第 24 回日本バーチャルリアリティ学会大会論文集（2019 年 9 月）

博物館等の多数展示品を対象としたリアルタイム特定物体検出手法の開発

The Development of Method for the Real-Time Detection of Specific Objects in Museum Exhibits

松永聖明 [1)]，赤嶺有平 [2)]，根路銘もえ子 [3)]
Masaaki MATSUNAGA, Yuhei AKAMINE, and Moeko NEROME

1) 琉球大学 理工学研究科 (〒 903-0213 沖縄県中頭郡西原町字千原)
2) 琉球大学 工学部 (〒 903-0213 沖縄県中頭郡西原町字千原)
3) 沖縄国際大学経済学部（〒 901-2701 沖縄県宜野湾市宜野湾二丁目 6 番 1 号)

概要: 博物館における複数の展示された展示品を対象に、AR(Augmented Reality) による情報提示を行う場合、対象物体の位置及びインスタンス情報をユーザーのスマートフォンなどにおける画面上よりリアルタイムで取得する必要がある。深層学習を利用した物体検出技術が多数提案されており、高速かつ精度の高い物体検出が可能なことが分かっている。しかし、博物館の展示物を対象にする場合、識別対象となるクラス数が膨大となる上学習データを全てユーザーが用意する必要がある為、認識精度の確保、学習コストの問題が発生する。本研究ではこのようなアプリケーションにおいて、実用性の高い物体検出手法の開発を行う。

キーワード： Deep Learning, Object Detection

1. はじめに

近年 CNN[1] を応用した物体検出手法の研究が数多くされており物体領域候補の抽出に Selective Search[2] を用いる R-CNN[3]、Selective Search ではなく RPN(Region Proposal Network) を用いる Faster R-CNN[5] など、複数の物体検出手法が提案されている.

また SSD[6] 及び YOLO[7] では、物体領域候補の抽出及びクラス分類タスクを 1 つのネットワークで完結させる事で、物体領域候補を求めた後に分類器にかける必要性を省き、より高速かつ高精度な物体検出を実現している.

一方で上記した Bounding Box ベースの物体検出手法では検出対象を学習させる際に人手を介して対象物ごとに正解クラス及び Bounding Box をアノテーションすることが必要となることからコスト面で学習データセット構築に大きな課題を残している。

加えてリアルタイム性が望まれる AR アプリケーションにおいて、予測時生成される Bounding Box による膨大なパラメータの削減は、検出速度の改善を行うにあたり解決すべき問題である。

そこで本稿では従来の検出モデルに必要となる Bounding Box によるアノテーション情報を必要とする事無く、対象物体の中心位置情報を推定することが可能な、物体座標検出手法の開発を行い、データセット構築コスト削減並びに検出速度の向上を目指す。

さらに本手法の開発は特に十分な学習データの収集が困難な特定物体検出 (Instance-level Object Detection) においても有効であることを目指し、後述する検証実験では本手法の有効性を検証する。

2. 関連研究

前述の通り、現行の State-Of-The-Art な物体検出手法の問題点として推定時多量に生成される Bounding Box によるハイパーパラメータの増加がボトルネックとなる。そこで Bounding Box ベースではない物体検出手法の研究が進められている。Point Linking Network[8] では画像をグリッドに分割し、対象物体において中心座標と左上・右上・左下・右下 4 点のペアをそれぞれ個別に推定する事で物体特定を効率的に行っており、これにより対象物体の形状に最適な領域の特定を実現している。

また Bounding Box ベースでは検出対象を学習させる際に人手を介して対象物ごとに正解ラベル及び Bounding Box を指定しアノテーションを行う必要があるため非常に学習コストが高くなることが問題となる. この点に関して CornerNet[9] では対象物体の左上座標及び右下座標をヒートマップとして出力する CornerNet を提案している。中心座標を回帰する上で 4 点の Bounding Box 情報を使用せず、2 点での推定を行う事でデータセット構築並びにハイパーパラメータの削減を実現している。本研究においては Bounding Box 情報自体を使用せず、対象物体の中心座標を推定する

事で上記の問題点解決を計る。

3. 提案手法

本提案手法は物体領域候補の抽出と対象物体における中心座標の特定の二段階構造で処理を行う。概要を図 1 に示す。

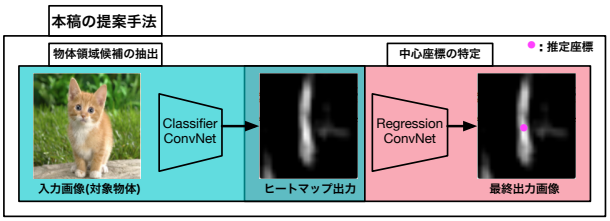

図 1: 提案手法概略図

図 1 左に示す CNN ベースの分類モデル (Classifier-ConvNet) について、画像分類を行う Fully Convolutional Network(FCN) を用意し検出対象と背景画像を分類するモデルを学習する。FCN は入力画像サイズの制約がないため検出対象が含まれる画像を入力すると検出対象領域示すヒートマップが得られる (図 1 中央)。続いてヒートマップ入力に対し物体中心座標を出力する回帰モデルを用いてを中心座標を推定する (図 1 右)。

提案手法では、正規分布により生成された擬似ヒートマップを用いてあらかじめ回帰モデルを学習しておく。そのため回帰モデルについては学習データを用意する必要がなく、中心座標の学習データが不要となる事からデータセット構築にかかるコストの削減が見込まれる。

4. 検証実験

特定物体を対象に、個人レベルで収集可能な少量の学習データに限定し物体検出を行う場合、本提案手法を利用することで実際に対象物体の中心座標を検出することが可能であるのか検証を行う。

4.1 データセット

図 1 左に示す Classifier-ConvNet を訓練するにあたり、学習させる対象物体の動画を対象物体を中心に回る様に 1 本撮影し、撮影した動画に対しトラッキングツールを用いて対象物体領域 (ポジティブデータ) 及び背景領域 (ネガティブデータ) を訓練用データセットとして生成する．今回対象とする物体には屋外に設置された陶器製のシーサー像を撮影したものを使用し、これら 2 つのクラスを分類対象とする．ポジティブデータ及びネガティブデータの一部を図 2 に示す.

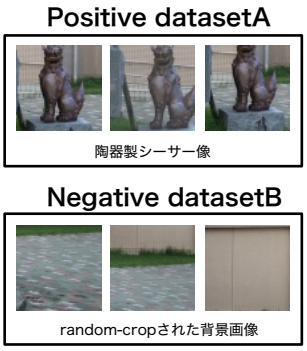

図 2: Classification-Model に使用する学習データ例

図 1 右に示す Regression-ConvNet の学習について、正規分布を用いて作成したヒートマップ画像を用いて画素値の最大値を学習する．ヒートマップ画像生成時には大きさを変化させ，かつノイズを混ぜた．ヒートマップ画像一部を図 3 に示す。

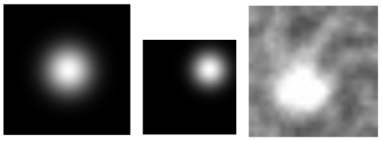

図 3: Regression-ConvNetl において学習時に使用するヒートマップ画像例

5. 実験結果

上記の学習データを元に Classifier-ConvNet を学習させ、対象物体の領域を捉えたヒート画像が取得できることを確認した。入力画像と出力したヒートマップを比較したものを図 4 に示す.

図 4: 入力画像 (左) 及び Heat-Map Model による出力マップ (右)

更に Classifier-ConvNet 及び Regression-ConvNet を組み合わせることにより対象物体における中心座標を取得できることを確認した。結果を図 5 に示す。

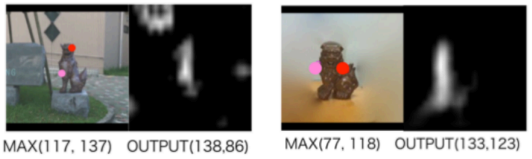

図 5: 対象物体において中心座標の検出に成功した例

6. 考察

検証実験の結果より Classifier-ConvNet の実行結果よる物体候補領域の取得に関して、モデルによる分類が正しく行われた場合に出力結果を可視化したところ候補領域取得ができている事を確認した.

一方で図 5 左図ヒートマップ出力画像に示すように対象物体以外の領域にも反応を示す場合があり，背景情報に少ながらず悪影響を受けている事が考えられる。そこで対象物以外の背景領域に対しピクセル値を平均化する事でぼかしをかけ、追加検証を行ったところ図 4 右に示すヒートマップ画像に比べ対象物体以外への反応が消えた事が確認された。ぼかし処理を加えた入力画像と出力ヒートマップを図 6 に示す.

また図 7 に示すように Regression-ConvNet の精度は良かったもののシーサーのヒートマップを使用した際の予測座標の精度が悪い事が認められ、Classifier-ConvNet の訓練に関し改善の余地がある．原因として Regression-ConvNet において正規分布を用いた擬似ヒートマップを学習しているため、ネットワークは正規分布を用いたヒートマップを期待しているが実際にはシーサーの形状をしていることが影響していると想定される。

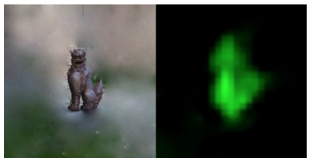

図 6: 背景情報をぼかした入力画像 (左) 及び Classifier-ConvNetl による出力マップ

図 7: 対象物体において中心座標の検出に失敗した例

7. 結び

本稿では Bounding Box によるアノテーション情報を使用せず、対象物体の中心座標を推定することでデータセット構築にかかるコストを削減し、かつリアルタイムに AR アプリケーション上で動作することを目的とした深層学習ベースの特定物体検出手法について検証を行った。問題点として Classifier-ConvNet からの出力は取得できたものの、対象物体の見えの変化や背景に影響される場合がある事が確認された.

今後の課題としてデータセットを数種類用意し対象物体の見えの変化と背景の誤検出を抑えることが可能かどうか検証する．Regression-ConvNet において対象物体の形状に近いヒートマップを学習させた際の精度の改善について検証する.

更には検出速度で高い成果を残す SSD 及び YOLO など他物体検出器との速度及び精度比較や本提案手法における博物館等の展示物を対象としたマルチクラス物体検出への有効性を追加検証していく.

謝辞 本研究の一部は JSPS 科研費 19K01142 の助成によるものである.

参考文献

[1] Alex Krizhevsky, ILya Sutskever, Geoffrey E. Hinton : ImageNet Classification with Deep Convolutional Neural Networks, Advances in neural information processing systems,25(2),2012

[2] J.R.R. Uijlings, K.E.A. van de Sande, T. Gevers, A.W.M. Smeulders : Selective Search for Object Recognition, International Journal of Computer Vision Volume 104 Issue 2, September 2013 Pages 154-171, 2013

[3] Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik : R-CNN Rich feature hierarchies for accurate object detection and semantic segmentation, Girshick2013RichFH, IEEE Conference on Computer Vision and Pattern Recognition, 580-587, 2014

[4] Ross Girshick : Fast R-CNN,IEEE International Conference on Computer Vision (ICCV),10.1109/ICCV.2015.169, 2015

[5] Ren, S. He, K., Girshick, R. B. ,Sun, J. : Faster R-CNN towards real-time object detection with region proposal networks. In NIPS 2015, 91-99.

[6] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg : SSD Single Shot MultiBox Detector,Computer Vision ECCV 2016 pp 21-37, 2016

[7] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi : YOLO (You Only Look Once) Unified, Real-Time Object Detection, IEEE Conference on Computer Vision and Pattern Recognition (CVPR),10.1109/CVPR.2016.91,2016

[8] Xinggang Wang, Kaibing Chen, Zilong Huang, Cong Yao, Wenyu Liu : Point Linking Network for Object Detection, 2017

[9] Hei Law, Jia Deng : CornerNet Detecting Objects as Paired Keypoints, Proceedings of the European Conference on Computer Vision (ECCV), 734-750 ,2019