" Multiclass Object detection And Counting Using Classification"

Asmita Sonawane[1], Ninad khedkar[1], Shubham Jadhav[1], Aaditya giri[1]

Prof. A.m.Karanjkar[2]

[1]*Department of Computer Engineering, SCOE, Pune, Maharashtra, India*

[2]*Assistant Professor, Department of Computer Engineering, SCOE, Pune, Maharashtra, India*

Abstract

This paper designed how to recognize and count objects in a real time manner in a highspeed inspection environment with large volumes of data, so as to verify the concept (smart camera with GPU cores) we proposed. A wide, active, and challenging field of computer vision is real-time object detection. Real-time object detection and counting is a vast. vibrant yet inconclusive area of computer vision. Image localization refers to the process of finding a single object in an image, while object detection refers to the process of finding several objects in an image. This recognizes a class of semantic items in digital photos and movies. Smart camera is equipped with processor, memory, communication interface and operating system, so it can process large amounts of data in advance to assist follow-up automatic inspection and judgment. Real-time object detection has a variety of uses, including object tracking, video surveillance, people counting, pedestrian detection, selfdriving automobiles, facial recognition, sports ball tracking, and more. When detecting objects with the aid of OpenCV a library of programming functions primarily geared toward real-time computer vision, Convolution Neural Networks is a representative technique of deep learning.

Keywords: *Multi-class object counting Multi-class object counting dataset, Crowd Counting Counting with classification*

1.Introduction:

The goal of this study is to determine how many of a particular 3D object there are in an image. An essential intermediate step in solving the item counting problem is object recognition, which is often carried out using two planar photographs of the same object taken from various angles. For the management of crowds, traffic, and environmental wildlife, counting objects in photos is crucial. We provide an end-toend deep learning method for general object counting as an alternative to creating an object-specific method.The challenge of generic object counting is challenging. The difficulty of the challenge may be the driving force for the additional annotations that modern counting techniques add to the final object count. A computer vision technique known as object detection and counting with classification identifies and counts the items in an image or video stream by combining object detection with classification.

Drawing bounding boxes around things in an image or video stream is the process of object detection Convolutional neural networks (CNNs), which are trained on big datasets of labelled images, are a common deep learning approach used for this. An object is classified when a label or category is given to it based on its characteristics. To determine the type of thing within the bounding box, such as a person, car, or animal, classification is employed in the context of object detection and counting Combining object detection with classification enables for a variety of automated object identification and counting applications, including surveillance, traffic monitoring, and object recognition in medical imaging. This method can also be applied to more complicated situations, like object tracking, where the position of the object is tracked throughout time continually

In the grand scheme of things, object recognition and counting with categorization is a potent tool in the field of computer vision, with numerous useful applications in various domains

2. Related Work:

Junyu Gao, Qi Wang, Yuan Yuan

Recently, crowd counting is a hot topic in crowd analysis. Many CNN based counting algorithms attain good performance. However, these methods only focus on the local appearance features of crowd scenes but ignore the large-range pixel-wise contextual and crowd attention information. To remedy the above problems, in this paper, we introduce the **S**patial-/**C**hannel wise **A**ttention Models into the traditional

Regression CNN to estimate the density map, which is named as "SCAR". It consists of two modules, namely Spatial-wise Attention Model (SAM) and Channel-wise Attention Model (CAM). The former can encode the pixel-wise context of the entire image to more accurately predict density maps at the pixel level. The latter attempts to extract more discriminative features among different channels, which aids model to pay attention to the head region, the core of crowd scenes. Intuitively, CAM alleviates the mistaken estimation for background regions. Finally, two types of attention information and traditional CNN's feature maps are integrated by a concatenation operation. Furthermore, the extensive experiments are conducted on four popular datasets, Shanghai Tech Part A/B, GCC, and UCF CC 50 Dataset.
The results show that the proposed method achieves state-of-the-art results.

Qi Wang,Junyu Gao,Wei Lin,Yuan Yua School of Computer Science and Center for

OPTical IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an, Shaanxi, P. R.

Recently, counting the number of people for crowd scenes is a hot topic because of its widespread applications (e.g. video surveillance, public security). It is a difficult task in the wild: changeable environment, large-range number of people cause the current methods can not work well. In addition, due to the scarce data, many methods suffer from overfitting to a different extent. To remedy the above two problems, firstly, we develop a data collector and labeler, which can generate the synthetic crowd scenes and simultaneously annotate them without any manpower. Based on it, we build a large-scale, diverse synthetic dataset. Secondly, we propose two schemes that exploit the synthetic data to boost the performance of crowd counting in the wild: 1) pretrain a crowd counter on the synthetic data, then finetune it using the real data, which significantly prompts the model's performance on real data; 2) propose a crowd counting method via domain adaptation, which can free humans from heavy data annotations. Extensive experiments show that the first method achieves the state-of-the-art performance on four real datasets, and the second outperforms our baselines. The dataset and source code are available at https://gjy3035.github.io/GCC-CL/.

Weizhe Liu ,Mathieu Salzmann, Pascal Fua Computer Vision Laboratory, Ecole

Polytechnique F'ed´ erale de Lausanne (EPFL)´

State-of-the-art methods for counting people in crowded scenes rely on deep networks to estimate crowd density. They typically use the same filters over the whole image or over large image patches. Only then do they estimate local scale to compensate for perspective distortion. This is typically achieved by training an auxiliary classifier to select, for predefined image patches, the best kernel size among a limited set of choices. As such, these methods are not endto-end trainable and restricted in the scope of context they can leverage.

In this paper, we introduce an end-to-end trainable deep architecture that combines features obtained using multiple receptive field sizes and learns the importance of each such feature at each image location. In other words, our approach adaptively encodes the scale of the contextual information required to accurately predict crowd density. This yields an algorithm that outperforms state-of-the-art crowd counting methods, especially when perspective effects are strong

Yuhong Li[1,2], Xiaofan Zhang[1], Deming Chen[1] 1

We propose a network for Congested Scene Recognition called CSR Net to provide a datadriven and deep learning method that can understand highly congested scenes and perform accurate count estimation as well as present high quality density maps. The proposed CSR Net is composed of two major components: a convolutional neural network (CNN) as the front-end for 2D feature extraction and a dilated CNN for the back-end, which uses dilated kernels to deliver larger reception fields and to replace pooling operations.

CSR Net is an easy-trained model because of its pure convolutional structure. We demonstrate CSR Net on four datasets (ShanghaiTech dataset, the UCF CC 50 dataset, the WorldEXPO'10 dataset, and the UCSD dataset) and we deliver the state-of-the-art performance. In the ShanghaiTech Part B dataset, CSR Net achieves 47.3% lower Mean Absolute Error (MAE) than the previous state-of-theart method. We extend the targeted applications for counting other objects, such as the vehicle in TRANCOS dataset. Results show that CSR Net significantly improves the output quality with 15.4% lower MAE than the previous stateofthe-art approach.

3.Algorithm:

a)Yolov3

Joseph Redmon and Ali Farhadi created the object identification method known as YOLOv3 (You Only Look Once version 3). It uses a cutting-edge deep learning system to recognise items instantly.

YOLOv3 has the following salient characteristics

Faster and more precise: YOLOv3 is quicker and more precise than YOLOv2, its forerunner. It is useful for real-time applications because of its great precision and quickness in object detection

Multi-scale detection: YOLOv3 can detect objects at various scales thanks to its multi-scale detection method. As a result, it is more resistant to changes in object size.
Deep convolutional neural networks (CNNs) are used by YOLOv3 to extract features from the input image. 53 convolutional layers make up this network, which is trained on a a big image collection.

Anchor boxes: To increase its accuracy, YOLOv3 makes use of anchor boxes. The bounding boxes of objects in a picture are predicted using anchor boxes, which are pre-defined boxes of various sizes and aspect ratios.
Non-maximum suppression: To get rid of duplicate detections, YOLOv3 employs nonmaximum suppression. This algorithm evaluates overlapping detections' confidence values and keeps the detection with the greatest score.

Overall, YOLOv3 is a potent object identification system with numerous real-world uses in industries including robotics, surveillance, and autonomous vehicles.

b) Dataset

A common dataset for object detection, segmentation, and captioning applications is the Common Objects in Context (COCO) dataset. It has more than 330,000 photos and more than
2.5 million instances of labelled objects in 80 different categories.

A polygonal segmentation format is used for the annotations in the COCO dataset to offer accurate outlines of the objects in the photos. This makes it a useful dataset for tasks like instance segmentation and keypoint recognition that call for the precise localisation of objects

The COCO dataset's primary characteristics include:

vast and diverse amount of examples: The COCO dataset's more than 2.5 million item instances offer a vast and varied group of examples for training object detection algorithms.

Annotations for object segmentation: COCO's polygonal segmentation annotations give clear descriptions of objects, making it a useful dataset for activities like instance segmentation and keypoint identification.

80 different item categories are covered by the annotations in COCO, which span a variety of typical things

Annotations gathered through crowdsourcing: The COCO dataset was annotated by crowdsourcing, which helps to ensure that the annotations are correct and diverse

Modern models have been trained using the COCO dataset, which has become a benchmark for assessing object detection algorithms. On the COCO website and through a number of deep learning frameworks, including TensorFlow and PyTorch, the dataset can be downloaded

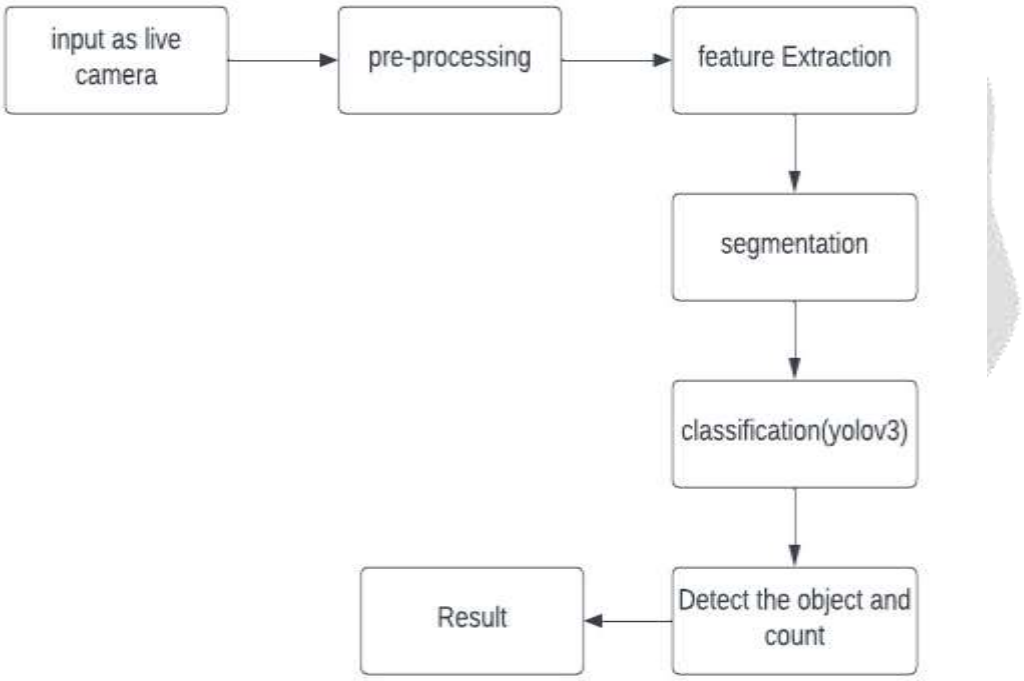


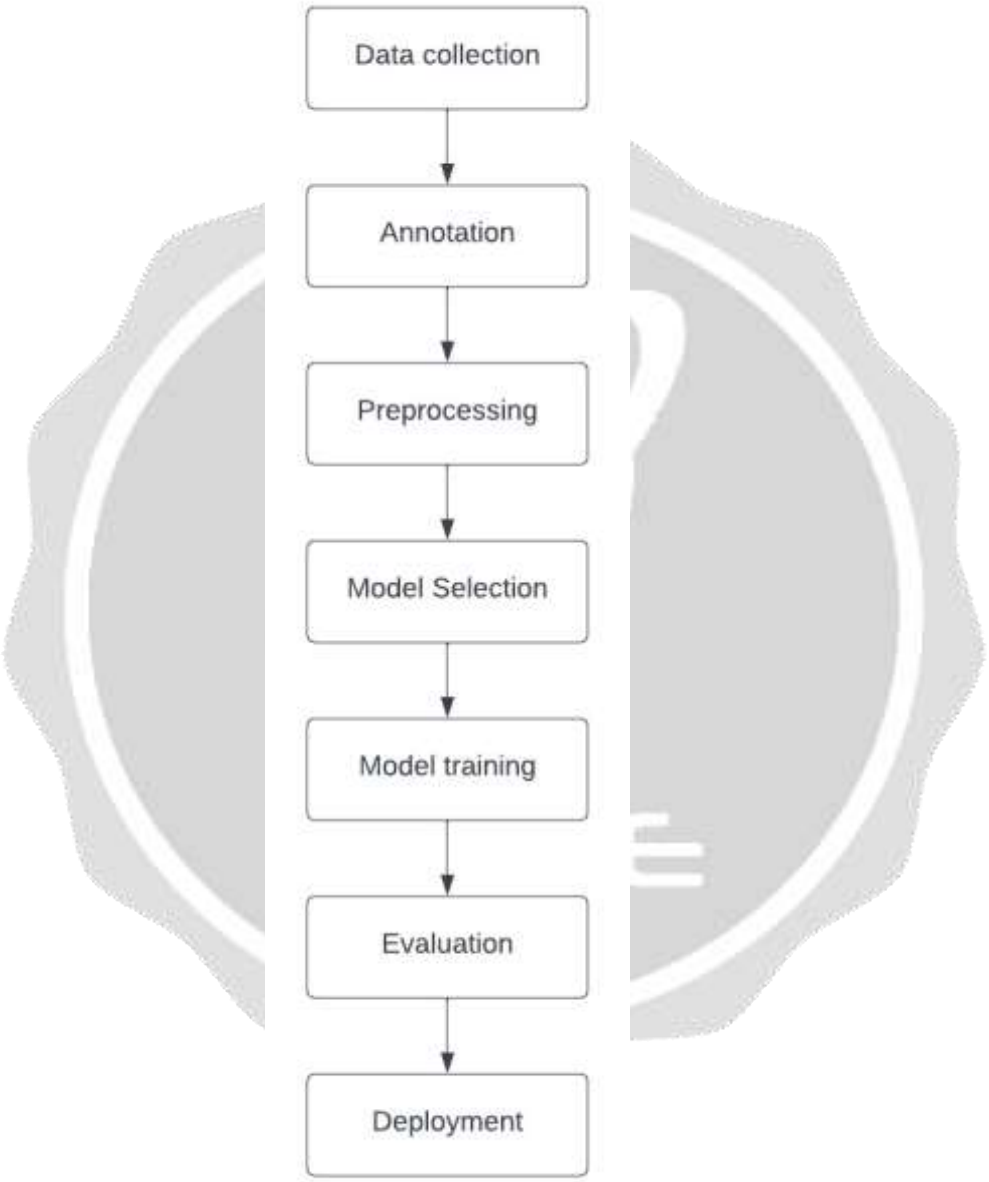
Figure: Proposed System

4. Implementation:

Overall, the implementation process for object detection and counting involves a combination of data collection, annotation, preprocessing, model selection, training, evaluation, and deployment. It requires expertise in computer vision, machine learning, and programming. **Implementation involves the following steps:**

1) Data collection
2) Annotation
3) Preprocessing

4) Model Selection
5) Model Training
6) Evaluation
7) Deployment



1)Data Collection: The first step is to collect a dataset of images that contains the objects that need to be detected and counted. The dataset should be diverse and contain a wide variety of objects in different sizes, orientations, and lighting conditions.

2)Annotation: The images in the dataset should be annotated with bounding boxes around the objects that need to be detected and counted. This annotation process can be done manually or using automated tools.

3)Preprocessing: The images in the dataset should be preprocessed by resizing, cropping, and normalizing the images to improve the performance of the model.

4)Model Selection: The next step is to select a suitable object detection model that can accurately detect and count the objects in the images. Some popular models include YOLO (You Only Look Once), Faster R-CNN (Region-based Convolutional Neural Network), and SSD (Single Shot MultiBox Detector).

5)Model Training: The selected model should be trained on the annotated dataset to learn how to detect and count the objects. This involves feeding the images and their corresponding annotations into the model and adjusting its parameters to minimize the prediction errors.

6)Evaluation: The trained model should be evaluated on a test set of images to measure its performance in terms of accuracy and speed. The evaluation metrics can include precision, recall, F1 score, and mean average precision (mAP).

7)Deployment: Once the model has been trained and evaluated, it can be deployed in a real-world application to detect and count objects in new images. The deployment can be done on a local machine or on a cloud-based platform depending on the requirements of the application.

5.Result and Discussion:

Figure : Specific Object

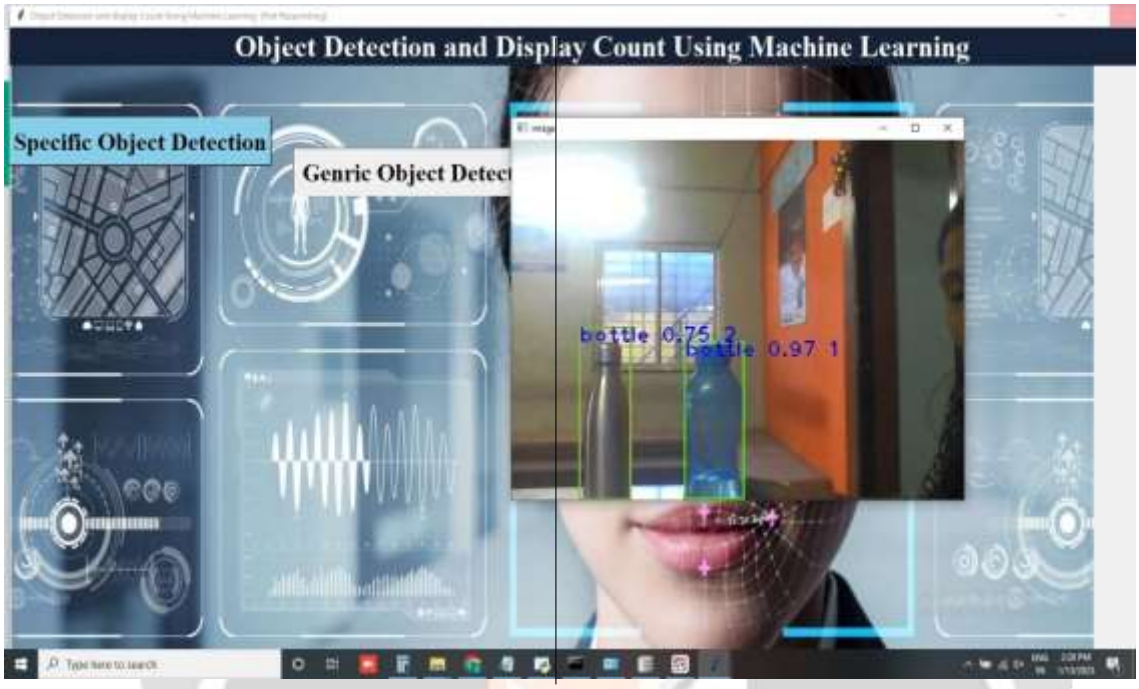


Figure: Generic O

6.Conclusion and Future Scope:

In this application, we are detecting object using a live camera.Yolo v3 algorithm is used for detection of object. We can also keep a count of object after the detection.This work proposes an approach towards generic object counting with unsupervised local image information. Moreover, we propose to learn from local image features, and predict global image object counts**.**

In future wide , active and challenging stream of computer vision is real time object detection . we can detect a single object in image or virtually and this can recognize class of symantic item in digital photos and movies.the platform like yolov3 algorithm is a programming function it can handle the real time computer vision easily in future there is variety of usess like object tracking, video surveillance,people counting,self driving automobile,and many more.currently this platform detect the images which are in dataset but in future we can find unique techniques for better results**.**

7. Acknowledgement:

We sincerely appreciate Prof. A.M. KARANJKAR, our mentor, for his support, advice, and encouragement in this work.

8.References:

1.Junyu Gao, Qi Wang, Yuan Yuan School of Computer Science and Center for OPTical IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an 710072, P. R. China

2. Qi Wang, Junyu Gao, Wei Lin, Yuan Yuan School of Computer Science and Center for OPTical IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an, Shaanxi, P. R. China

3. Weizhe Liu Mathieu Salzmann Pascal Fua Computer Vision Laboratory, Ecole Polytechnique F ´ ed´ erale de Lausanne (EPFL

4. Yuhong Li1,2 , Xiaofan Zhang1 , Deming Chen1 1University of Illinois at Urbana-Champaign 2Beijing University of Posts and Telecommunications

5. Junyu Gao, Qi Wang, and Yuan Yuan, "Scar: Spatial-/channel-wise attention regression networks for crowdcounting," Neurocomputing, vol. 363, pp. 1–8, 2019

6. Qi Wang, Junyu Gao, Wei Lin, and Yuan Yuan, "Learn-ing from synthetic data for crowd counting in the wild,"in Proceedings of the IEEE conference on computer vi-sion and pattern recognition, 2019, pp. 8198– 8207

7. Yingying Zhang, Desen Zhou, Siqin Chen, Shenghua Gao, and Yi Ma, "Single-image crowd count via multi-column convolutional neural network," in Pro ceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 589–59