

# 時系列深層学習と

# 一般物体認識ネットワーク

# を用いた物体追跡

情報文化学部自然情報学科  
複雑システム系 渡辺研究室 村手翼

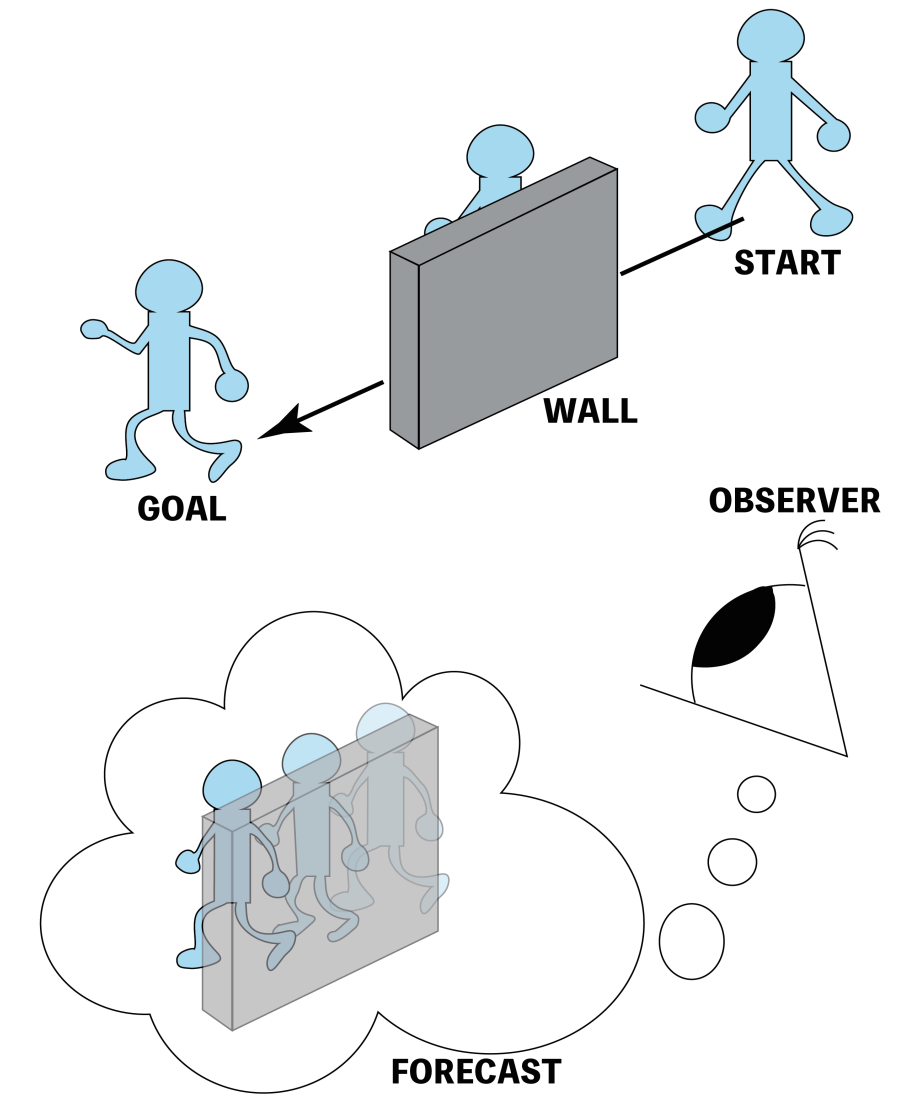
## 1. 物体追跡に対するアプローチ

物体追跡とは動画像において対象物体の存在領域を推定する問題である。

本研究では人間の物体追跡方法を模倣した。人間は物体追跡する際、視覚情報だけでなく、視覚情報をベースとしたある程度の記憶と予想を用いて物体追跡を行う。そこで、視覚的情報に基づいた物体追跡手法に記憶予測機能を付加し、追跡精度向上を目指す。

本研究では視覚的物体追跡に一般物体認識を用いた物体追跡手法の FMST\* を用い、記憶予測機能を畳み込み層を持った LSTM(ConvLSTM) にて再現した。

\* Feature Map Selection Tracking, 2017 年卒 山田真生氏 修士論文より



←人間の行う物体追跡イメージ

## 2. 一般物体認識を用いた物体追跡

FMST は一般物体認識ネットワークの 16 層学習済み VGG net を用いて追跡対象の特徴を抽出し、予測マップを生成。その後、予測マップから対象の存在領域を推定するという手法。(fig 1.)

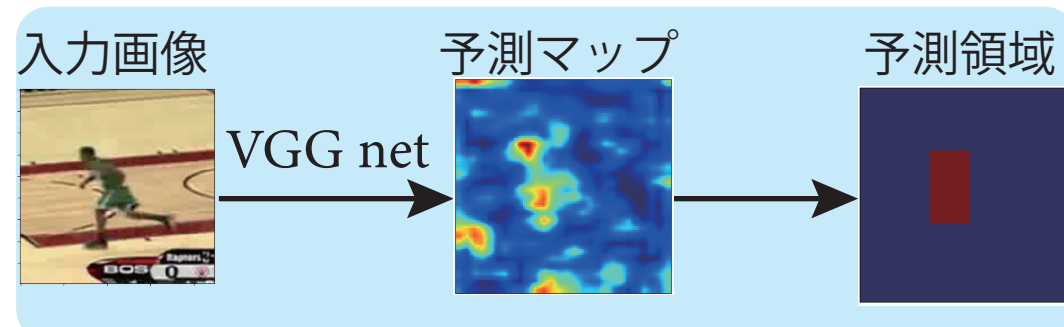


fig 1. 予測領域生成の流れ

FMST は高精度かつ高速な処理を行うが、一般物体認識ネットワークを用いていることから、同カテゴリ物体に予測が転移する場合があります。

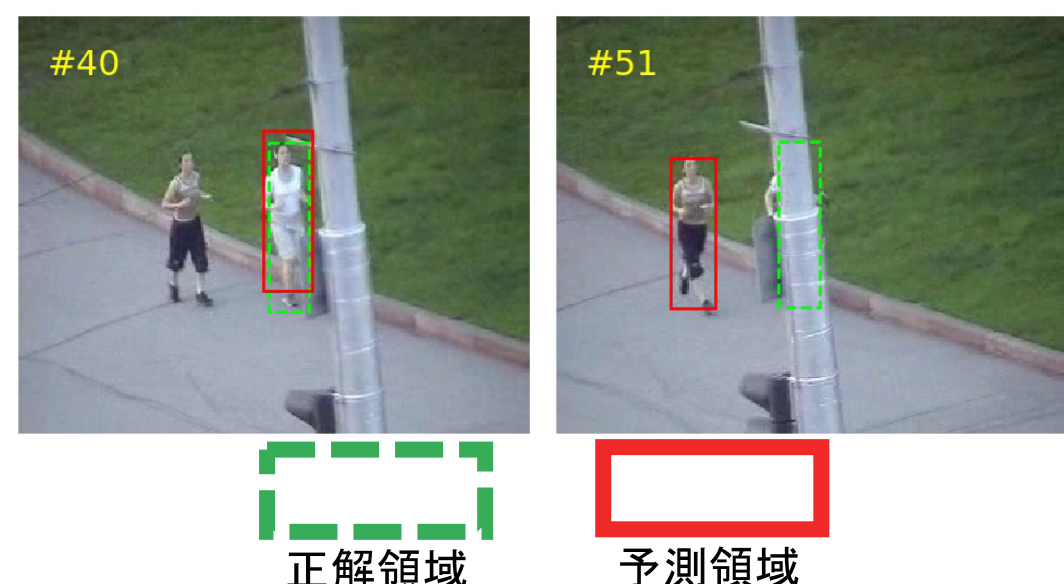


fig 2. 予測転移の例。電柱により予測が転移している。

fig 2. は 40 フレーム時点では予測できているが、対象が電柱の影になる事で予測転移している例である。

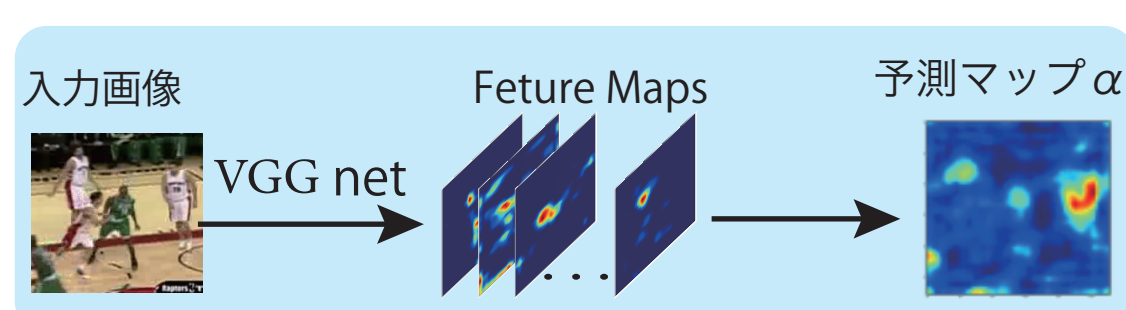


fig 5. FMST より得る 視覚的予測マップ  $\alpha$

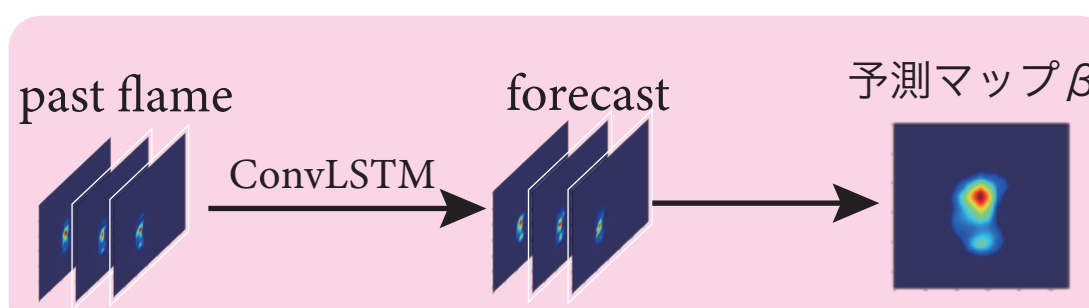


fig 6. 記憶予測機能が生成する予測マップ  $\beta$

提案手法では 2 種類の予測マップを生成する。FMST の手法にて視覚的情報をベースに存在領域を推定する予測マップ  $\alpha$  (fig 5.) と、ConvLSTM を用いて過去のフレーム情報の流れに基づいて存在領域を推定する予測マップ  $\beta$  (fig 6.) の 2 種類を用いる。これら 2 種類の予測マップを足し合わせた予測マップ  $\gamma$  を最終的な予測マップ (fig 7.) とし、そこから予測領域を決定する。

このような手法にて、現在の視覚的情報と過去の記憶に基づく予想を用いる人間の物体追跡方法を模倣している。

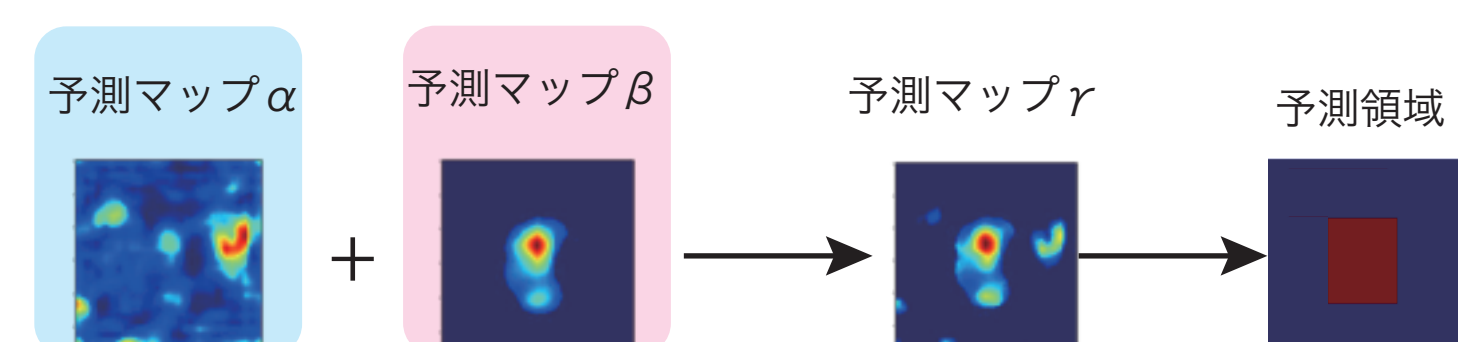


fig 7. 提案手法での予測領域決定の流れ

## 5. 実験結果

FMST と提案手法を 50 の動画からなるベンチマーク (VTB13) で比較した。(fig 8) 処理速度は 20% ほど低下したが、正解と予測の距離に基づく距離精度は 3.37%, 正解と予測の重なりの指標は 2.50% 向上した。

実際の追跡例での比較を fig 9. に示す。

FMST では電柱によって追跡対象が見えなくなることによって、予測が対象以外の人物に転移していることが確認できる。

提案手法では電柱で追跡対象が視界から消失しても、追跡が行えていることが確認できる。実際に人間に近い挙動の追跡ができています。

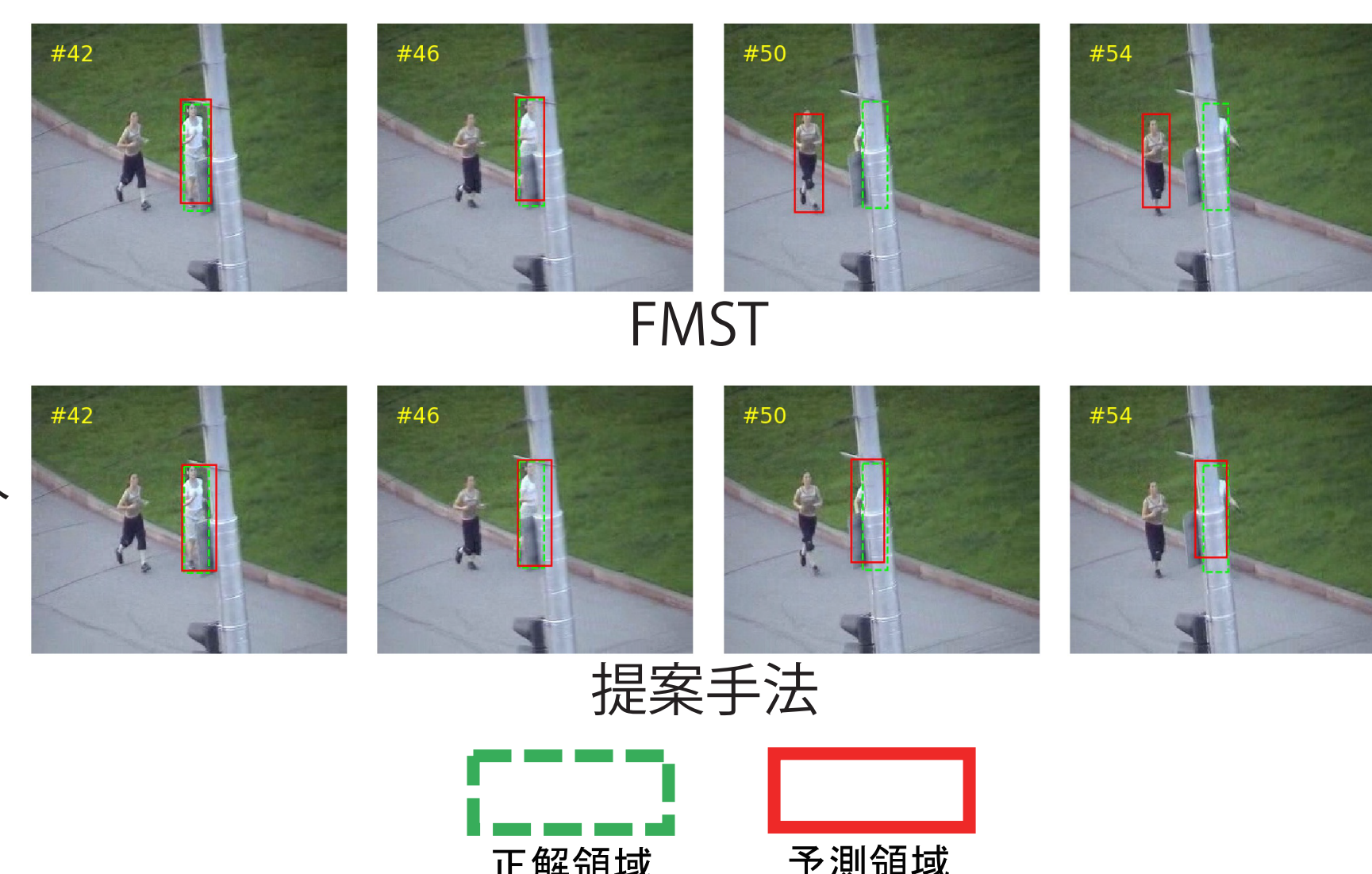


fig 9. VTB13 内の Jogging というタスクの追跡例

	距離指標(%)	オーバーラップ(%)	処理速度 (fps)
FMST	83.68	57.86	42.13
提案手法	87.05	60.36	34.87

fig 8. ベンチマーク VTB13 での精度比較