Topic Title : Object Detection based on deep learning

Published Online : 12 June 2020

Abstract:

The abstract provides a summary of deep learning techniques, like deep convolutional neural networks (DCNNs) using for object detection. Compared with traditional handcrafted feature-based methods, the deep learning -based object detection methods can learn both low-level and high-level image features.

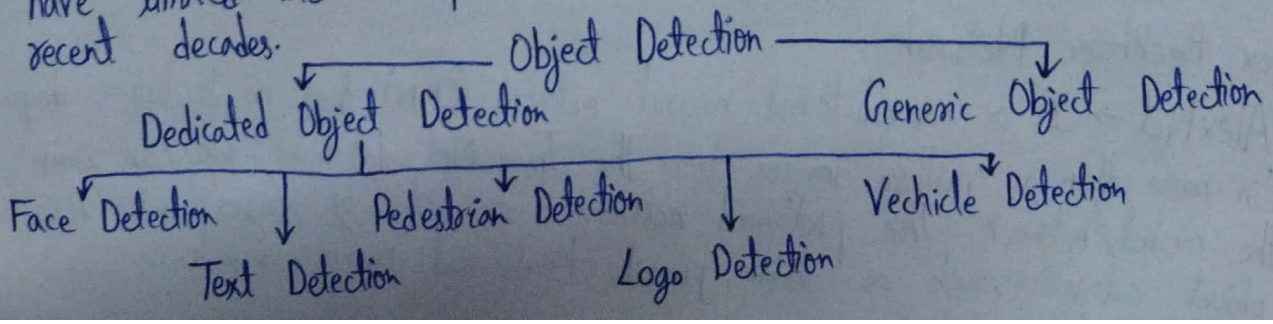
In recent years, the deep learning-based object detection techniques includes → backbone networks, loss functions and training strategies, classical object detection architectures, complex problems, datasets and evaluation metrics, applications and future development directions.

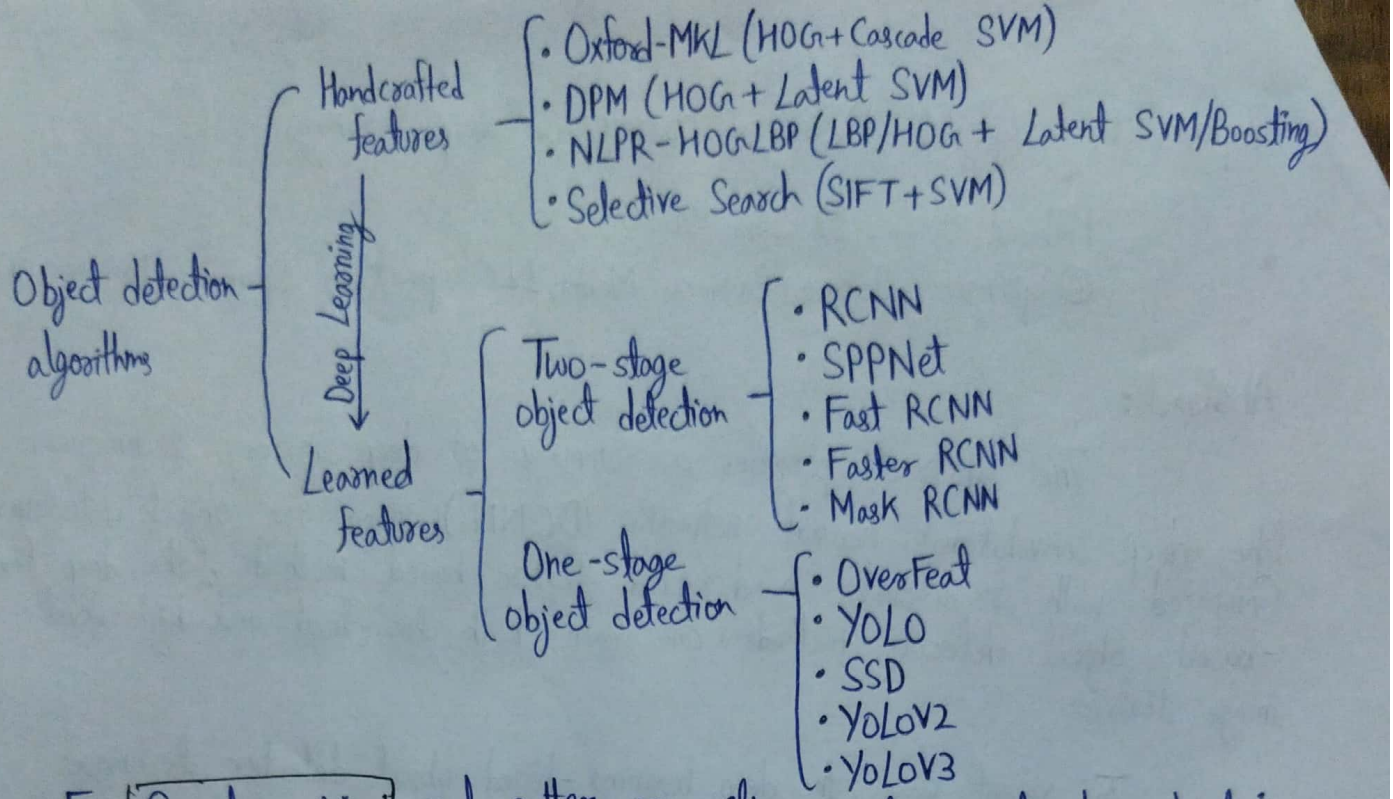
Database :

The essence of object detection is to locate and classify objects, which uses rectangular bounding boxes to locate the detected objects and classify the categories of the objects. It has some relations with object classification, semantic segmentation and instance segmentation.

Object detection is an important area of computer vision and has important applications in scientific research and practical industrial production, such as → Face Detection, Text Detection, Pedestrian Detection, Logo Detection, Video Detection, Vechile Detection and Medical Image Detection, etc.

The limitation of the computing resources, the datasets, and the basic theories have limited the development and application of deep neural networks in recent decades.





→ For "Computer vision" and pattern recognition, a large-scale hierachical image database is used, from ~~Ting~~ ImageNet.
→ Exploring a large collection of categorical data in Sun database.
→ Over a 10 million image database is used for recognition.
→ Introduction to WordNet: an on-line lexical database, of semantic relations between words that links words into semantic relations including synonyms, hyponyms, and meronyms. This help support NLP tasks such as sentiment analysis, automatic language translation, text similarity, and more.
→ IEEE Xplore Digital Library Database.
→ PASCAL VOC, MS COCO

Machine Learning Algorithms:

The DCNNs are the backbone network for object detection.
→ Accuracy of the complex backbone networks (CBNs) can be improved by increasing the depth of the network. It can be also be that reducing the parameters in reasonable ways, which do not affect the accuracy of the lightweight backbone networks (LBNs).

- Complex Backbone Network

① AlexNet → Is a convolutional neural network (CNN) that is 8 layers deep. On more than a million images from the ImageNet database, we can train the model/network. The pretained network can classify images into 1000 object categories, such as table, pencil, utensils and many animals.

② ZFNet → ZFNet improved the accuracy by fine-tuning AlexNet ①, is used to visually determine which parameters should be tuned to get better accuracy instead of trial and error.

③ GoogLeNet → For various Computer Vision applications, including Object Detection, Image Classification, etc.

④ VGGNet → The VGG architecture is the basis of ground-breaking object recognition models, is used for image recognition architectures.

⑤ ResNet → To solve complex problems, we stack some additional layers in the (DNN) which results in improved accuracy and performance, the layers progressively learn more complex features.

⑥ DetNet → DetNet ensures deterministic, low-latency communication over Ethernet networks, guaranteeing that critical data packets reach their destination predictably, benefiting industries like manufacturing and autonomous vehicles. It achieves this through techniques like time-sensitive networking (TSN) and resource reservation.

■ Lightweight backbone network

① SqueezeNet → It is a (CNN) that employs design strategies to reduce the no. of parameters

② Xception → It is a (CNN) with 71 layers deep. (like AlexNet)

③ MobileNet → Class of small, low-latency, low-power models that can be used for classification, detection, etc.

④ MobileNetV2 ⑤ ShuffleNet ⑥ ShuffleNetV2 ⑦ PeleeNet

■ Loss function for object detection

① Classification Loss

• Hinge Loss → is a proxy function 0-1 loss function. Used for max-margin problem in ML or DL, and can be extended to multi-class (SVM) loss.

$$L(y) = \max(0, 1 - t \cdot y) \qquad y = wx - b$$

$$t \in \{-1, 1\}$$

$(w, b) \rightarrow$ hyperplane parameters

- Cross entropy loss → is a metric used in ML to measure how well a classification model performs, used in the softmax classifier → more suitable than hinge loss.

(Distance b/w predicted value and target true value)

$$L(p,w) = -\sum_{i=0}^{K} u_i \log p_i$$

② Location Loss → RSS (Residual Sum of Squares), MSE, SAD (Sum of Absolute Diff.), MAE (Mean Absolute Error.)

③ Loss-based training strategies
- Stage-wise training method
- Multi-task training method

- Architecture of object detection
  - ① Two-stage object detection Method → RCNN, SPPNet, FastRCNN, FasterRCNN, MaskRCNN; → FPN, YOLO, YOLOV2, YOLOV3, SSD, DSSD
  - ② One-stage object detection Method

Types of Results
1. Face Detection
2. Salient Object Detection
3. Pedestrian Detection
4. Remote Sensing Image Detection
5. Medical Image Detection
6. Contextual Information Detection
7. Video Stream Object Detection
8. Depth image (RGB-D) Detection

Results include accuracy metrics, such as precision, recall and average precision, which measures the performance of Object Detection algorithms.

Speed and Frame Rate acheived by different algorithms, which is important for real-time applications.