Different Object Detection Methods

Date: 21.05.2021

Object detection:

Object detection is technology that includes computer vision and image processing. Used to detect objects in images and videos.

Artificial intelligence is the basic principle that drives object detection. The process is obtained in data collected through computer vision, then it builds the models using machine learning algorithms or deep learning algorithms.

The camera collects the images and sends them to the image processing unit and then data is used to find the solution by using various object detection methods.

Machine learning based model approaches computer vision techniques are used to look at various features of images, such as the color histogram or edges, to identify groups of pixels that may belong with its label.

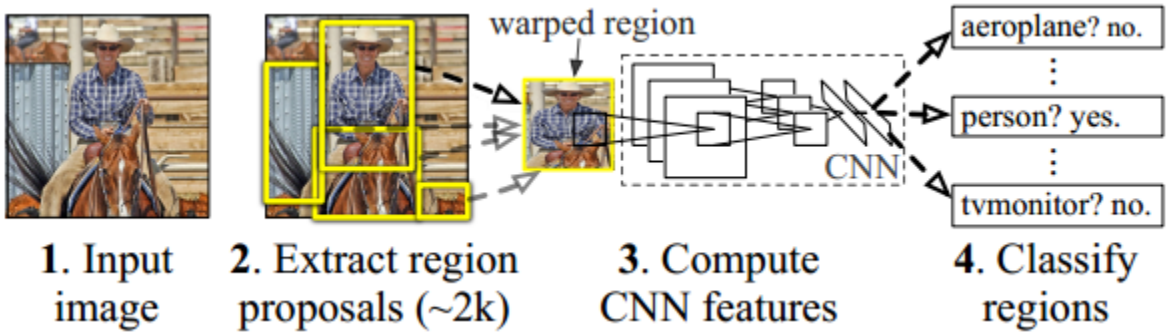
On another hand deep learning based approaches employ convolutional neural networks(CNNs) to perform end to end, unsupervised object detection, in which features don't need to be defined and executed separately.

Deep learning based object detection:

1. R CNN :

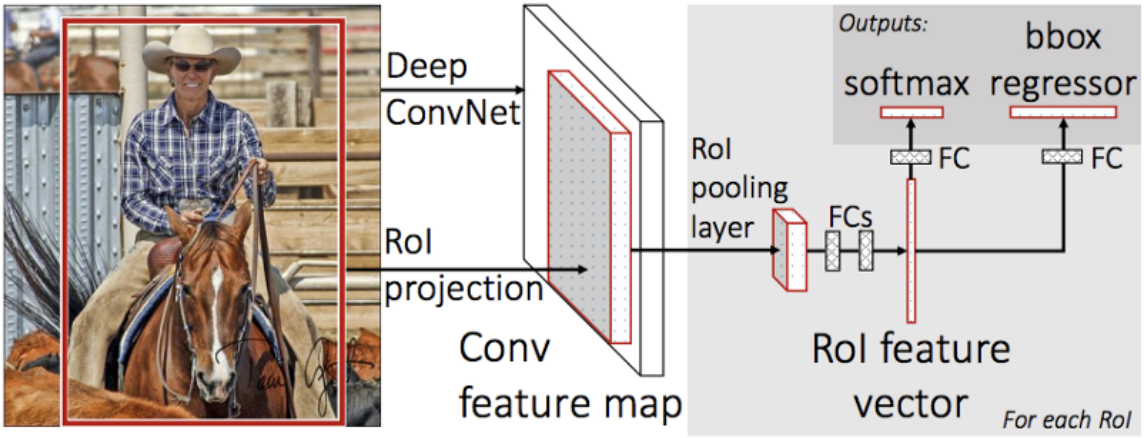
   The R CNN consists of three main modules:

   a. The first module generates 2000 region proposals using the selective algorithm.
   b. After being resized to a fixed predefined size , the second module extracts a features vector length 4,096 from each region proposal.
   c. The third module used a pre-trained SVM algorithm to classify the region proposal to either the background or one of the object classes.
   d. As an extension of RCNN the fast RCNN Model is proposed , to overcome some limitations.
   e. It still takes a huge amount of time to train the network as you would have to classify 2000 region proposals per image. It cannot be implemented real time as it takes around 47 seconds for each test image. That's why the fast RCNN model is proposed.
   f. High computation time as each region is passed to the CNN separately also it uses three different models for making predictions.



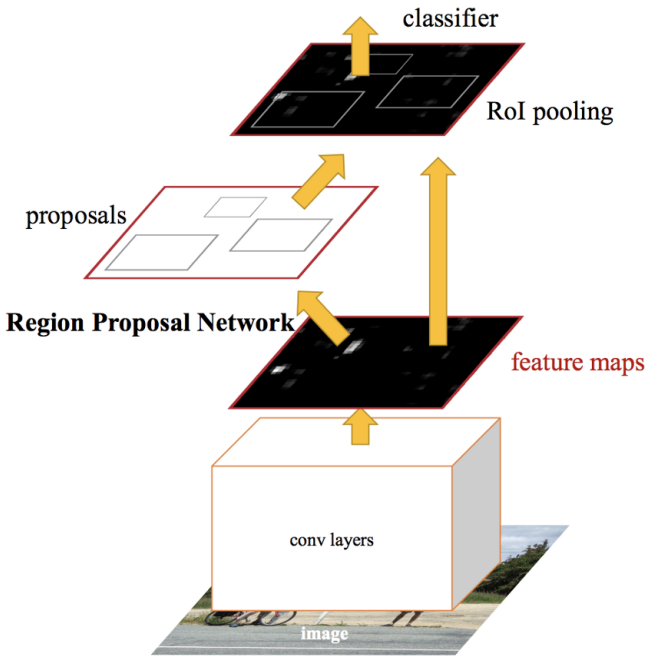
1. The method takes an image as input and extracts around 2000 region proposals from the image(Step 2 in the above image).
2. Each region proposal is then warped(reshaped) to a fixed size to be passed on as an input to a CNN.
3. The CNN extracts a fixed-length feature vector for each region proposal(Step 3 in the above image).
4. These features are used to classify region proposals using category-specific linear SVM(Step 4 in the above image).
5. The bounding boxes are refined using bounding box regression so that the object is properly captured by the box.

2. Fast R-CNN:
    a. R-CNN to build a faster object detection algorithm and it was called Fast R-CNN. The approach is similar to the R-CNN algorithm. But, instead of feeding the region proposals to the CNN, fast R-CNN feeds the input image to the CNN to generate a convolutional feature map.
    b. The reason “Fast R-CNN” is faster than R-CNN is because you don’t have to feed 2000 region proposals to the convolutional neural network every time. Instead, the convolution operation is done only once per image and a feature map is generated from it.
    c. Fast RCNN is more accurate than RCNN.
    d. Compared to RCNN, which has multiple stages(region proposal generation, feature extraction and classification using SVM), faster RCNN builds a network that has only a single stage.
    e. Each image is passed only once to the CNN and feature maps are extracted. Selective search is used on these maps to generate predictions. Combines all the three models used in RCNN together.
    f. Selective search is slow and hence computation time is still high.



g. Fast RCNN has certain problem areas. It also uses selective search as a proposal method to find the Regions of Interest, which is a slow and time consuming process. It takes around 2 seconds per image to detect objects, which is much better compared to RCNN. But when we consider large real-life datasets, then even a Fast RCNN doesn't look so fast anymore.

3. Faster RCNN:
   a. Similar to Fast R-CNN, the image is provided as an input to a convolutional network which provides a convolutional feature map. Instead of using a selective search algorithm on the feature map to identify the region proposals, a separate network is used to predict the region proposals.
   b. This model introduces Region proposal network(RPN) which is a network that is responsible for the region proposal computation and also for object detection.
   c. Faster RCNN shares computations across all proposals rather than doing the calculations for each proposal independently. This is done by using the ROI pooling layer, which makes faster R-CNN faster than fast R-CNN.
   d. The RPN generates regional proposals.
   e. For all region proposals in the image, a fixed-length feature vector is extracted from each region using the ROL pooling layer.
   f. The extracted feature vectors are then classified using the fast RCNN.
   g. The class scores of the detected objects in addition to their bounding-boxes are returned.
   h. Replaced the selective search method with a region proposal network which made the algorithm much faster.
   i. Object proposal takes time and as there are different systems working one after the other, the performance of systems depends on how the previous system has performed.



4. Mask R-CNN:
    a. An extension to the faster R-CNN model includes another branch that returns a mask for each detected object.
    b. This is not real time performance, but it is good enough for a detector needed for industrial purposes or as a worker assisting tool.
    c. Mask RCNN extends the Faster RCNN framework by additionally including a branch for prediction of an object mask in parallel on each region of interest.
    d. This model is simple to train and it incorporates a three part loss that takes into account the detection loss, as well as the mask loss that leads to better performance.
    e. Instead of RoIPool operation for extracting small feature maps from each region of interest in the Fast RCNN architecture, they present a new operation called RoI Align that removes the hard quantization of RoIPool, aligning the extracted features with the input.
    f. Replaced the selective search method with a region proposal network which made the algorithm much faster.

5. YOLO:
    a. Yolo is a standard way of detecting objects in the field of computer vision.YOLO stands for You Only Look Once. this
    b. The base YOLO model processes images in real-time at 45 frames per second, while the smaller version of the network, Fast YOLO processes an astounding 155 frames per second while still achieving double the mAP of other real-time detectors.

c. This algorithm outperforms the other detection methods, including DPM and R-CNN, when generalising from natural images to other domains like artwork.
d. This method that frames object detection as a regression problem to spatially separated bounding boxes and class probabilities was considered.
e. In this method, a single network predicts the bounding boxes and the probabilities directly in just one evaluation.
f. This enabled it to take into account the direct detection performance and made it easier to optimize the whole process, such that impressive speed gains were possible.
g. Disadvantages:
    i. The model has a hard time detecting small object, especially the ones that appear in groups (eg. birds, fish) because of the strong spatial constraints applied on bounding box predictions because each grid cell predicts just 2 boxes and can have just 1 class
    ii. the fact that the model struggles to generalize to objects in a new setting, a new configuration.

6. Single shot detectors:
    a. Single Shot Detector (SSD) is a method for detecting objects in images using a single deep neural network. The SSD approach discretises the output space of bounding boxes into a set of default boxes over different aspect ratios. After discretizing, the method scales per feature map location. The Single Shot Detector network combines predictions from multiple feature maps with different resolutions to naturally handle objects of various sizes.
    b. SSD completely eliminates proposal generation and subsequent pixel or feature resampling stages and encapsulates all computation in a single network.
    c. Easy to train and straightforward to integrate into systems that require a detection component.
    d. SSD has competitive accuracy to methods that utilise an additional object proposal step, and it is much faster while providing a unified framework for both training and inference.
7. Region-based Convolutional Neural Networks (R-CNN):
    a. The Region-based Convolutional Network method (RCNN) is a combination of region proposals with Convolution Neural Networks (CNNs).
    b. R-CNN helps in localising objects with a deep network and training a high-capacity model with only a small quantity of annotated detection data. It achieves excellent object detection accuracy by using a deep ConvNet to classify object proposals.

c. R-CNN has the capability to scale to thousands of object classes without resorting to approximate techniques, including hashing.

Classification Algorithms :

a. Logistic Regression:
   i. Logistic regression is a machine learning algorithm for classification. In this algorithm, the probabilities describing the possible outcomes of a single trial are modelled using a logistic function.
   ii. Logistic regression is designed for this purpose (classification), and is most useful for understanding the influence of several independent variables on a single outcome variable.
   iii. Works only when the predicted variable is binary, assumes all predictors are independent of each other and assumes data is free of missing values.

b. Naïve Bayes :
   i. Naive Bayes algorithm based on Bayes' theorem with the assumption of independence between every pair of features. Naive Bayes classifiers work well in many real-world situations such as document classification and spam filtering.
   ii. This algorithm requires a small amount of training data to estimate the necessary parameters. Naive Bayes classifiers are extremely fast compared to more sophisticated methods.
   iii. Naive Bayes is known to be a bad estimator.

c. Stochastic Gradient Descent :
   i. This is a simple and very efficient approach to fit linear models. It is particularly useful when the number of samples is very large. It supports different loss functions and penalties for classification.
   ii. Efficiency and ease of implementation.
   iii. Requires a number of hyper-parameters and it is sensitive to feature scaling.

d. K-Nearest Neighbours:
   i. Neighbours based classification is a type of lazy learning as it does not attempt to construct a general internal model, but simply stores instances of the training data. Classification is computed from a simple majority vote of the k nearest neighbours of each point.
   ii. This algorithm is simple to implement, robust to noisy training data, and effective if training data is large.

iii. Need to determine the value of K and the computation cost is high as it needs to compute the distance of each instance to all the training samples.

e. Decision Tree :
   i. Given a data of attributes together with its classes, a decision tree produces a sequence of rules that can be used to classify the data.
   ii. It is simple to understand and visualise, requires little data preparation, and can handle both numerical and categorical data.
   iii. Decision trees can create complex trees that do not generalise well, and decision trees can be unstable because small variations in the data might result in a completely different tree being generated.

f. Random Forest :
   i. classifier is a meta-estimator that fits a number of decision trees on various sub-samples of datasets and uses average to improve the predictive accuracy of the model and controls over-fitting. The sub-sample size is always the same as the original input sample size but the samples are drawn with replacement.
   ii. Reduction in over-fitting and random forest classifier is more accurate than decision trees in most cases.
   iii. Slow real time prediction, difficult to implement, and complex algorithm.

g. Support Vector Machine:
   i. is a representation of the training data as points in space separated into categories by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall.
   ii. Effective in high dimensional spaces and uses a subset of training points in the decision function so it is also memory efficient.
   iii. The algorithm does not directly provide probability estimates, these are calculated using an expensive five-fold cross-validation.

Data augmentation:

Data augmentation is a strategy that enables practitioners to significantly increase the diversity of data available for training models, without actually collecting new data. Data augmentation techniques such as cropping, padding, and horizontal flipping are commonly used to train large neural networks.

These techniques improve and scale up the training set to expose the model to scenarios that would have otherwise been unseen. Data augmentation in computer vision is key to getting the most out of your dataset, and state of the art research continues to validate this assumption.

Steps of data augmentation:

1. Image rotation.
2. Add noise
3. Image flipping
4. Add many others there

Generative Adversarial Nets (GAN)

Generative Adversarial Nets (GAN) is a method of data augmentation. The generative model in the GAN framework tries to produce realistic images to fool the discriminative model, while the discriminative model attempts to distinguish the generated samples from the real images.

The emergence of GAN has attracted the attention of many researchers, and many variants of GAN have been proposed to improve the quality of the synthetic image. The GAN approaches are rarely used for real-world scene detection problems, as it becomes much more difficult to generate an image with many object instances placed in a relevant background than to generate an image with only one object.

Methods for Fault Detection in Wind Turbine

1. Manual Annotation:

   We have to manually distribute the faults in different classes like:

   1. Leading Edge erosion (LE erosion):
   2. Vortex Generator panel (VG panel):
   3. VG panel with missing teeth:
   4. Lightning receptor:

2. Image augmentation:

1. Normal augmentation:
   a. Perspective transformation for the camera angle variation simulation.
   b. Left-to-right flip or top-to-bottom flip to simulate blade orientation: e.g., pointing up or down.
   c. Contrast normalization for variations in lighting conditions.
   d. Gaussian blur simulating out of focus images.

2. Pyramid and patching augmentation:
   a. The bottom level of the pyramid scheme is defined as the image size where either the height or width is 1000 pixels.
   b. In the pyramid, from top to bottom, images are scaled from1.00x to 0.33x, simulating from the highest to the lowest resolutions.
   c. Sliding windows with10% overlap were scanned over the images at each resolution to extract patches containing at least one object.

d. Using the multi-scale pyramid and patching scheme on the acquired high-resolution training images, scale-varied views of the same object were generated and fed to the neural network.
e. In this scheme, the main full-resolution image was scaled to multiple resolution images like 1.00x, 0.67x, 0.33x ,and on each of these images, patches containing objects were selected with the help of a sliding window with 10% overlap. The selected patches were always 1000×1000 pixels.

Examples of tower inspection:

Some use-cases in which the UAVs were employed for inspection of towers:

1. Telecommunication tower inspection for designing new tower mount:

Drone was used by Ryka UAS, a Seattle based company, for identifying, analyzing and designing new antenna mount. Due to the usage of drones, this process took only 1 day for capturing all images and creation of high definition 3D models.

2. Cell tower inspection:

The Unnamed Vehicle Technologies team used drones to collect clear, accurate visual data on the state of cell power phones. Using a high quality camera, they were able to fly significant distances and collect clear images as required. This process is said to have needed a single fly.

3. Cell phone tower inspection:

SkyVue Solutions used drones and Pix4Dinspect to slash the inspection time of a cell phone tower in Douala, Cameroon. This presented amazing results. It was observed that the inspection time was decreased by one-third.