THINK INDIA JOURNAL

ISSN:0971-1260
Vol-22- Issue-14-December-2019

Object Recognition Using Shift Invariant Artificial Neural Network

Dr B. Annapurna
Assoc. Prof. in Computer Science and Engineerig , Aditya College of Engineering Surampalem , East Godavari Dt., AP, India

Mr. Raushan Kumar Singh
Technical Director, Spectrum Solutions Pondicherry, India

Abstract:

An Object classification issue is predicting the label of a photograph from many of the predefined labels. It assumes that there may be single item of identification in the picture and it covers a sizable portion of image. Detection is just no longer simplest finding of the magnificence of item however additionally localizing the quantity of an item in the photograph. The object may be lying randomly in the photo and can be of any scale. So object classification isn't any extra beneficial when there are Multiple targets in photo, Objects are minute and Exact location and size of object in photograph is desired. Traditional methods of detection worried the use of a block-wise orientation histogram(SIFT or HOG) feature which couldn't acquire high accuracy in fashionable datasets consisting of PASCAL VOC. These techniques encode a compratelily lower stage characteristics of the objects and therefore are not capable to differentiate efficiently among various special labels. Shift invariant or space invariant artificial neural networks which are based on their shared-weights architecture also termed as Convolutional networks has implementable strategies , have turn out to be the state of the art in object detection in photograph. They assemble a representation in a hierarchical manner with increasing order of abstraction from lower to higher levels of neural network. One may want to perform detection by carrying out a class on distinctive sub-windows or patches or regions extracted from the image. The patch with excessive probability will no longer simplifies the magnificence of that area however also implicitly offers its area too in the image. Most of the tactics vary at the type of technique used for selecting the home windows. Any Deep Neural Network will encompass 3 varieties of layers particularly The Input Layer, The Hidden Layer and The Output Layer. We can say Deep Learning is the most recent term within the area of Machine Learning. It's a way to implement Machine Learning. Deep Learning is located and proves to have the excellent strategies with latest performances. Thus, Deep Learning surprise us and could maintain to achieve this in the near destiny. Recently, researchers are continuous in exploring Machine Learning and Deep Learning.

Keywords: *SIFT, HOG,* PASCAL VOC, Machine Learning and Deep Learning, Convolutional networks

ISSN:0971-1260
Vol-22- Issue-14-December-2019

1. Introduction

Convolutional Neural Networks (CNN) Approach:

When the data is small, Deep Learning algorithms don't perform well. This is the only reason DL algorithms need a large amount of data to understand it perfectly. Deep Learning depends on high-end machines while traditional learning depends on low-end machines. Thus, Deep Learning requirement includes GPUs. That is an integral part of it's working. They also do a large amount of matrix multiplication operations. Feature Engineering is a general process. here, domain knowledge is put into the creation of feature extractors to reduce the complexity of the data and make patterns more visible to learn the algorithm working. Although, it's very difficult to process. Hence, it's time consuming and expertise. DL algorithms needs to break a problem into different parts to solve them individually. And to get a result, combine them all.

Scenario:

Let us suppose we have a task of multiple object detection. In this task, we have to identify what the object is and where is it present in the image. Deep Learning takes more time to train as compared to Machine Learning. The main reason is that there are so many parameters in a Deep Learning algorithm.

This will be a tedious process from computational time point of view as each sub-window would require passing it through CNN and calculating the feature for that region. R-CNN therefore uses a object proposal algorithm like selective search in its pipeline which gives out a number(~2000) of TENTATIVE object locations and extents on the basis of local cues like color rgb, hsv etc. This does not use any fancy supervised algorithm and therefore is class-agnostic which can be used independent of the domain. These object regions are warped to a fixed sized(227X227) pixels and are fed to a classification convolutional network which gives the individual probability of the region belonging to background and classes. The tricky part is feeding the appropriate regions labelled as background during the training of convolutional network. If random regions that do not have anything to do with the object classes are fed as background the network wont be able to distinguish between the object regions and regions which are partially containing the objects. Therefore regions which have an IOU greater than 0.5 with the objects are marked with class of that object and those with overlap < 0.3 are marked as background. As in the classification training, SGD is used to train the network end to end. The second stage of RCNN involves improving the localization(coordinates of the extent of object) accuracy by minimizing the error of predicted coordinates against the ground truth coordinates. This is required because SS need not necessarily produces regions which can encompass the objects perfectly. For this a linear regression layer is optimized on top of Conv5 layer after fine-tuning the network for classification. Since this training is done independent of classification training, conv5 layer and layers preceding it cannot be fine-tuned once their weights have been optimized for classification(because we will be using the same network for both to reduce computation time). In the test time an additional technique Non Maximal Suppression is used to merge highly overlapping regions which are predicted to be of same class.

ISSN:0971-1260
Vol-22- Issue-14-December-2019

2. Widespread Evaluation of CNN

Image class is the venture of taking an enter picture and outputting a category (a cat, dog, etc) or a probability of training that nicely describes the picture. For humans, this project of reputation is one of the first talents we study from the moment we're born and is one that comes evidently and effortlessly as adults. Without even thinking two times, we're able to quick and seamlessly identify the environment we're in as well as the items that surround us. When we see an image or simply while we look at the sector around us, maximum of the time we're capable of symbolizing the scene and supply each item a label, all without even consciously noticing. These skills of being able to quickly apprehend patterns, generalize from earlier understanding, and adapt to exclusive picture environments are ones that are not present is machines.

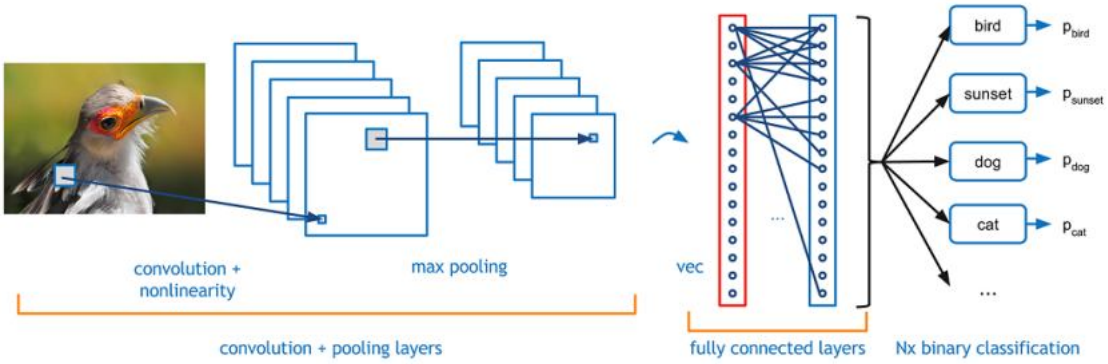


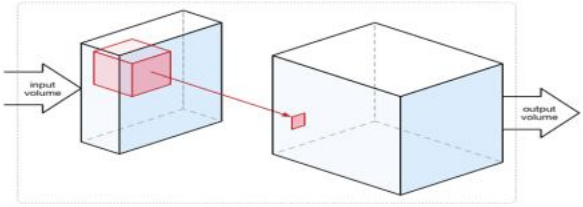
Fig.1 – Convolution Neural Networ Approach

When a PC sees a picture (takes an photo as enter), it's going to see an array of pixel values. Depending on the resolution and size of the picture, it will see a 32 x 32 x 3 array of numbers (The three refers to RGB values). Each of those numbers is given a value from zero to 255 which describes the pixel depth at that point. These numbers, whilst meaningless to us whilst we perform photograph classification, are the only inputs available to the PC. The idea is which you deliver the PC this array of numbers and it's going to output numbers that describe the chance of the photograph being a certain elegance (.Eighty for cat, .15 for dog, .05 for bird, and many others). Now that we realize the problem to identify the object in a scene by machines. What we need the computer to do is so one can differentiate between all of the pics and discern out the precise capabilities that make a dog a dog or that make a cat a cat. This is the process that goes on in our minds subconsciously as well. When we look at a image of a dog, we can classify it as such if the photo has identifiable features along with paws or 4 legs. In a similar way, the computer is able perform picture category through looking for low degree features including edges and curves, and then building up to greater summary principles via a chain of convolutional layers. This is a widespread evaluation of what a CNN does.

ISSN:0971-1260
Vol-22- Issue-14-December-2019

3. Convolutional layers

The real power of these deep learning networks, especially for image recognition, comes from convolutional layers.

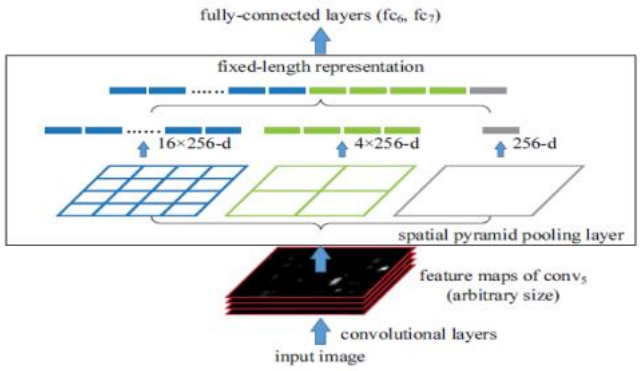


Experimental Analysis and Results

With a fully-connected layer, the input and output are one-dimensional vectors of numbers. However, a convolutional layer works on three-dimensional volumes of data (also called tensors). The output usually has the same width and height as the input volume but a larger depth. You can think of an image as a three dimensional "cube". The width and height are as usual, but each of the RGB components gets its own plane. Convolutional layers are great for image recognition because they scan the input much like a human eye does. Just as with the fully-connected layer, what we learn is weights — but here these weights are inside the convolution kernels.

Region based convolution neural networks (R-CNNS) approach:

Spatial Pyramid Pooling is one of the turning points for making a highly accurate RCNN pipeline feasible in run-time. RCNN had to pass all the ~2000 regions from SS independently through CNN and is therefore a very slow algorithm. Spatial Pyramid Pooling allows the whole image to be passed through the convolutional layer only once. This saves a lot of time because same patch may belong to multiple regions and convolutions on them are not calculated multiple time as done in RCNN thereby enabling a shared computation of convolution layers among the regions. Since major chunk of time(~90%) is spent on the convolutional layers it reduces the computation time drastically. After passing the image through convolution layers, independent feature need to be calculated for each of the regions generated by SS. This can be done by performing a pooling type of operation on JUST that section of the feature maps of last convolution layer that corresponds to the region.



ISSN:0971-1260
Vol-22- Issue-14-December-2019

The rectangular section of convolution layer corresponding to a region can be calculated by projecting the region on convolution layer by taking into account the down sampling happening in the intermediate layers. Normally a max pooling layer follows the final convolution layer of CNN, but the feature vector produced by max pool layer depends upon the size of region and therefore vector obtained cannot be fed into the following FC layers which require a fixed sized vector as input. Spatial Pyramid Pooling solves this problem by replacing max pooling layer with spatial pooling layer. Spatial Pyramid Pooling layer divides a region of any arbitrary size into constant number of bins and max pool is performed on each of the bins. Since number of bins remain the same, a constant size vector is produced. The training windows for positive and negative samples are chosen in the same manner as RCNN. The only difference in SPP net is that since windows of arbitrary size are used for pooling operation, back-propagation through Spatial Pyramid Pooling layer and fine tuning the network end-to-end is not trivial and therefore SPP net fine tunes just FC part of the network and gradients are not propagated across the Spatial Pyramid Pooling layer. The second part follows on the same lines as RCNN wherein it fits a linear regression layer for localization on top of conv5 layers.

4. CONCLUSION

Thus we created a new concept of convolutional neural network which is a class of deep neural networks, most commonly applied to analyzing visual imagery. They are also known as shift invariant or space invariant artificial neural networks, based on their shared-weights architecture and translation invariance characteristics. We discussed the problem faced by the object detection system which severely hapmpers the accuracy contents. All these methods concentrate on increasing the run-time efficiency of object detection without compromising on the accuracy. Faster-Rcnn has become a state-of-the-art technique which is being used in pipelines of many other computer vision tasks like captioning, video object detection, fine grained categorization etc.

REFERENCES

1. Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik: Rich feature hierarchies for accurate object detection and semantic segmentation
2. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun: Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition
3. Ross Girshick: Fast R-CNN
4. Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks http://www.cs.utoronto.ca/~fidler/slides/CSC420/lecture17.pdf
5. Matthijs Hollemans. Convolutional neural networks on the iPhone with VGGNet. http://matthijshollemans.com/2016/08/30/vggnet-convolutional-neural-network-iphone/, 2016.

ISSN:0971-1260
Vol-22- Issue-14-December-2019

Dr B. Annapurna, Assoc. Prof. in Computer Science and Engineerig , Aditya College of Engineering , Surampalem , East Godavari Dt., AP, India. She has an excellent teaching experience of 24 years. She acquired her PhD from Acharaya Nagarjuna University on Data Mining Domain. She is convenor for a number of International Conferences. She organised many National seminars and Workshops. To her thrust of interest in Research she published 23 research papers on various domains in several International conferences and in International Journals.

Raushan Kumar Singh is the M.Tech Gold Medalist from the department of Embedded System. He serves as the Technical Director of Spectrum Solutions, which is a Pondicherry, India based government approved academic event organising company. He chaired and organised several International Conferences / FDPs / Workshops / Project Expo / Guest Lectures and Robowars. He is the Editorial board member for several International Journals. He was Internationally Awarded by the First Vice President and the then Chief Justice of Supreme Court of Nepal for his excellence in Academic Events. His area of interest is immensely multidimensional.