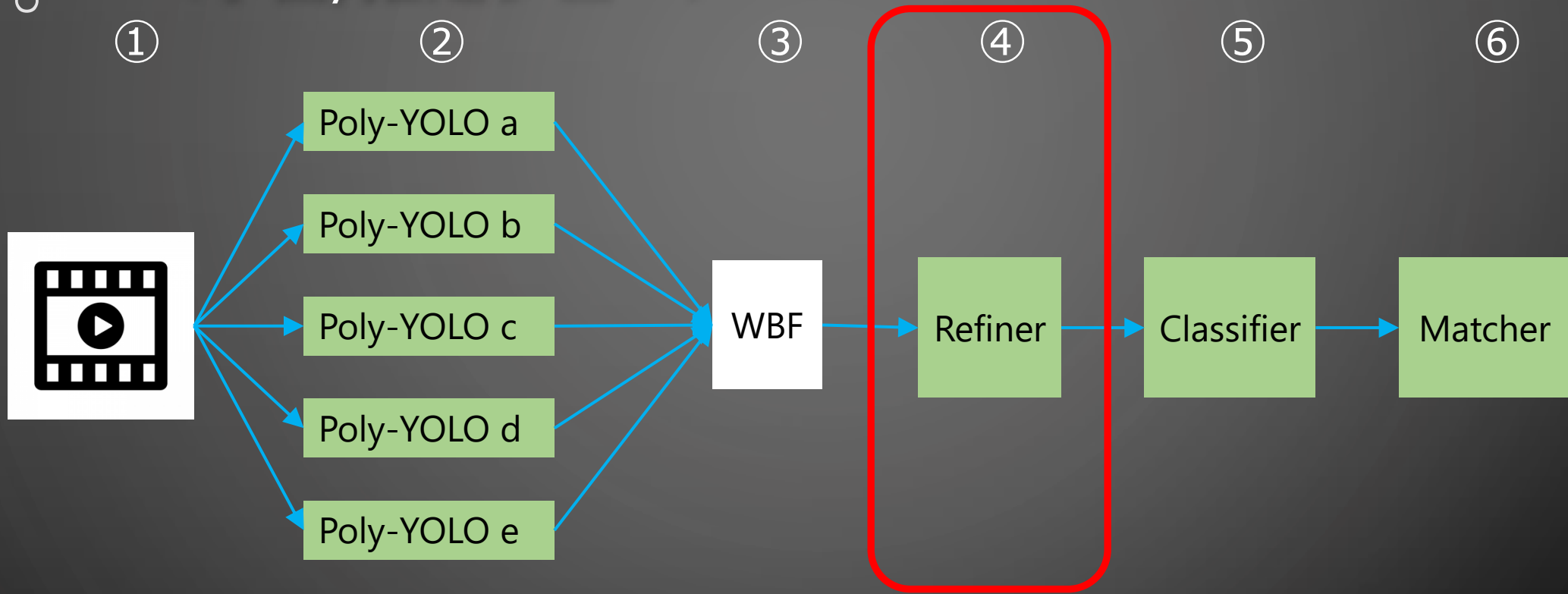
- a)  $448 \times 864$  (自動車、歩行者)
- c)  $352 \times 704$  (自動車、歩行者)
- c)  $224 \times 448$  (自動車、歩行者)
- d)  $960 \times 1952$  (歩行者のみ)
- e)  $544 \times 1120$  (自動車、歩行者)

上記a)～e)の5種類の解像度で学習を行い、WBF(Weighted Box Fusion)によって物体検知のアンサンブル学習を行う。

(WBFの参考)

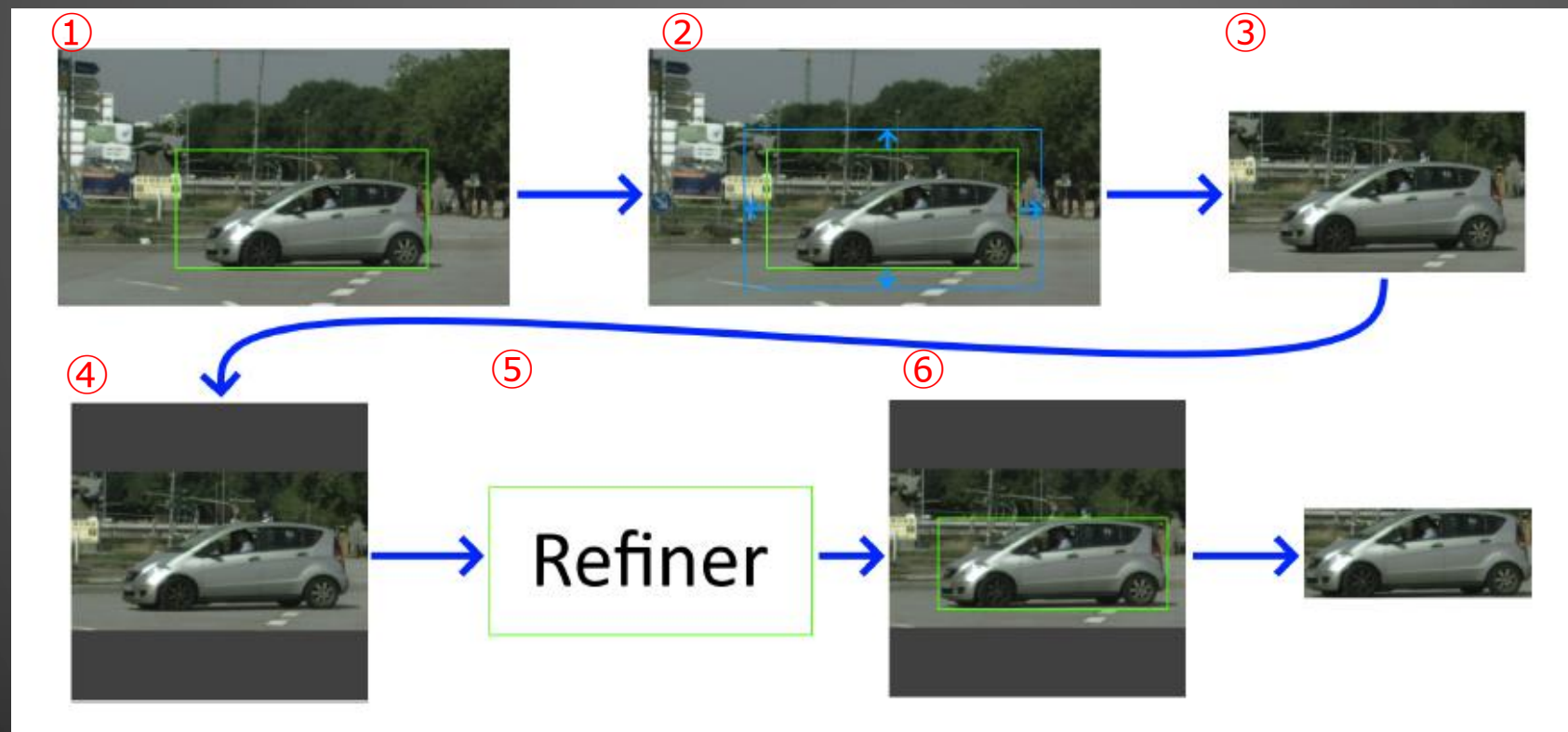
[物体検出で重なったバウンディングボックスを除去・集約するアルゴリズムのまとめ \(NMS, Soft-NMS, NMW, WBF\)](#)

# <学習/推定フロー>



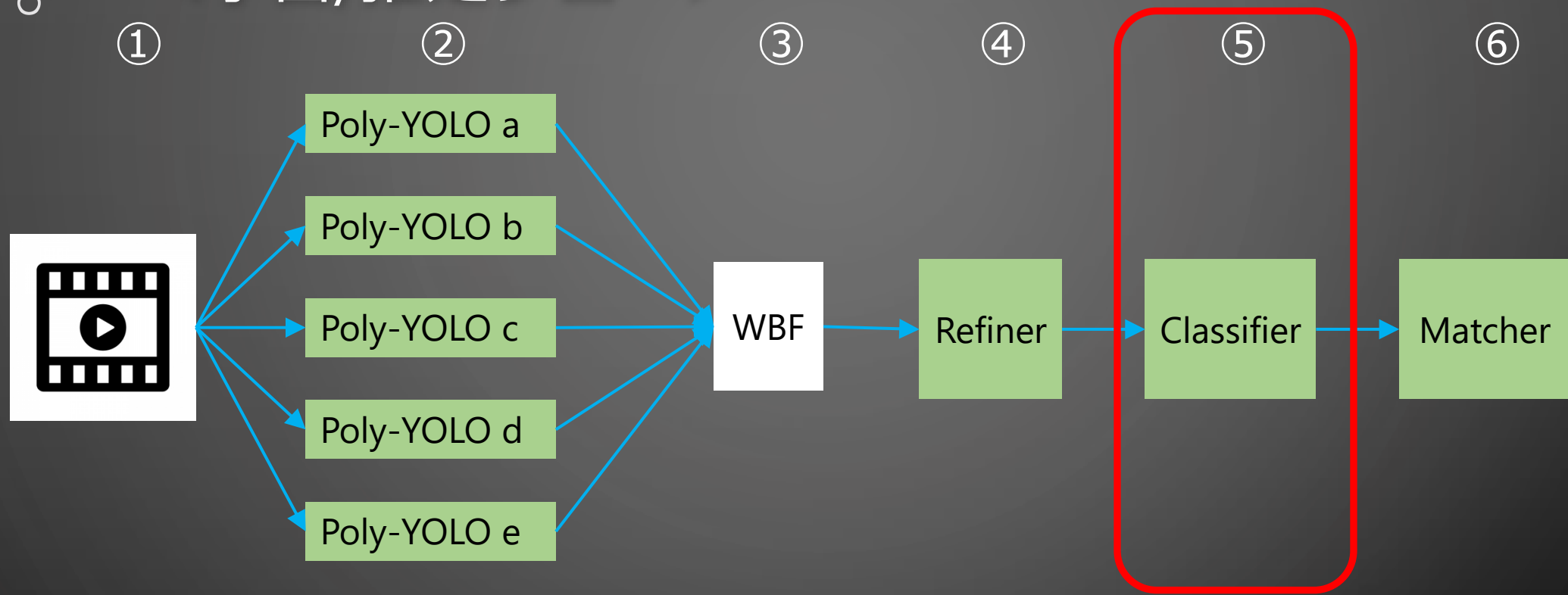
- ①入力動画をフレーム単位に切り出し
- ②複数のPoly-YOLO検出器で学習
- ③WBF(Weighted Box Fusion)でアンサンブル学習
- ④Refiner:検出境界を高精度化
- ⑤クラス分類(自動車、歩行者 分類)
- ⑥フレーム間オブジェクトの関連付け(追跡処理)

## <Refiner (effnetb3バックボーン)>



- ①まず物体検知を行いバウンディングボックスを取得
- ②①の範囲より一回り大きな領域を選択
- ③②の領域で切り抜き
- ④Refiner入力の正方形ウィンドウに一致するようサイズ変更
- ⑤Refinerへ入力
- ⑥より高精度なバウンディングボックスが取得される

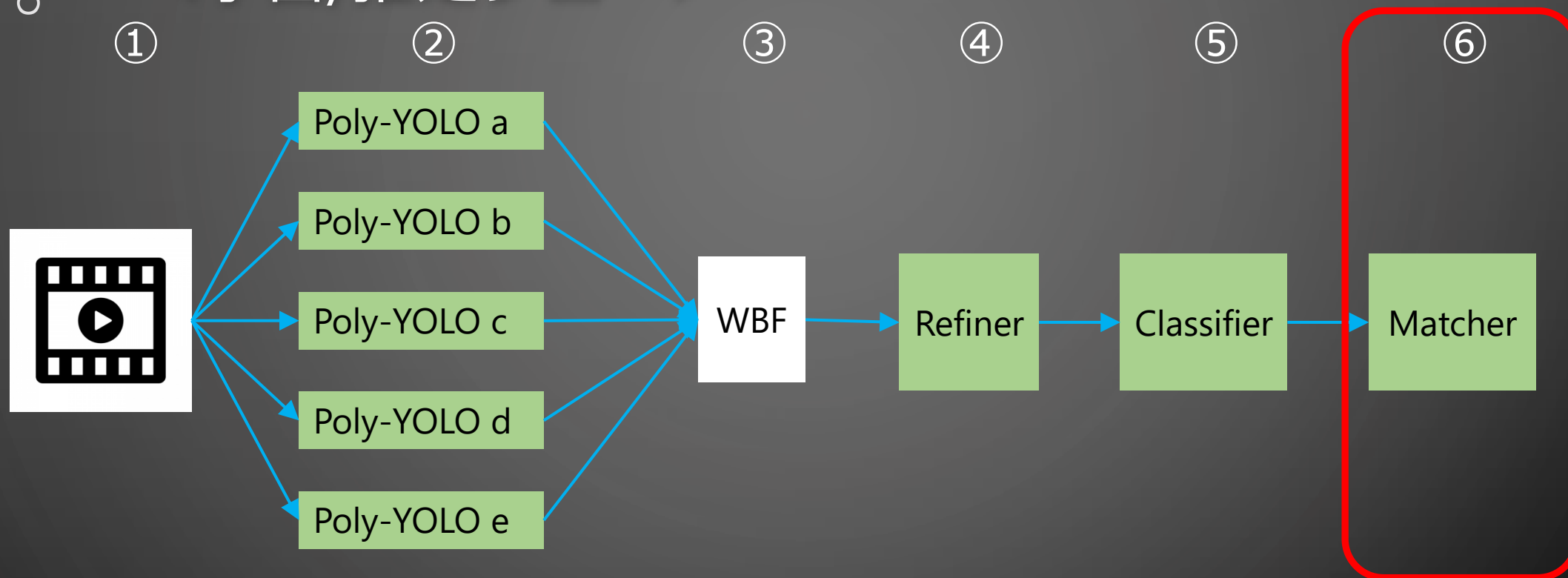
# <学習/推定フロー>



- ①入力動画をフレーム単位に切り出し
- ②複数のPoly-YOLO検出器で学習
- ③WBF(Weighted Box Fusion)でアンサンブル学習
- ④Refiner:検出境界を高精度化
- ⑤クラス分類(自動車、歩行者 分類)
- ⑥フレーム間オブジェクトの関連付け(追跡処理)

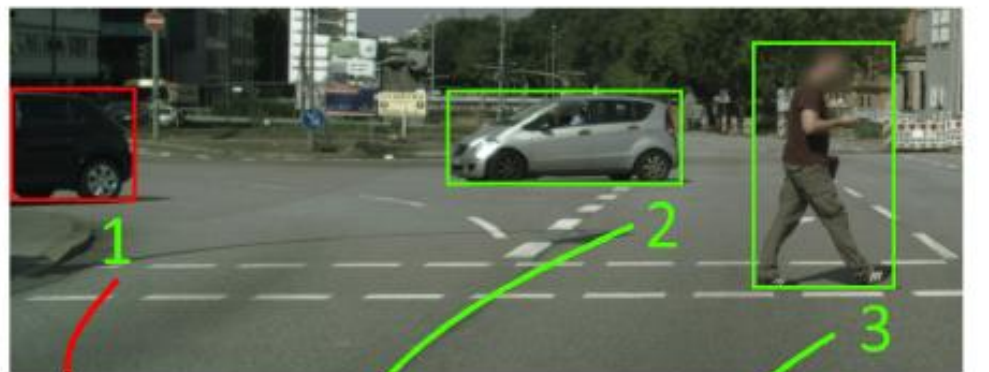


# <学習/推定フロー>



- ①入力動画をフレーム単位に切り出し
- ②複数のPoly-YOLO検出器で学習
- ③WBF(Weighted Box Fusion)でアンサンブル学習
- ④Refiner:検出境界を高精度化
- ⑤クラス分類(自動車、歩行者 分類)
- ⑥フレーム間オブジェクトの関連付け(追跡処理)

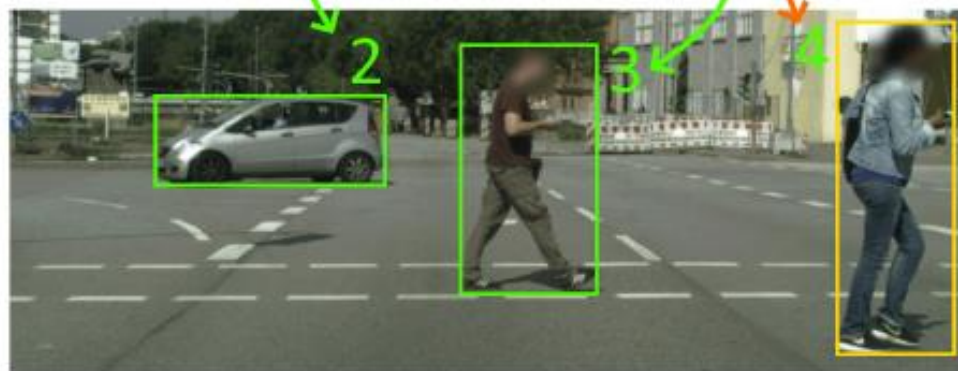
# <Matcher>



Frame t-1

Matcher

(new)

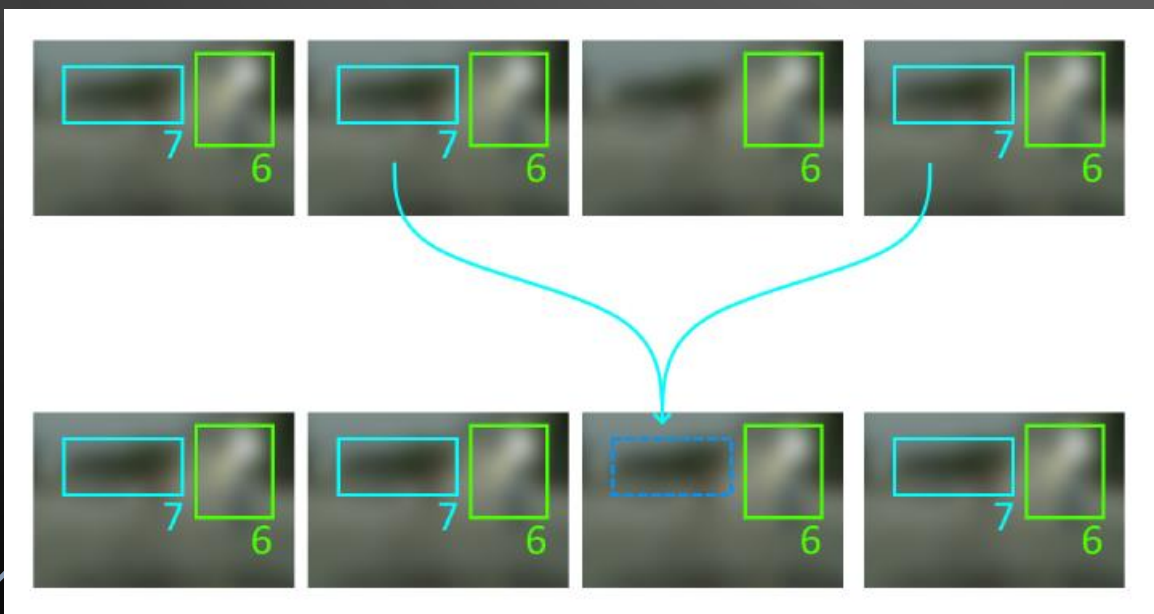
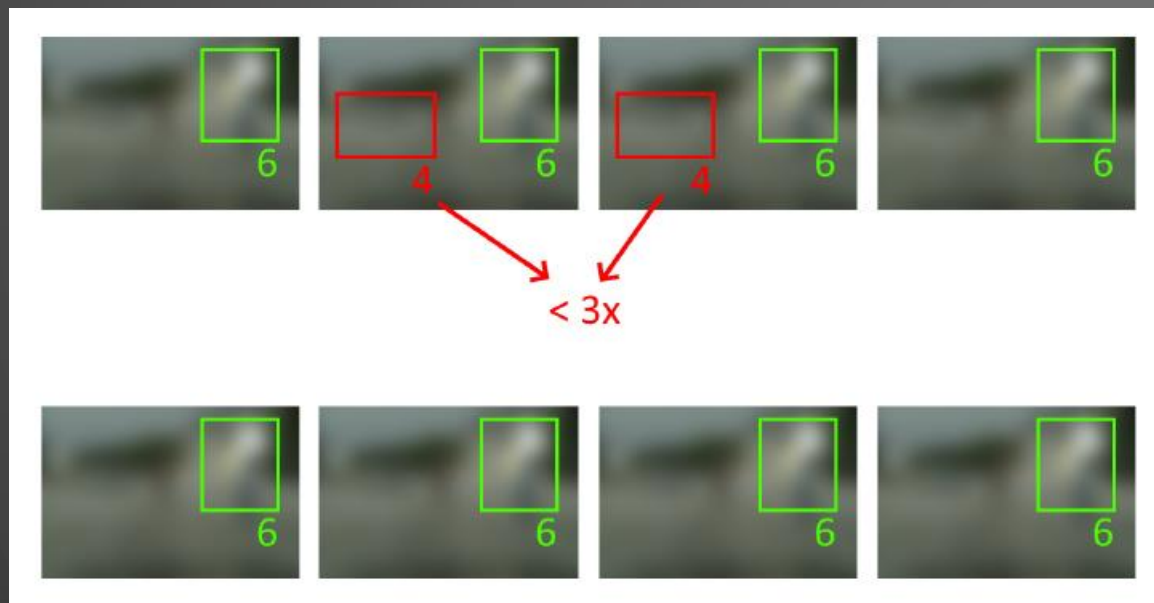


Frame t

Matcherによってボックス間の類似性を計算し、直前のフレームに映っている物体と同一物体かどうかをチェックします。



## <Matcher(続き)>



対象の物体が2フレーム以下しか映りこんでいない場合はボックスを削除する。

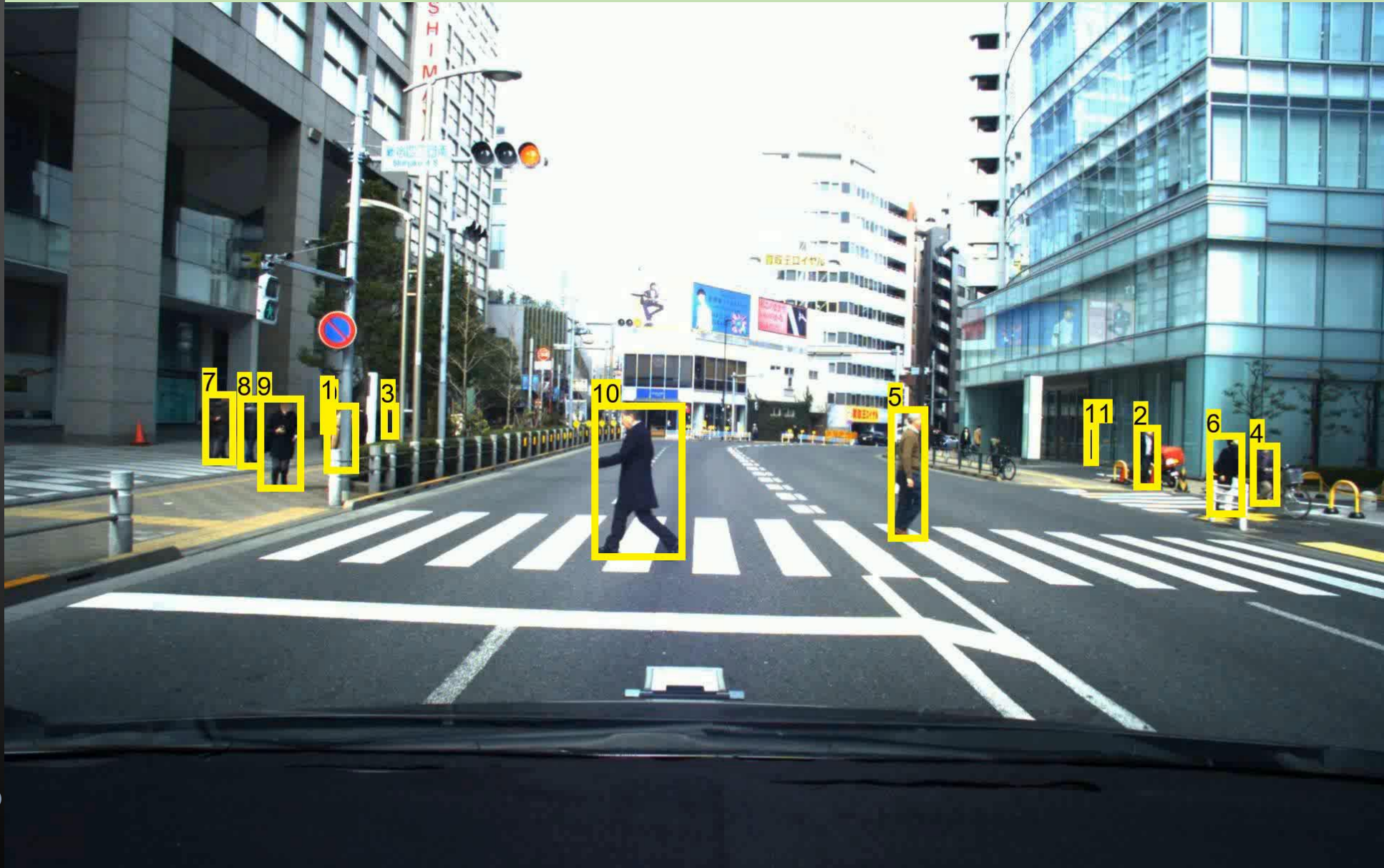
連続フレーム中に特定のフレームで該当ボックスが欠落していた場合、その前後のフレーム情報から欠落しているボックスを生成する。





## <推定結果>

<https://drive.google.com/file/d/1tEmudXjZKc9yvDMViESS05doUjMwZbvX/view?usp=sharing>





# <その他動画（検知前）>

秋葉原

29.97fps

<https://drive.google.com/file/d/135sBO9hEmDGJ10jMsfRqp1oqUMK9tyfd/view?usp=sharing>



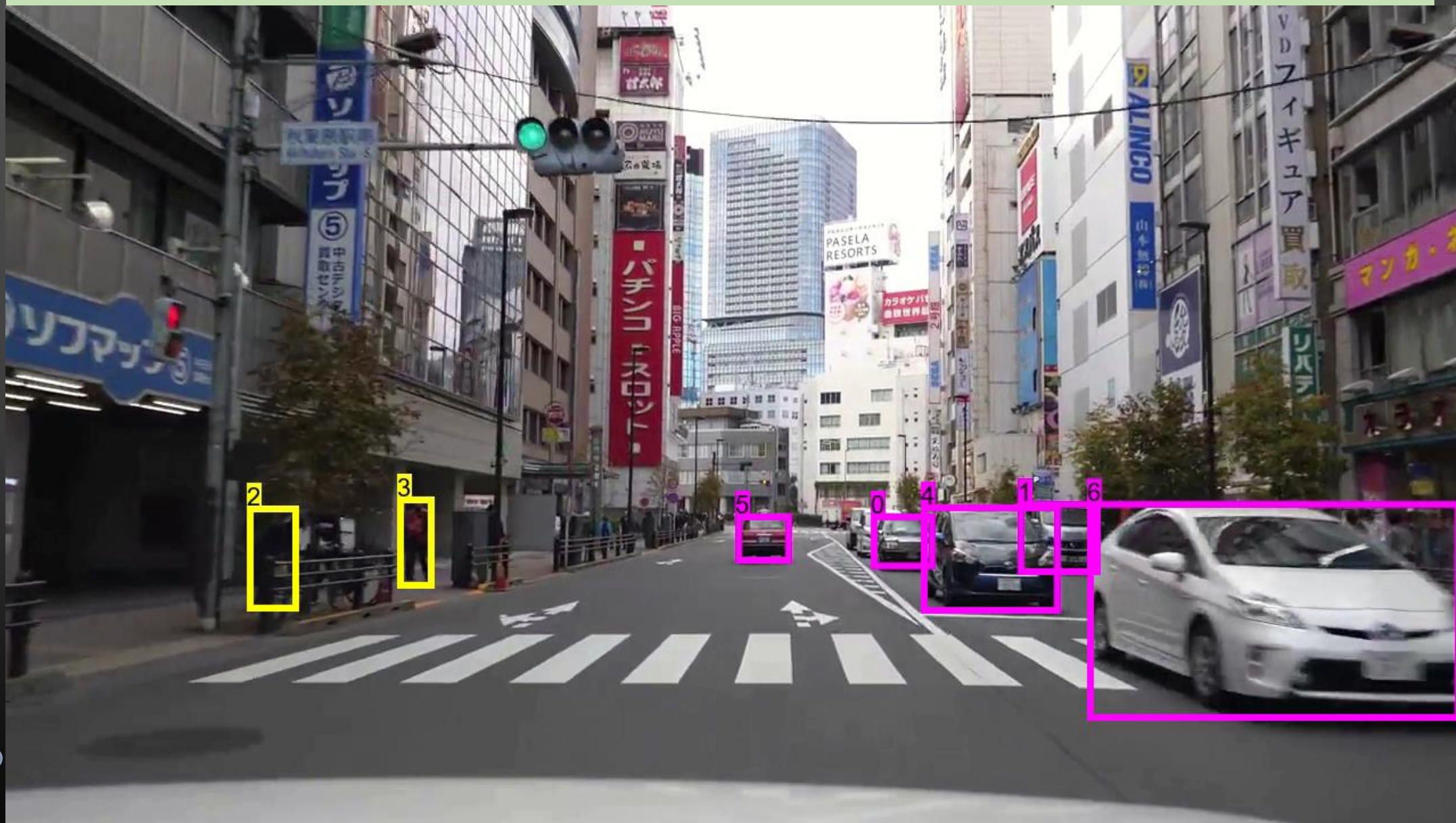


# <その他動画（検知後）>

秋葉原

29.97fps

<https://drive.google.com/file/d/1wLM7L2mtyeZxkqL-jp1YJI7eN36DTxur/view?usp=sharing>



# <感想・まとめ>

## (問題点)

- ①モデルの選定をミスし、学習に多大な時間を要するモデルを選択してしまった。
- ②コンペ参加者が公開されているモデルの為、学習時間より精度重視のモデルの手法となっていた。
- ③学習/推定 共に時間がかかる為、リアルタイムの物体追跡は難しい。
- ④多種のモデルが数珠つなぎになった構成の為、学習/推定 処理が複雑になった。
- ⑤コンペの課題に伴い、検知対象を歩行者と乗用車の2つのみとしたが、信号機も対象に含め、カラーヒストグラムより信号機の色判定等も行いたかったが、学習に時間がかかるので断念した。

## (良かった点)

- ①課題タスクに応じたモデル選定が重要であることを知ることが出来た。
- ②物体追跡を行うアルゴリズムを知ることが出来た。
- ③今回用いたモデルは、他の物体検知のタスクに応用可能であり、知見を蓄積することが出来た。
- ④海外のコンペ参加者と連絡を取ることを覚えた。