International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 5, pp: 800-804 www.ijaem.net ISSN: 2395-5252

Pedestrian Detection by using deep learning

Varsha Shingade, pratiksha Dargude, Usha Bansode, Prof. Dr. Meenakshi Pawar
Corresponding author: Varsha Shingade

Date of Submission: 30-08-2020 Date of Acceptance: 12-09-2020

ABSTRACT: Pedestrian detection method widely used in autonomous driver assistance and security surveillance systems. Pedestrian detection is a practical application of object detection. pedestrian detection algorithm require design features to describe the pedestrian characteristics. With the development of convolutional neural networks (CNNs), deep feature can be learned. Training images will be divided into sub regions to reduce the influence of human appending, such as bags. The remaining regions are almost fixed regions. These fixed regions will be fed to convolutional neural network. To accept different type of images, an arbitrarily-sized pooling layer is used in CNN. Deep learning and convolutional neural network have made great success in image and audio, which is important component of deep learning.

Keywords: Pedestrian detection, convolutional neural network (CNN), deep learning, feature extraction.

I. INTRODUCTION

Pedestrian Detection is to determine whether the image contains a pedestrian and mark the specific area of people traveling. Where people flow much, such as train station, airport. Pedestrian tracking is meaningful for keeping security. Histogram of Oriented Gradient (HOG) feature is widely used in pedestrian recognition and tracking. In recent years, CNNs is mostly used in computer vision and image processing technique. It learns the deep features of training data. In generally, CNNs architecture always contains convolution layers which extract features of local regions, and pooling layers which reduce the number of features. It also contain fully-connected layers which is for classification. Convolution layers combined with pooling layer can extract main feature and reduce the number of feature simultaneously. CNNs architecture also used in image classification, image segmentation. Training processing used to learning parameters for training images. Weights and offsets are learned in the training processing. With the development of deep learning, deep learning based pedestrian detection has achieved great success without computing cost. With the development of computer vision, pedestrian detection in the intelligent auxiliary driving system, intelligent monitoring system, pedestrian analysis and intelligent robot, and other fields have been widely used.

Deep learning:

Deep learning is a subset of machine learning that uses multiple layers to progressively extract higher level features from the raw input.

Detection models can be mainly divided into three stages: informative region selection, feature extraction and detection.

A.Informative region selection :As person appear in any positions of the image and have different aspect ratios or sizes. It is a compulsory task to scan the whole image with a multi-scale sliding window. Although this exhaustive strategy can find out all possible positions of the person. Due to a huge number of candidate windows, it is computationally expensive and produces too many redundant windows. However, if only a fixed number of sliding window templates are applied there is a possibility of unsatisfactory regions in the given image.

B.Feature Extraction :To identify different objects, we need to extract visual features which can provide a semantic and robust representation. This is due to the features which can produce representations associated with complex cells in human brain. However, due to the diversity of appearances, illumination conditions and backgrounds, its difficult to manually design a robust feature descriptor to perfectly describe all kinds of objects.

C.Detection :A detector distinguishes a target perperson from all the other categories and to make the presentations more hierarchical and informative for visual recognition.

Usually, the Supported Vector Machine

International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 5, pp: 800-804 www.ijaem.net ISSN: 2395-5252

(SVM) and Deformable Part-based Model (DPM) are good choices. Among these classifiers, the DPM is a flexible model by combining object parts with deformation cost to handle severe deformations. In DPM, with the support of a graphical model, carefully designed low-level features and kinematically inspired part decompositions are combined and discriminative learning of graphical models permits for building high-precision part-based models for a variety of classes.

II. LITERATURE SURVEY

C. Yang, X. Yao, L. Guo.J. Han G. Cheng [1]

Pedestrian tracking technique is used in many intelligent systems, like video surveillance, security regions. But many methods suffer from illumination, human posture or human appendant. With the help of Convolutional Neural Networks (CNNs), deep feature can be learned. In this paper, training images will be divided into sub regions to reduce the influence of human appendant, such as bags. The remain regions are almost fixed regions. Then these fixed regions will be fed into our CNNs for learning deep features. In order to copy with different sizes of training images, an arbitrarily-sized pooling layer is developed in our CNN architecture. Then, these deeply-learned feature vector can be used in pedestrian recognition. Optical flow is used for pedestrian tracking. Experimental results show our proposed method can achieve pedestrian tracking effectively.

P.C. Mouli R. Soundrapandiyan [2]

Infrared pedestrian classification plays an important role in advanced driver assistance systems. However, it encounters great difficulties when the pedestrian images are superimposed on a cluttered background. Many researchers design very deep neural networks to classify pedestrian from cluttered background. However, a very deep neural network associated with a high computational cost. The suppression of cluttered background can boost the performance of deep neural networks without increasing their depth, while it has received little attention in the past. This study presents an automatic image matting approach for infrared pedestrians that suppresses the cluttered background and provides consistent input to deep learning. The domain expertise in pedestrian classification is applied to automatically and softly extract foreground objects from images with cluttered backgrounds.

X. Yin , X. Liu G. Brazil [3]

A set of images that contain objects from a common category, object co-segmentation aims at automatically discovering and segmenting such common objects from each image. During the past few years, object co-segmentation has received great attention in the computer vision community. However, the existing approaches are usually designed with misleading assumptions, unscalable priors, or subjective computational models, which do not have sufficient robustness for dealing with complex and unconstrained real-world image contents.

D. Zhang, G. Cheng, N. Liu, D. Xu, J. Han [4]

Multispectral images of color-thermal pairs have shown more effective than a single color channel for pedestrian detection, especially under challenging illumination conditions. However, there is still a lack of studies on how to fuse the two modalities effectively. In this paper, we deeply compare six different convolutional network fusion architectures and analyze their adaptations, enabling a vanilla architecture to obtain detection performances comparable to the state-of-the-art results..

R. Quan, D. Zhang, F. Nie, J. Han [5]

Pedestrian tracking technique is now widely used in many intelligent systems, such as video surveillance, security regions. But many methods suffer from illumination, human posture or human appendant. With the development of Convolutional Neural Networks (CNNs), deep feature can be learned. In this paper, training images will be divided into sub regions to reduce the influence of human appendant, such as bags. The remain regions are almost fixed regions. Then these fixed regions will be fed into our CNNs for learning deep features.

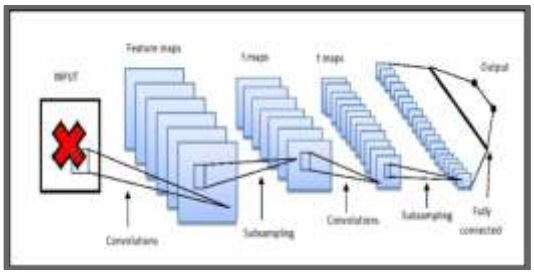
III. INTRODUCTION TO DEEP LEARNING

The deep learning is an artificial intelligence function that imitates the workings of the human brain in processing data and creating patterns for decision making. It is a subset of machine learning in artificial intelligence which has networks capable of learning unsupervised from data.

A.CNN: The CNN is nothing but Convolution Neural Network. A CNN includes an input,an output layer and multiple hidden layers. The hidden layers of a CNN have a series of convolutional layers that convolve with a multiplication or other dot product. The activation function is generally a RELU layer which is followed by additional convolutions such as

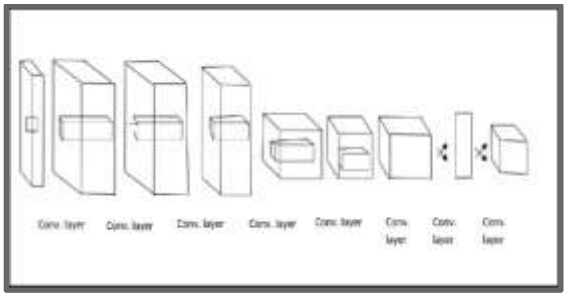
International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 5, pp: 800-804 www.ijaem.net ISSN: 2395-5252

pooling layers, fully connected layers and normalization layers, referred to as hidden layers as their inputs and outputs are masked by the activation function and final convolution. It is generally a sliding dot product or cross-correlation. These CNN layers convolve the input and pass the result to the next layer.



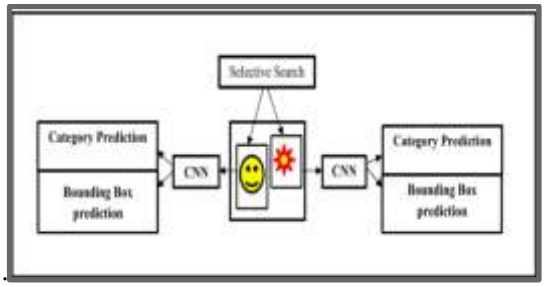
Fig(1): CNN

B.YOLO: The YOLO (You Only Look Once) uses deep learning and convolutional neural networks (CNN) for object detection. It only needs to see each image once. It allows YOLO to be one of the fastest detection algorithms. It can detect objects in real time up to 30 FPS. For the detection, the image is divided in a grid of S*S (left image). Each cell will predict N possible bounding boxes and the level of probability of each one of them. This means S*S*N boxes are calculated.



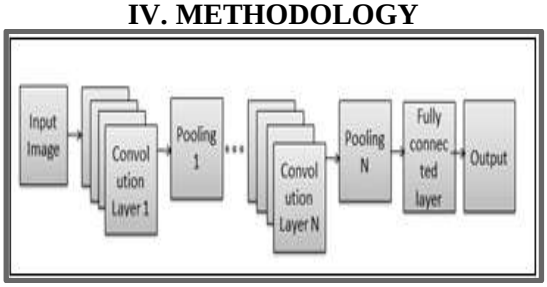
Fig(2): Yolo

C.R-CNN : The R-CNN is nothing but Region based CNNs. The R-CNN model first selects several proposed regions from given image and then label their categories and bounding boxes. It uses a CNN to perform forward computation to extract features from each proposed area. Then we use the features of each proposed region to predict their categories and bounding boxes



Fig(3): R-CNN

IV. METHODOLOGY



Fig(4): Basic architecture of CNN model

Basic architecture of CNN model consists of

1. Input Image
2. Convolution Layer
3. Pooling Layer
4. Fully Connected Layer
5. Output

1. Input image

The input image is the image given to the model to check the output by performing various functions on it. It is given to the block named as convolution layer.

2. Convolution Layer

Convolution is the first layer to extract features from the given input image. Convolution restores the relationship between pixels by learning image features using small squares of input data. It is a mathematical operation which takes two inputs such as image matrix and a filter.

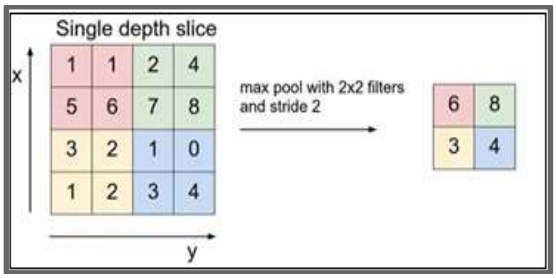
3. Pooling Layer

The section of pooling layers reduces the number of parameters when the images are too large. Spatial pooling also called subsampling or down sampling that reduces the dimensionality of each map without changing the important information. Spatial pooling can be of different types:

a) Max Pooling
b) Average Pooling
c) Sum Pooling

International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 5, pp: 800-804 www.ijaem.net ISSN: 2395-5252

Max pooling take the largest element from the rectified feature map. Taking the largest element could also take the average pooling. Sum of all elements in the feature map call as sum pooling.

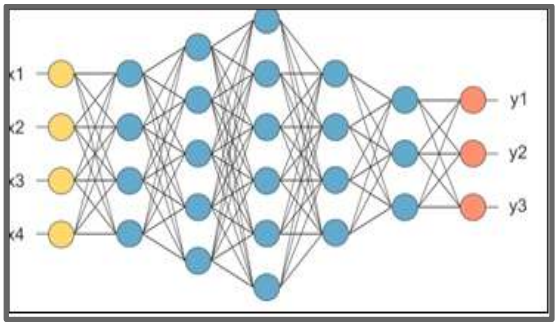


Fig(5). Pooling

4. Fully Connected Layer

The layer we call as FC layer, we flattened our matrix into vector and feed it into a fully connected layer like neural network.

The fully connected layer combines all the features together to create a model. Finally an activation function such as softmax or sigmoid is used to classify the outputs as cat, dog, car, truck etc.



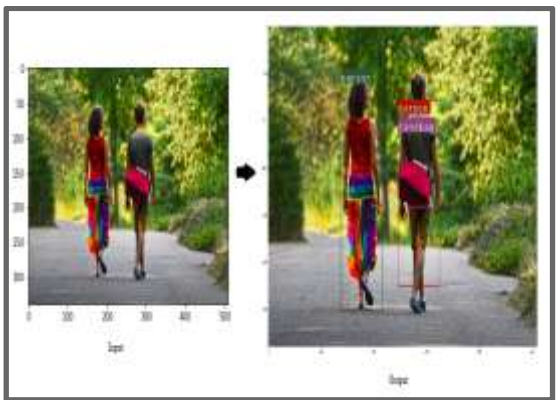
Fig(6): After pooling layer, flattened as FC layer

5.Output

The output image provides the result of the model. On the basis of the fully connected layer the output is carried out.

V. RESULT

The performance of a model for pedestrian detection is evaluated using the precision and recall across each of the best matchingbounding boxes for the known objects in the image.



VI. CONCLUSION

This paper firstly introduces the classical methodologies of pedestrian detection in deep learning. Then it clarifies the ideas of model design and the limitations of deep learning method by overviewing the early object detection methods based on deep learning. Afterwards, it elaborates on the common object detection model based on deep learning. Finally, this paper makes a further analysis of the challenges in object detection based on deep learning, and offers some solutions for reference.

VII. ACKNOWLEDGEMENT

We feel happy in forwarding this project report as an image of sincere efforts. The successful project reflects our work, effort of our guide.

We give special thanks to our respected guide Dr.M.M.Pawar for her constant interest and constant encouragement throughout the completion of our project.

We are also equally indebted to our principal Dr. B. P. Ronge for his valuable help whenever needed.

We express deep gratitude to all staff members who lend us their valuable support and co-operation to enable us to complete our project successfull

REFERENCES

[1]. C. Yang, X. Yao, L. Guo .J. Han G. Cheng,"When deep learning meets metric learning : remote sensing image scene classification via learning discriminative CNNs ,"IEEE Trans. Geosci Remote Sens., vol. 56, no. 5, pp. 0196-2892, may 2018.

[2]. P.C. Mouli R. Soundrapandiyan , "Adaptive pedestrian detection in infrared images using fuzzy enhancement and top hat transform, "International J.Comput. Vis. Robot, vol. 7, no. 1-2, pp. 49-67, 2017.

[3]. X. Yin , X. Liu G. Brazil, "Illuminating pedestrians via simultaneous detection

&segmentation,"Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 37, no. 8, pp. 4950-4959, 2017.

[4]. D. Zhang, G. Cheng, N. Liu, D. Xu, J. Han, "Advanced deep-learning techniques for salient and category specific object detection ,"IEEE Signal process, vol. 35, no. 1, pp. 84-100, Jan 2018.

[5]. R. Quan, D. Zhang, F. Nie, J. Han, "Robust object co-segmentation using background prior,"IEEE Trans. Image Process, vol. 27, no. 4, pp. 1639–1651., 2018