IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

OBJECT DETECTION AND CLASSIFICATION USING SPARSITY REGULARIZED PRUNING ON LOW QUALITY IMAGE/VIDEO

[1]Shakil Abdul Rajjak Shaikh, [2]Laxami Rama Wadekar

[1]Assistant Professor, [2]PG Student

[1]Electronics Engineering, Electronics and Telecommunication Engineering

[1,2]Pravara Rural Engineering College, Loni, Dist: Ahmednagar, Maharashtra, India

Abstract: In the context of computer vision, object tracking is a crucial challenge. Image noise, variations in backdrop illumination, dynamic object motion, and partial and complete occlusion of objects all add to the complexity of object tracking. i.e. video of poor quality. Despite the impressive outcomes of deep learning, there has been less emphasis on low-resolution object recognition as compared to high-resolution images. We proposed a framework for reducing model size state-of-the-art single stage detector i.e. YOLOv4 with pruning and Kalman filtering for tracking the low quality video. The framework addresses compression in the following stages: Sparsity Induction, Filter Choice, and Filter Pruning. The detector model for detecting object is sparsified in the Sparsity by using a better global threshold. In the Filter Selection & Pruning stage, we choose and remove filters based on sparsity metrics of filter weights in two subsequent convolutional layers. As a result, the model is smaller than the majority of compact designs available now on the market. We test our framework's performance with a MSCOCO datasets. The model with compressing algorithm with having Kalman filter to produce good quality image representation on the low-quality input image or video feed analysis proposed for the condition. The models size examines for the inferences and the analytical framework for the better represent for the detection stages and evaluation results.

***Index Terms* - Single Shot Detector, Filter Pruning, Sparsity, Kalman filter, YOLOv4.**

I. INTRODUCTION

For years, automatic video recognition of moving objects has been one of the most rapidly developing topics in video image processing due to the great variety of fields that could potentially benefit from such advancement.. Robotic vision, medical imaging, space exploration, remote monitoring, and video surveillance are among the various fields that have attracted researchers to the problem of detecting and extracting information from moving objects. Several algorithms for object detection and tracking in low-quality video have been proposed over the last decade, but none has yet to clearly surpass the human vision system, leaving potential for new researchers to come up with new ideas on how to enhance existing approaches or invent new ones.

Despite deep learning's widespread success, a conventional deep model is difficult to install on resource-constrained devices, such as embedded devices mobile phones or mobile phones. A resource restricted situation is one in which a computational activity must be completed with a finite amount of resources, such as processing time, memory space, battery power, and so on. One of the most key limitations of deep neural networks is their high computing and memory overhead, which poses a significant problem for a mobile device. Pruning is one of the most popular methods to reduce network complexity which has been widely studied in the model compression community [2]. In the domain of visual surveillance, multiple object tracking is a extensively invastigated topic. The major source of its difficulty is because observed data are commonly corrupted by noise, missing information, and clutter, for example, in a blob-based entity, similar circumstances can arise, and detection approaches relying on background subtraction [6], where under-segmentation, over-segmentation and incorrect recognition are frequent problems, as a result, when tracking is done, failures occur. In

In the area of computer vision, object tracking is a critical problem. The difficulty in tracking of object is related to image noise, changes in background illumination, intricate object motion, and partial and complete occluded objects, i.e. low quality video. Despite the impressive outcomes of deep learning, there has been less emphasis on low-resolution object recognition than on high-resolution images. The average resolution of the images used in ILSVRC is 482 × 415 pixels [10]. Despite the presence of background and several items in such images, deep networks were able to extract rich visual features and achieve significant classification performance by extracting enough information about each object. Deep networks created for high-resolution object recognition may or may not perform well when categorising extremely low-resolution images in which most of the crucial object-related data has been lostcrushed. To improve the object recognition performance on low resolution images, image super-resolution (SR) techniques can be applied before the recognition step. In the past, computational barriers have limited the complexity of real-time video processing applications. As a result, most systems were either too slow to be useful, or they succeeded by limiting themselves to very specific conditions. Researchers have now been able to consider more complex, robust models for real-time analysis of streaming data because to newer technology. Researchers can now begin modelling real-world processes under a variety of situations using these new methodologies. Take, for example, the issue of video surveillance and monitoring. A reliable system should not be reliant on camera positioning. It should also be able to withstand anything is in its viewing field as well as any lighting effects. It should be able to handle movement through congested areas, overlapping objects in the visual field, shadows, lighting changes, and the impacts of moving parts of the scene (e.g. swaying trees), objects that move slowly and newly added things or withdrawn from the scene. In these instances, traditional procedures relying on back grounding methods usually fail. [15].

The use of video cameras for security reasons has increased in recent times. Identify a person with automatic face detection systems have greater importance today; but the low-quality of the videos make it difficult and are still an open problem that many researchers are trying to solve. Video cameras used in surveillance are becoming increasingly important in large and small organizations. Those seek to keep their physical assets or capital safe from the possibility of theft, robbery or illicit activities of their staff. Cities are increasingly reliant on video monitoring to prevent or detect criminal actions, traffic issues, or accidents which could occur. One task performed in surveillance videos is face recognition. It is carried out by analyzing external characteristics of people's faces. Although all current face recognition systems have reached a certain level of maturity, the development in video remains limited due conditions presented in real environments. For example, the process of recognize faces in images obtained from outdoor videos changes the illumination conditions, occlusion with others objects, angle of view or low-resolution of the acquired images. All characteristics of these frames make more difficult apply techniques such as face detection or recognition which were originally designed for images in semi-controlled environments [19] In this paper, we presented a new method to automate the process of image analysis and objects detection in real low quality video. We use Yolo-tiny model with backbone is CSP Darknet 53 which is replaced with Resnet 152 with the pruned method. The Kalman filter help to find the output detected classes in the frames to frame for the tracking the distance of the identified object and provide best result to surpass the reidentification in the sudden frame. .

II. Related Work

Sankar K. Pal et al. [1] The challenges in the tracking task are mostly caused by: i) improper or imperfect object recognition, ii) determining if an object is a true incomer or not, iii) proper linkage among detection and tracking and iv) false alarms occurrence. There are further certain issues concerning tracking as follows: Although a large number of studies has been done to solve the MOT problem for a single class, the same for multi-class problems is not yet much explored. Jian-Hao Luo et al. [2] proposed ThiNet, an effective channel-wise pruning method for deep model acceleration and compression. We showed that the proposed scheme can significantly improve model pruning performance over existing methods. Laith Alzubaidi et al. [3] explain the difficulties and remedies to assist investigators in understanding the research gaps that exist. It's followed by a list of the most popular DL programs. The impact of computational technologies such as FPGA, GPU, and CPU on DL is summarized. Sheng Ren et al. [4] demonstrated an efficient video detecting object using super-resolution for public security using a deep fusion system. Firstly, designed a super-resolution framework for video detection objects. Authors suggested a deep learning-based intelligent video identification object super-resolution (SR) method by combining object detection algorithms, video key frame selection algorithms, and super-resolution reconstruction methods. Secondly, we designed a regression-based object detection algorithm and a key video frame selection algorithm. Police and security personnel use the object detection algorithm to track suspected things in real time.

Tanvir Ahmad et al. [5] described humans can easily discover and recognise several objects in images or videos, regardless of their appearance, but computers find it difficult to identify and discriminate between things. For object detection, a modified YOLOv1 based neural network is proposed by authors. A. Ali and K. Terada [6] presented a solid framework for tracking multiple humans, which combines the Kalman filter and the accelerated mean shift method. The system begins by recognizing the object using a change detection technique and then tracks it by tracking the discovered object's blob center. Jeevith S. H. and Lakshmikanth S. [7] demonstrated moving detection and tracking of object for dynamic situations, particularly in the design of video surveillance systems, is a difficult task in Computer Vision, according to the author. the updated BGS proposes adaptive threshold detection for efficiently identifying moving objects in the foreground utilizing background. The Kalman filter is utilized to track a moving object in this suggested work. Sanjivani Shantaiya et al. [8] presented a tracking method for processing video data in order for a machine vision system to track it. It sums up by allowing several things to be tracked at the same time. It can solve several multi-object tracking issues, such as object appearance and disappearance, as well as object missing.

Emadeldeen Noureldaim et al. [9] attempt within the context of Kalman Filtering, to better tracking of several moving objects Given that the combined PCA-GMM beats the traditional GMM in terms of segmentation performance. The use of the PCA-GMM-KF approach to each case and under a succession of photos successfully captured the objects and produced good results, according to the experimental results of tracking objects. Jeongin Seo and Hyeyoung Park [10presents a collaborative training method for recognising very low resolution objects that includes an IEN and an object recognition network. The IEN can generate images with better quality in terms of appearance and perception using training signals from the object recognition network. C I Patel and R Patel [11] [17] demonstrated many methods for locating moving objects in video sequences, followed by a new strategy based on the Gaussian mixture model. The algorithm's usefulness is demonstrated by preliminary testing results in a variety of challenging conditions, including changes in illumination, object displacement, comparable background and foreground colour, shadow objects, and non-static backgrounds. Yehui Tang et al. [12] suggested a manifold regularised dynamic pruning strategy to excavate the redundancy of neural networks to the greatest extent possible. Further investigate the lot of data in the sample space to uncover the relationship between various instances from two perspectives, namely, complexity and similarity, and then preserve the association in the associated sub-networks. The instance complexity and network complexity are aligned using an adaptive penalty weight for network sparsity, but the similarity relationship is retained by matching the similarity matrices.

Porikli, F. and Yilmaz, A. [13] draw attention to key trends and provide a taxonomy of current methodologies in the goal of enabling the integration of detection and tracking of object for more effective business-oriented video analytics. Kenshi Saho [14] demonstrated use of Kalman filters to track moving objects is covered, as well as how to design them effectively using steady-state performance analysis. A dynamic analysis for tracking is initially defined, including only position and both position-velocity measurements. The author then summarizes issues with Kalman filter design in tracking systems and proposes an efficient steady-state performance index. Chris Stauffer and W.E.L Grimson [15] presented a unique, probabilistic background subtraction method It includes creating a separate mixture model for each pixel. The authors developed a robust and reliable real-time approximate technique. The procedure just needs two parameters: α and T. These two factors are unaffected by different cameras or scenarios.

Manoj Attri et al. [16] suggested background subtraction method is used to detect the item, and the identified results are improved utilising morphological operators. The rate of misclassification was lowered because to these operators. The proposed approach works better in terms of qualitative results after morphological operation, as shown by the experimental findings. Hitesh A Patel, and Darshak G Thakore [18] a visual surveillance system with moving object recognition and tracking capability using Kalman filters has been proposed. Rolando Jesus Cardenas T et al. [19] demonstrated based on the parallel Gunnar Farneback algorithm for image sequences (optical flow), Haar Cascades, and Local Binary Patterns, an unique method for detecting faces on low-quality videos was proposed. The model does not use techniques like illumination normalization or super-resolution, which are extensively used in the literature. In the future, each technique utilized will be examined in greater depth in order to process video with GPU in real-time. Zuojin Li et al. [20] researchers provides a new approach for grouping SURF features based on a cascade framework algorithm and the complicated background of fishing environment surveillance.

III. Proposed Methodology

Figure 1 shows architecture of Proposed Methodology for Object Detection and Classification using Sparsity Regularized Pruning on Low Quality Image/Video.

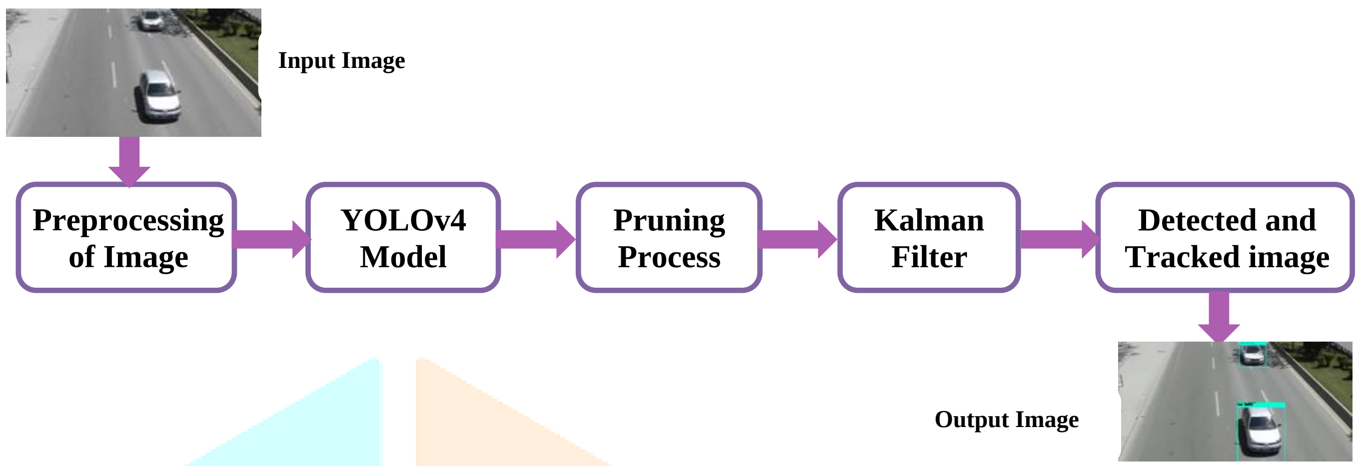


Fig. 1: Proposed Methodology for Object Detection and Classification using Sparsity Regularized Pruning on Low Quality Image/Video

We employed a single stage detector called YOLOv4 for object identification and classification, pruning to reduce the complexity of YOLOv4 (and hence the size of the model), and a Kalman filter to detect and track the object in low-quality video in the suggested methodology. The actions taken to format images before they are utilised in model training and inference are known as image preprocessing. This covers resizing, orienting, and colour corrections, among other things. Processing of image might be as basic as scaling an image. To feed a convolutional network a dataset of images, they must all be the same size. Other processing tasks, such as geometric and colour transformations, colour conversion to grey scale, and so forth, are possible. The goal of pre-processing is to improve the image's quality so that we can better evaluate it. We can reduce unwanted distortions and boost some properties that are important via preparing for the application we are developing. Those functionalities may differ depending on the application. For preparing the image, we employed pixel brightness transformations/brightness adjustments, image filtering, and scaling. The resized image is applied to YOLOv4 model for object detection and classification.

YOLOv4 model:

The architecture of YOLOv4 model is as shown in figure 2.

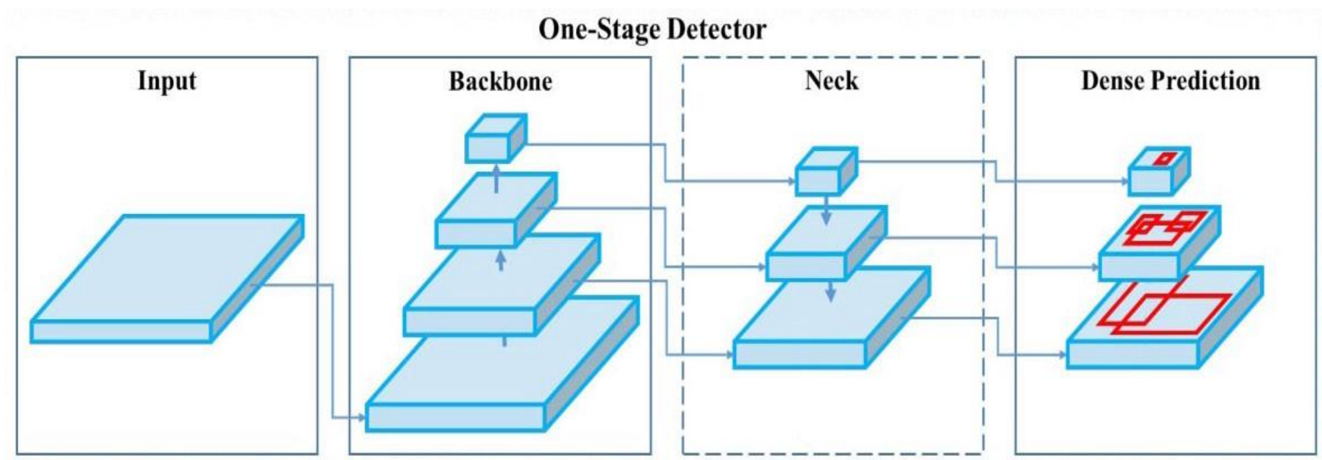


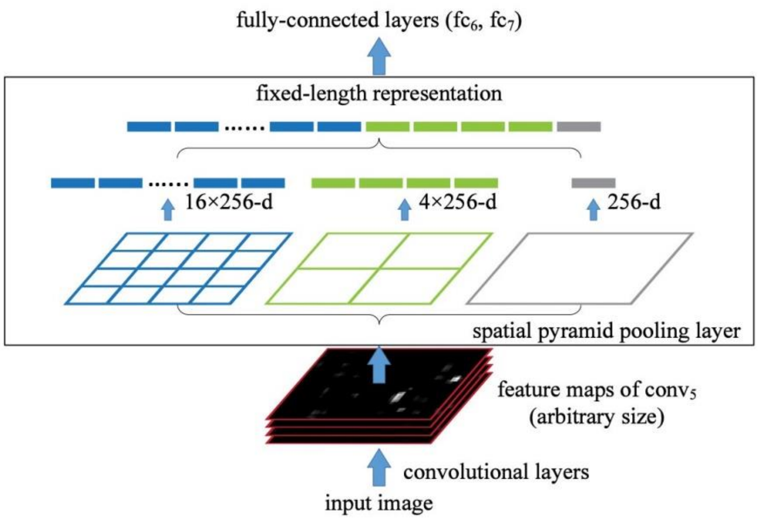
Fig. 2: Architecture of YOLOv4 model [22]

In feature extraction, YOLOv4 uses the CSP connections above and the Darknet-53 below as the backbone. Even while ResNet-based designs offer greater classification performance, the CSPDarknet53 model has higher object detection accuracy. However, using Mish and other techniques, the classification accuracy of CSPDarknet53 can be enhanced. As a result, CSPDarknet53 is the best option for YOLOv4. Object detectors consist of a backbone that extracts features and a head that detects objects. A hierarchy structure is created with the head probing feature maps at multiple spatial resolutions to detect things at different scales. Before feeding into the head, neighboring feature maps from the bottom-up and top-down streams are element-wise combined together or concatenated to enhance the information that feeds into the head. As a result, the head's input will include both spatially rich and semantically rich information from the bottom-up and top-down streams. The neck is the name for this section of the system.

In order to detect objects of various sizes, SPP uses a slightly different method. A spatial pyramid pooling layer takes the place of the last pooling layer (after the last convolutional layer). The feature maps are separated into m x m bins, with m equaling 1, 2, or 4 in this case. After that, each bin for each channel is given a maximum pool. These results in a fixed-length representation that can be studied further using FC-layers. Figure 3 illustrates YOLOv4 with SPP. Many CNN-based models include FC-layers and, as a result, only accept input images with particular dimensions. SPP, on the other hand, accepts photos of various sizes. However, there exist methods such as fully convolutional networks (FCN) that do not use FC-layers and can handle images of various dimensions. This architecture is very effective for image segmentation where spatial information is crucial. As a result, converting 2-D feature maps into a fixed-size 1-D vector is not always desired for YOLO.

YOLO with SPP:

The SPP is changed in YOLO to keep the output spatial dimension. A sliding kernel of size 1x1, 5x5, 9x9, or 13x13 is given a maximum pool. There is no loss of spatial dimension. The output is then concatenated from the features maps from different kernel



sizes.

Fig, 3: YOLOv4 with SPP Layer [22].

Regularization and Pruning Process:

Regularization: One of its most typical issues that data researchers face is avoiding overfitting. Have you ever had a model that performed brilliantly on train data but failed to predict test data? Regularization is an approach for increasing a learning system's generalization by making modest adjustments. The model's performance on unexpected data also improves as a result. The phrase "regularizes" is the activity of bringing something into a regular or acceptable state. This is exactly how we use it. Regularizations are methods for decreasing error and preventing overfitting by fitting a function to the supplied training set correctly. Regularization is a method of fine-tuning by adding a new function term to the function that produces errors. The term that was added prevent the coefficients from reaching extreme values by managing the excessively fluctuating function. Methods of decreased weight decline are terms for a way of controlling or lowering the error coefficients' value in neural networks. The term "regularization" refers to a process for reducing errors and preventing overfitting by ensuring the correct fit function to the provided training set.

Pruning: The unbiased or weightless layers, which represent very little or null value in the memory, are deleted in this process. Pruning is the removal of weight connections from a network to speed up inference and save model storage space. In general, neural networks have too many parameters. Pruning a network involves removing unnecessary parameters from an over-parameterized system. Weights pruning: Pruning by weights, also referred as model pruning, is a set of techniques for increasing the sparsity of a network's weights (the no. of zero-valued components in a tensor). The term 'parameters' refers to a model's weights and bias tensors in general. Biases are rarely, if ever, pruned because there are far fewer bias elements than weights elements, and the effort is just not worth it. Pruning necessitates pruning techniques criteria to determine which pieces to prune. The absolute value of each element is the more widely used factor for pruning: the element's absolute value is compared to a threshold value, and if it is less than the threshold, the element is set to zero (i.e. pruned).

We focused on the following advancements:

1. Weight Pruning: The neural network that produces a smaller model is pruned for a better result in the weights values, which benefits in achieving a good accuracy and providing a dynamically network for architecture development. This is in reference to the features maps' sparsity aspect and the model's neuron pruning's, which aren't active and aren't employed very often. 2. Filter Pruning: This is done to lower the amount of memory space required for storage while enhancing model inference. To approximate the changes in the cost function produced by pruning, a Taylor expansion-based pruning criteria is created employing independent pruning filtering method, and a Taylor expansion-based pruning criterion is developed. It guides the significance of evaluation of filters using statistics from the preceding layer. Pruning low magnitude weight filters is comparable to pruning low weight sum filters. When all of a filter's kernel weights fall below a given threshold, magnitude-based weight pruning, it is possible to get eliminate of a whole thing. It does, however, demand precise threshold tuning, and forecast the exact number of filters that will be pruned in the end. Furthermore, it generates sparse convolutional kernels, which may be difficult to speed up due to a lack of fast sparse libraries, particularly in low-sparsity scenarios.

In the last stage, a Kalman filter is used to estimate the object's next state, using the head angle for direction. From the estimated values, a cost function is minimized, allowing the creation of trajectories for each object. It's possible that there are weaknesses in the assessment of the object's state, causing the trajectory to be lost; as a result, the interactions of the trajectories are modelled to communicate the disparate paths, resulting in a correct object path, that is, the reconnection of the missing path with next began path, evaluated by the smallest distance.

The Kalman filter is a mathematical procedure that provides a computationally efficient (recursive) means of estimating the state of a process in several ways: It can estimate past, present, and even future states, even if the particular features of the represented system are unknown. The Kalman filter employs feedback control to estimate a process. The filter evaluates the condition of the process at a given point in time and subsequently receives noisy measurements as feedback.Kalman filter equations are split into two groups categories: time update equations and measurement update equations. The time update formulas are in responsible of calculating the a priori estimation for next time step by projecting the present condition and error covariance estimations ahead (in time). The feedback is controlled by the measurement update equations. This is used to improve an a posteriori estimate by incorporating a fresh measurement into the a priori estimate. Predictor equations are sometimes referred to as time update equations, whereas corrector equations are referred to as measurement update equations. The final estimation technique, as shown in Figure 4, is similar to a predictor-corrector a method for addressing numerical problems. The time update extrapolates the present condition estimation into the future. The measurement update corrects the expected estimate based on current measurements [21].

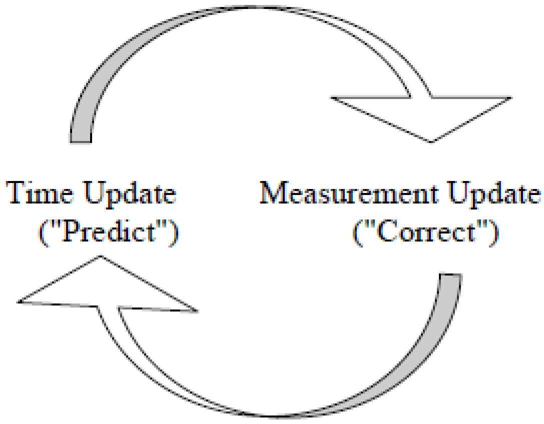


Fig. 4: Discrete Kalman filter cycle [21]

IV. Experimentation

For experimentation and implementation of proposed technique we used the hardware and software for training and testing the model is as i7-7700k CPU and Nvidia 2060 GPU,Ram – 32 GB,HDD – 2 TB ,Windows 11 CUDA 11.2,cuDNN v8.4.5,TensorRT-7.0.1.5,Tensorflow-GPU 2.4.1.

We used MSCOCO dataset for training the model.

Dataset MSCOCO:

COCO is a dataset that includes object detection, segmentation, and captioning on a huge scale.

COCO dataset features:

- 330K photos (>200K tagged)
- 1.5 million object instances
- 80 object categories
- 91 stuff categories
- 5 captions per image
- 250,000 individuals with keypoints
- Object segmentation
- Recognition in context
- Super pixel stuff segmentation

Figure 5 shows sample images of MSCOCO dataset.

Fig. 5: MSCOCO dataset sample images

V. RESULTS AND DISCUSSION

Figure 6 shows a plot of total Loss vs. Epochs during Training stage

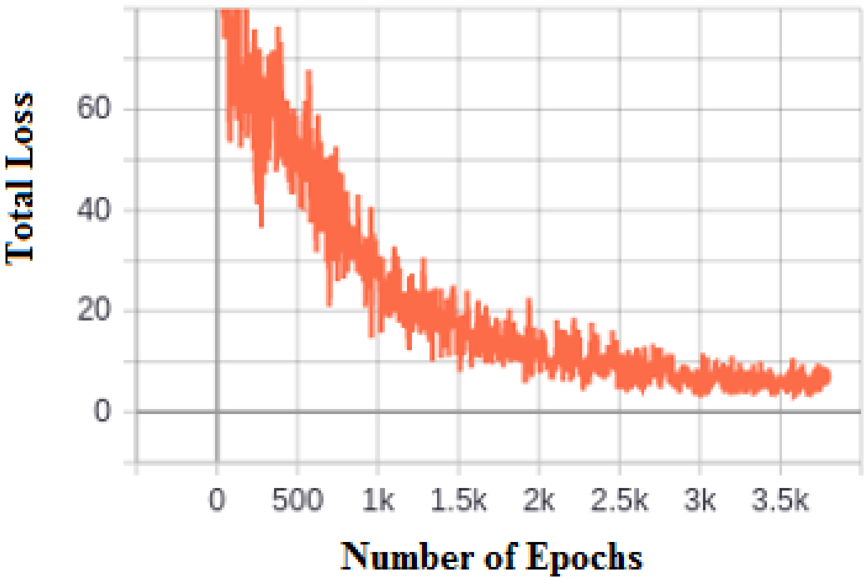


Fig 6: Plot of Loss vs Epochs during Training stage

Table 1: Comparisons of various detection techniques with respect to mAP and FPS

Methods	Datasets	mAP50	mAP95	FPS
Resnet150	MS COCO	76.2%	62.8%	15
Yolov3	MS COCO	82.6%	58.2%	34
Yolov4	MS COCO	86.23%	69.87%	42
Yolov4 Our model	MS COCO	88.72%	70.43%	45

Table 1 shows comparisons of various detection techniques with respect to mAP and FPS

VI. CONCLUSION

The task of recognition and tracking of object is crucial in the computer vision applications. The complexities of object tracking is caused dueto image noise, lighting variations, complicated object motion, and partial and complete occlusion of objects. i.e. low quality video. We proposed a framework consisting of YOLOv4, pruning and Kalman filtering for tracking the low quality video. Due to the pruning of YOLOv4 (single stage detector), which reduces the complexity of model and gives fast processing of image. The MSCOCO dataset is used to train and evaluate the proposed model.For experimentation and implementation of proposed technique we used Google colab and used free GPU, the hardware and software for training and testing the model is as i7-7700k CPU and Nvidia 2060 GPU,RAM, 32 GB,HDD 2 TB ,Windows 11 CUDA 11.2,cuDNN v8.4.5,TensorRT-7.0.1.5,Tensorflow-GPU 2.4.1. The experimental results for our proposed model shows mAP50=88.72, mAP95=70.43and Frames per Second 45 which is more than the state-of-the-art techniques.

REFERENCES

[1] S. K. Pal, A. Pramanik, J. Maiti, and P. Mitra, “Deep learning in multi-object detection and tracking: state of the art,” Applied Intelligence, vol. 51, no. 9, pp. 6400–6429, Apr. 2021.

[2] J.-H. Luo, H. Zhang, H.-Y. Zhou, C.-W. Xie, J. Wu, and W. Lin, “ThiNet: Pruning CNN Filters for a Thinner Net,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 10, pp. 2525–2538, Oct. 2019.

[3] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, “Review of deep learning: concepts, CNN architectures, challenges, applications, future directions,” Journal of Big Data, vol. 8, no. 1, Mar. 2021.

[4] Sheng Ren , Jianqi Li , Tianyi Tu , Yibo Peng , and Jian Jiang” Efficient Video Detection Object Super-Resolution with Deep Fusion Network for Public Safety” Security and Communication Networks, 2021, pp.1–14.

[5] T. Ahmad, Y. Ma, M. Yahya, B. Ahmad, S. Nazir, and A. ul Haq, “Object Detection through Modified YOLO Neural Network,” Scientific Programming, vol. 2020, pp. 1–10, Jun. 2020.

[6] A. Ali and K. Terada, “A framework for Human tracking using Kalman filter and fast mean shift algorithms,” 2009

IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops, Sep. 2009.

[7] J. S. H. and L. S., “Detection and tracking of moving object using modified background subtraction and Kalman filter,” International Journal of Electrical and Computer Engineering (IJECE), vol. 11, no. 1, p. 217, Feb. 2021.

[8] Sanjivani Shantaiya, Kesari Verma, and Kamal Mehta, “Multiple Object Tracking using Kalman Filter and Optical Flow”, European Journal of Advances in Engineering and Technology, 2015, p. 34-39

[9] E. Noureldaim, M. Jedra, and N. Zahid, “Multiple Tracking of Moving Objects with Kalman Filtering and PCA-GMM Method,” Intelligent Information Management, vol. 05, no. 02, pp. 42–47, 2013.

[10] J. Seo and H. Park, “Object Recognition in Very Low Resolution Images Using Deep Collaborative Learning,” IEEE Access, vol. 7, pp. 134071–134082, 2019.

[11] C. I. Patel and R. Patel, “Gaussian mixture model based moving object detection from video sequence,” Proceedings of the International Conference & Workshop on Emerging Trends in Technology - ICWET ’11, 2011.

[12] Y. Tang, Y. Wang, Y. Xu, Y. Deng, C. Xu, D. Tao, and C. Xu, “Manifold Regularized Dynamic Network Pruning,” 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2021.

[13] F. Porikli and A. Yilmaz, “Object Detection and Tracking,” Video Analytics for Business Intelligence, pp. 3–41, 2012.

[14] K. Saho, “Kalman Filter for Moving Object Tracking: Performance Analysis and Filter Design,” Kalman Filters - Theory for Advanced Applications, Feb. 2018.

[15] C. Stauffer and W. E. L. Grimson, “Adaptive background mixture models for real-time tracking,” Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149).

[16] M. Attri, R. Tanwar, Narender, and N. Nandal, “Evaluation of an Efficient Method for Object Detection in Video,” Communications in Computer and Information Science, pp. 442–451, 2020.

[17] D. Hari Hara Santosh, P. Venkatesh, P. Poornesh, L. Narayana Rao, N. Arun Kumar, “Tracking Multiple Moving Objects Using Gaussian Mixture Model”, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-3, Issue-2, pp. 114-119,May 2013

[18] Hitesh A Patel, Darshak G Thakore, “Moving Object Tracking Using Kalman Filter”, International Journal of Computer Science and Mobile Computing, Vol. 2, Issue. 4, April 2013, pg.326 – 332.

[19] R. J. Cardenas T., C. A. B. Castañón, and J. C. G. Cáceres, “Face Detection on real Low Resolution Surveillance Videos,” Proceedings of the 2nd International Conference on Compute and Data Analysis - ICCDA 2018, 2018.

[20] Zuojin Li, Liukui Chen, Jun Peng and Lei Song, “An Approach to Interesting Objects Detection in Low Quality Image Sequences for Fisheries Management”, Wseas Transactions on Signal Processing, Volume 11, 2015, pp. 1-8.

[21] Welch, G. and Bishop, G.: An introduction to the Kalmann Filter, In University of North Carolina at Chapel Hill, Department of Computer Science. Tech. Rep. 95-041, available at http://russelldavidson.arts.mcgill.ca/e761/kalman-intro.pdf,2004.

[22] https://becominghuman.ai/explaining-yolov4-a-one-stage-detector-cdac0826cbd7