

1. はじめに

2 フレーム間での移動量を活用した Convolutional Neural Network(CNN) ベースの物体検出手法として, Detect to Track and Track to Detect (D&T) [1] がある. D&T は各フレームで物体検出し, そしてフレーム間での移動量を推定する. しかし, 移動量の推定では, 画像全体または局所的な画素の変化を捉えず, 検出領域の類似性から移動量を推定している. そのため, 移動量推定が正しく出来ないこともある. そこで, 本研究ではオプティカルフローを同時に推定し, 画像全体の動きを加味することで, 物体検出を高精度化する手法を提案する.

2. Detect to Track and Track to Detect(D&T)

D&T は 2 フレームにおける移動量推定と物体検出を同時に行う手法である. D&T では, 各フレームの画像を R-FCN へ入力し, 物体候補領域とクラススコアを得る. それと同時に, 途中の特徴マップを相関層へ入力し, 画像間の微小な変化量を推定する. このように, 物体の移動量を活用して物体検出精度の向上を図っている.

3. FlowNetC

FlowNetC は CNN を用いてオプティカルフローを推定する手法である. はじめに時刻 t 及び $t+1$ の入力画像に対して畳み込み処理を複数回行い, 特徴マップ $f_1(x_1), f_2(x_2)$ を得る. これを相関層へ入力し, 移動幅により一致した座標の特徴マップ $f_1(x_1 + o), f_2(x_2 + o)$ を式 (1) により結合する.

$$c(x_1, x_2) = \sum_{o \in [-k, k] \times [-k, k]} \langle f_1(x_1 + o), f_2(x_2 + o) \rangle \quad (1)$$

ここで, x_1, x_2 はフレーム間の各座標, k は x_1, x_2 を中心とした注目領域, o は移動幅である. 結合した特徴マップから変化量を捉える特徴マップを得る. そして, deconvolution を繰り返すことで特徴マップを拡大し, 元画像サイズのオプティカルフローを推定する.

4. 提案手法

D&T と FlowNetC は, 2 枚の入力画像に対する変化量を求める処理が共通している. 両手法では共に相関層を用いる点においても類似しており, マルチタスク化することで両タスクの精度向上を図ることが可能である. また, D&T は物体ごとに変化量を算出するが, その際, FlowNetC による画素ごとのフロー情報は, 補助情報として物体検出に有効であると考えられる. そこで, 図 1 のように相関層を共通化し, 物体検出結果とオプティカルフローを同時に出力するネットワークを提案する.

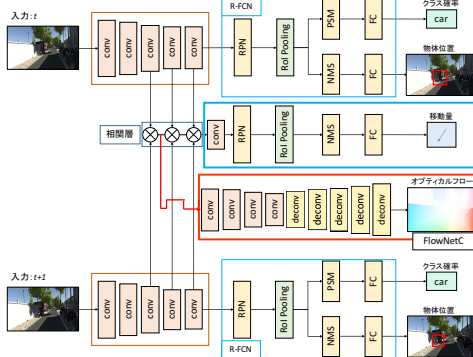


図 1: D&T と FlowNetC のマルチタスク学習モデル

また, 物体候補領域の推定精度を向上させるために, 図 2 のような候補領域の移動量とオプティカルフローとの角度誤差を新たな損失関数として定義する. 提案手法及び損失関数 L_{rad} は図 2, 式 (2) のようになる.

$$L_{rad} = \begin{cases} M_{rad} - F_{rad} & (|M_{rad} - F_{rad}| < \pi/4) \\ 0 & (otherwise) \end{cases} \quad (2)$$

ここで, M_{rad} は相関層から出力された物体領域の移動量, F_{rad} は物体領域内のオプティカルフローの中央値である. 提案手法では M_{rad} と F_{rad} の誤差が $\pm\pi/4$ 以内のパウディングボックスを物体追跡に有効なパウディングボックスとして角度誤差を学習する.

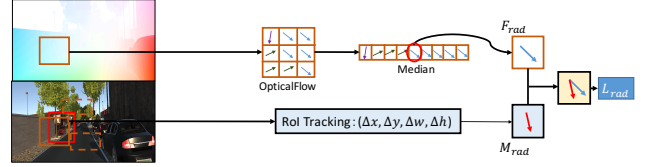


図 2: オプティカルフローを考慮した学習方法

5. 評価実験

本手法の有効性を示すため, D&T と提案手法の定量的な精度の比較を行う.

5.1. 実験概要

本実験には Virtual KITTI を用いる. データセットには天気や時間帯が異なる 8,940 枚の一般車道を走行している CG 画像, オプティカルフローの真値, 車の矩形領域情報が含まれている. 本実験では学習に 338 枚, 評価に 225 枚を用いて, 車の物体検出とオプティカルフローの推定を行う. 物体検出の精度指標には物体検出タスクにおいて使用される mean Average Precision(mAP) を用いる. また, オプティカルフローの精度指標には Average EndPoint Error(AEPE) を用いる.

5.2. 実験結果

従来手法として D&T のみ, FlowNetC のみ, D&T に FlowNetC を追加学習したモデルを用いる. 提案手法は物体追跡とオプティカルフローをマルチタスクとして学習するモデル (D&T+Flow), さらに角度誤差を用いて学習するモデル (D&T+Flow+ L_{rad}) である, 各結果を表 1 に示す. マルチタスク化した D&T+Flow は従来手法と比べ, mAP が 0.4% 低下した. 一方で, D&T+Flow+ L_{rad} は従来手法より mAP が 0.9% 向上し, AEPE も 0.5pixel 削減できた. これより, 本手法で提案した損失関数の有効性が確認できる.

表 1: 各モデルの精度比較

	mAP [%]	AEPE [pixel]
D&T	80.1	-
FlowNetC	-	22.1
D&T+Flow	79.7	23.0
D&T+Flow+ L_{rad}	81.0	22.5

提案手法における出力画像例を以下の図 3 に示す. ここで, 赤い矩形が真値, 青い矩形が推定値である. 図より従来の物体検出に加え, オプティカルフローも出力できていることが確認できる.

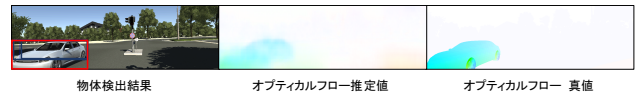


図 3: 提案手法における画像出力例

6. おわりに

本研究では, 物体検出とオプティカルフロー推定をマルチタスクとして学習, 推定する手法を提案し, 物体検出精度を向上させた. 今後は, 評価時にオプティカルフローを考慮した時の精度向上や学習コストの削減を目指す.

参考文献

- [1] C. Feichtenhofer, *et al.*, “ Detect to Track and Track to Detect ”, ICCV 2017.
- [2] A. Dosovitskiy, *et al.*, “ Learning Optical Flow with Convolutional Networks ”, ICCV 2015.