YOLOV を用いた物体予測検出の一検討

A Study on Future Object Detection Using YOLOV

渡部泰樹　進藤嵩紘　渡辺裕
Taiju Watanabe　Shindo Takahiro　Hiroshi Watanabe

早稲田大学基幹理工学部
School of Fundamental Science and Engineering, Waseda University

1．まえがき

物体検出とは，画像の中から定められた物体の種類，位置を正確に特定するタスクである．本稿では，動画を用いた物体予測検出手法を提案する．物体予測検出 (future object detection) とは，過去のフレームから将来の物体の種類，位置を特定するタスクである[1]．物体予測検出では，将来の物体の動向を把握することができるため，危険予測などに応用できる．提案手法では動画物体検出手法の YOLOV[2]を予測用に修正し，動画予測モデルである SimVP[3]の構造を付加することにより，時系列情報の取得を可能としている．

2．提案手法

物体予測検出とは，$t = 1$から$t = T$までの過去のフレーム$\{\mathbf{I}^t\}_{t=1}^{T}$を物体検出モデル $\mathbf{F}$ の入力として，将来の物体の位置，信頼度，クラス予測確率を含んだベクトル$\mathbf{O}^{T+\tau}$を出力するタスクである．この物体検出モデル $\mathbf{F}$ は，時刻$t = T$から$t = \tau$までの物体予測検出モデルであり，式(1)のように定式化できる．

$$\mathbf{O}^{T+\tau} = \mathbf{F}(I^1, I^2, I^3, \cdots, I^T) \tag{1}$$

本稿では，簡単のために$T = 3$, $\tau = 3$として，過去 3 フレームから将来の 3 フレームに存在する物体情報を推定する．物体予測検出手法として，動画物体検出手法の YOLOV を利用する．YOLOV は事前学習済みの YOLOX [4]を利用したモデルであり，動画物体検出タスクにおいて高い性能を示している．そこで，YOLOV を物体予測検出用に修正する．具体的には過去 3 フレームを入力として，正解を将来の 3 フレームの物体情報として学習させる．しかしながら，この YOLOV は時系列情報を保持できないという問題点がある．そこで，入力のフレーム間情報を取得するために動画予測モデルである SimVP の構造を取り入れる．SimVP では中間特徴量に対して，複数の Inception モジュールを適応させることにより入力フレームの時系列情報を保持する．提案する検出モデルの構造を図 1 に示す．

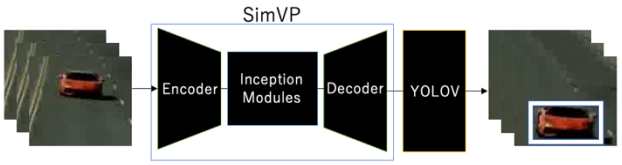

図 1．提案する検出モデルの構造

3．実験

提案手法の有効性を検証するために実験を行った．データセットは ImageNet VID [5]を用いる．ImageNet VID は動画の物体検出用のデータセットであり，訓練用として 3862 個の動画，検証用として 555 個の動画が用意されている．また，クラス数は 30 である．YOLOV では，パラメタ数によって，三つのモデル(YOLOV-S, YOLOV-L, YOLOV-X)が用意されており，それぞれについて SimVP を用いて時系列情報を含めた場合との比較を行う．いずれのモデルについても 10 エポック学習とする．評価指標として，IOU の閾値を 50 とした AP50 を各クラスについて計算し，その mAP を用いた．表 1 に物体予測検出結果を示す．表 1 から SimVP によって時系列情報を保持したモデルの方が，物体検出精度が向上していることがわかる．

表 1．物体予測検出結果

モデル	mAP
YOLOV-S	68.2
YOLOV-S + SimVP	**68.5**
YOLOV-L	74.2
YOLOV-L + SimVP	**74.4**
YOLOV-X	73.8
YOLOV-X + SimVP	**73.9**

4．まとめ

本稿では，YOLOV を用いた物体予測検出手法を提案した．SimVP を用いて時系列情報を保持させることで，物体検出精度が向上することを実験により確認した．

謝辞

本研究成果は，国立研究開発法人情報通信研究機構の委託研究 (No. 05101) により得られたものである．

参考文献

[1] A. Tonderski, et al. “Future Object Detection with Spatiotemporal Transformers”, arXiv: 2204.10321, 1-22, Apr. 2022.
[2] Y. Shi, et al. “YOLOV: Making Still Image Object Detectors Great at Video Object Detection”, arXiv:2208.09686, 1-11, Aug. 2022.
[3] Z. Gao, et al. “SimVP: Simpler yet Better Video Prediction”, CVPR, 3160-3180, Jun. 2022.
[4] Z. Ge, et al. “YOLOX: Exceeding YOLO Series in 2021”, arXiv:2107.08430, 1-7, Jul. 2021.
[5] O. Russakovsky, et al. “ImageNet Large Scale Visual Recognition Challenge”, IJCV, 211-252, Apr. 2015.