**Overview of Object Detection Algorithms Using Convolutional Neural Networks**

Algorithm/Model:

1. R-CNN: In 2014, Ross Girshick proposed R\_CNN, which uses a selective search algorithm to replace the sliding window, which solves the problem of window redundancy and reduces the time complexity of the algorithm.
2. SPPNet: In 2015, SPPNet [[18](https://www.scirp.org/journal/paperinformation.aspx?paperid=115011#ref18)] was published on IEEE. The spatial pyramid pooling layer is the core of SPPNet, and its main purpose is to generate a fixed size output for any size input.
3. Fast R\_CNN: Fast R-CNN uses the CNN network to first extract the features of the entire image instead of extracting multiple times for each image block. Then, we can apply the method of creating candidate regions directly to the extracted feature maps. Fast-RCNN does not repeatedly extract features.
4. Faster R\_CNN: The problem with Fast R-CNN: There is a bottleneck: selective search to find all candidate boxes, which is also very time-consuming. To obtain these candidate frames more efficiently, Faster R-CNN has added a neural network region proposal network RPN (region proposal network) that extracts edges.
5. Mask R\_CNN: Mask R-CNN is an extension of the original Faster-RCNN, adding a branch to use existing detection to predict the target in parallel.
6. Trident Net: Trident net was the first to propose the influence of receptive fields on objects of different scales in object detection tasks, and carried out relevant experimental verifications.
7. D2Det: Based on the two-stage method, Cao *et al.* improved the classification and regression branches to further improve the accuracy of object detection and instance segmentation. They proposed D2Det, a method that can both accurately locate and accurately classify.
8. Sparse R-CNN: Sparse R-CNN avoids the manual setting of a large number of hyperparameters for candidate boxes and many-to-one positive and negative sample allocation.
9. YOLOv1: Unlike the R-CNN series that needs to find the candidate area first, and then identify the objects in the candidate area, YOLO’s prediction is based on the entire picture, and it will output all detected target information at one time, including category and location.
10. SSD(Single Shot MultiBox Detector): SSD algorithm is a one-stage method, and MultiBox indicates that the SSD is a multi-frame prediction.
11. RetinaNet: RetinaNet is proposed with believe that the one-stage method is fast but not as accurate as the two-stage because the positive and negative samples are not balanced.
12. CornerNet: ConerNet proposed to solve the object detection problem as a key point detection problem, that is, to obtain the prediction frame by detecting the two key points of the upper left corner and the lower right corner of the target frame.
13. CenterNet: This algorithm is to predict the center point of the target, instead of the two corner points in CornerNet; CenterNet uses a heat map to achieve this, introducing the Gaussian distribution area of the predicted points Calculate the true predicted value.
14. EfficientNet: CVPR believe that the current object detection, either pursues more accurate detection results, but costs a lot, or is more efficient, but at the expense of accuracy. Therefore, the paper designs a set of object detection frameworks to adapt to different constraints, while satisfying high precision and high efficiency.

# Papers Name **:A sensor fusion system with thermal infrared camera and LiDAR for autonomous vehicles and deep learning based object detection**

## Abstract:

Under poor lighting conditions, dazzling sunlight, or bad weather an object might be difficult to be identified with general vision sensors.this paper propose a sensor fusion system that combines a thermal infrared camera and a LiDAR sensor that can reliably detect and identify objects even in environments with poor visibility, such as day or night.

Model/Algorithm:

1. **Thermal infrared camera calibration:**Convergence studies of cameras and LiDAR sensors are being actively conducted.
2. **Thermal image feature point extraction:** This paper propose an algorithm for the direct extrinsic parameter calibration of thermal infrared cameras and LiDAR sensors.
3. **LiDAR 3D point cloud feature point extraction:** In order to fit the field of view (FOV) of the thermal infrared camera without using the full 3D point cloud, primary filtering was performed leaving only the front 180°
4. **Thermal infrared cameras and LiDAR sensors:** If the unique parameters of the thermal infrared camera are known, the R|t between the world coordinate system of LiDAR sensor and the image coordinates of the thermal infrared camera is estimated using the PnP algorithm [[13]](https://www.sciencedirect.com/science/article/pii/S2405959521001818#b13) implemented in the OpenCV library.
5. **YOLOv4:** YOLOv4 is a representative one-stage-detector algorithm and has the advantage of fast detection because localization and classification are performed simultaneously.

# Papers name: **A Deep Learning-Based Object Detection Algorithm for Empty-Dish Recycling Robots**

Abstract:

Robotics is one of the solutions to replace humans and overcome this urgent problem. This paper introduces a deep learning-based object detection algorithm for empty-dish recycling robots to automatically recycle dishes in restaurants and canteens, etc.

Model/Algorithm:

1. YOLO-GD Model: To achieve high-speed and real-time dish detection, a lightweight detection network YOLO-GD is proposed. The network mainly consists of three parts: feature extraction, feature fusion, and result prediction.
2. YOLOv4 :YOLOv4 is a representative one-stage-detector algorithm and has the advantage of fast detection because localization and classification are performed simultaneously.
3. TensorRT: TensorRT is a high-performance deep learning inference optimizer that provides low-latency, high-throughput deployment inference for deep learning applications.

# Papers Name: **Object detection using YOLO**

Abstract:

Object detection is one of the predominant and challenging problems in computer vision. Over the decade, with the expeditious evolution of deep learning, researchers have extensively experimented and contributed in the performance enhancement of object detection and related tasks such as object classification, localization, and segmentation using underlying deep models. Broadly, object detectors are classified into two categories viz. two stage and single stage object detectors. Two stage detectors mainly focus on selective region proposals strategy via complex architecture; however, single stage detectors focus on all the spatial region proposals for the possible detection of objects via relatively simpler architecture in one shot.

Model/Algorithm:

1. ANN
2. SVM
3. Decision trees
4. KNN
5. Two Stage object Detection
6. R\_CNN

### **Challenges in object detection**

1. Multi-scale training
2. Foreground-Background class imbalance
3. Detection of relatively smaller objects
4. Necessity of large datasets and computational power
5. Smaller sized datasets
6. Inaccurate localization during predictions

# Papers Name: **A survey of modern deep learning based object detection models**

Introduction:

Early object detection models were built as an ensemble of hand-crafted feature extractors such as Viola-Jones detector,Histogram of Oriented Gradients etc. These models were slow, inaccurate and performed poorly on unfamiliar datasets. The re-introduction of convolutional neural network (CNNs) and deep learinging for image classification changed the landscape of visual perception.

**Key challenges:**

1. Intra class variation
2. Number of categories
3. Efficiency

Algoritm and Dataset:

1. **Pascal VOC 07/12 :** The Pascal Visual Object Classes (VOC) challenge was a multiyear effort to accelerate the development in the field of visual perception. Two versions of these challenges are mostly used as a standard benchmark. Pascal VOC introduced the mean Average Precision (*mAP*) at 0.5 IoU (Intersection over Union) to evaluate the performance of the models.
2. **ILSVRC:** The ImageNet Large Scale Visual Recognition Challenge (ILSVRC)  was an annual challenge running from 2010 to 2017 and became a benchmark for evaluating algorithm performance. The dataset size was scaled up to more than a million images consisting of 1000 object classification classes.
3. **MS-COCO:** The Microsoft Common Objects in Context (MS-COCO) is one of the most challenging datasets available. It has 91 common objects found in their natural context which a 4-year-old human can easily recognize. It has more than two million instances and an average of 3.5 categories per images.
4. **Open Image**

**Architecture:**

1. Backbone architectures
2. AlexNet
3. VGG
4. GoogLeNet/Inception
5. ResNets
6. ResNeXt:
7. CSPNet:  CSPNet which creates different paths for the gradient flow within the network. CSPNet separates feature maps at the base layer into two parts. It is easy to implement and general enough to be applicable on other architectures like ResNet , ResNeXt , DenseNet , Scaled-YOLOv4  etc. Applying CSPNet on these networks reduced computations from 10% to 20%, while the accuracy remained constant or improved.
8. EfficientNet: It summarized how altering network parameters like depth, width and resolution influence its accuracy. It showed how scaling any parameter individually comes with an associated cost. Increasing depth of a network can help in capturing richer and more complex features, but they are difficult to train due to vanishing gradient problem.