Study of Various Types of ML Algorithms

Prof.SayaliS. Kuchekar,DhanashriL.Pawar,ShreejeetV.Adsul
Department of E&TC, Sinhgad College of Engineering
Department of E&TC ,Sinhgad College of Engineering
Department of E&TC ,Sinhgad College of Engineering

---***---

Abstract - This paper introduces about various algorithms used in image processing, The field of image detection has advanced significantly with the development of several powerful algorithms designed to identify and localize objects in images. This study investigates three prominent algorithms in image detection: You Only Look Once (YOLO), Region-based Convolutional Neural Networks (RCNN), and Single ShotMultibox Detector(SSD).Eachofthesealgorithms employs distinct methodologies for achieving object detection, and their performance varies depending on factors such as speed, accuracy, and computational complexity. YOLO is renowned for its real-time detectioncapabilities,making itsuitablefor applications where speed is crucial. RCNN, on the other hand, offers high accuracy by employing region proposals and deep learning for classification, though it is computationally intensive. SSD strikes a balance between speed and accuracy by utilizing a single network to predict object locationsandclassscoresdirectly.Thisstudyprovidesan in-depth comparison of these algorithms based on key performance metrics, such as precision, recall, and inference time.

Key Words :computer vision, image processing, object detection, CNN, Accuracy.

1. INTRODUCTION

Image processing is a crucial field in machine learning, especially in tasks such as object detection, image classification, segmentation, and enhancement. The abilitytoextractmeaningfulinformationfromimageshas made image processing algorithms essential for a wide range of applications, from medical image analysis to autonomous driving. In recent years, machine learning (ML)modelshavesignificantlyadvancedthecapabilities of image processing. This paper explores the key algorithms used in image processing and how they integrate with machine learning models for enhanced performance.

Neural Networks (CNNs) for object detection: Convolutional Neural Networks (CNNs) are the cornerstone of modern image processing in machine learning. CNNs consist of multiple layers, such as convolutionallayers,poolinglayers,andfullyconnected layers.Theconvolutionallayersapplyfilterstotheinput image, extracting essential features such as edges, textures, and patterns. Pooling layers reduce the spatial dimensions,andfullyconnectedlayerscombine features for classification or regression tasks. CNNs are particularly well-suited for tasks like image classification, object detection, and semantic segmentation due to their ability to learn hierarchical features automatically from raw image data.

- ThefigureofCNN'Snetworksreferredfrom

J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, "A comprehensive review of artificial intelligence and machine learning for cybersecurity," *Journal of Big Data*, vol. 8, no. 1, Art. no. 53, 2021. [Online].Available:https://doi.org/10.1186/s40537-021-00434-7.

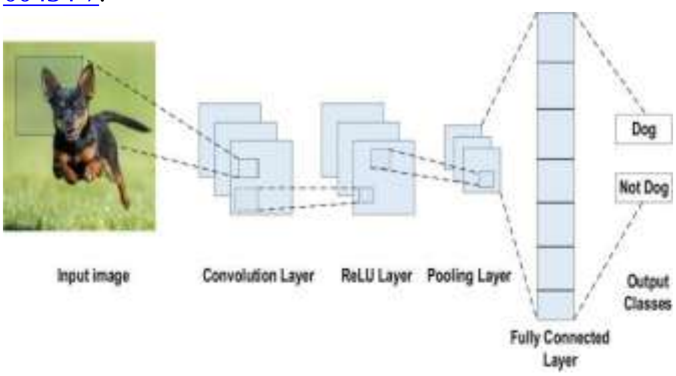


Figure1.DifferentlayersofCNN'S

- **ConceptsofConvolutionalNeuralNetwork:**

Convolutional Neural Network (CNN), also called ConvNet,isatypeofArtificialNeuralNet-work(ANN), which has deep feed-forward architecture and has amazing generalizing ability as compared to other networks with FC layers, it can learn highly abstracted featuresofobjectsespeciallyspatialdataandcanidentify them more efficiently. A deep CNN model consists of a finite set of processing layers that can learn various features of input data (e.g. image) with multiple level of abstraction. The initiatory layers learn and extract the high-level features (with lower abstraction), and the deeper layers learns and ex- tracts the low-level features (withhigherabstraction). The basicconceptualmodelof CNN was shown in figure 1.

2. ML ALGORITHM

1. YouOnlyLookOnce(YOLO)algorithm:

Generally,apersonusuallystaresatanimageandwillget to know what is there in the image and how it interacts with each other. Real-time object detection refers to the ability to detect and localize objects in a continuous stream of data. The background of real-time object detectionstemsfromtheincreasingneedforefficientand accurate analysis of visual data in real-worldscenarios. object detection plays the same role in detecting and differentiating images by neural networks.

Limitations:

1. Highcomputationaldemands
2. Poorperformanceinocclusionandclutter
3. Sensitivitytolightingandweatherchanges
4. Difficultydetectingsmallobjects

The following sample figure of yolo detection and differentversionsofyoloreferredfrom“G.Lavanyaand S. D. Pande, "Enhancing real-time object detection with YOLOalgorithm,"*EAIEndorsedTrans.IoT*,vol.10,no. 1, 2023.[Online].Available:https://doi.org/10.4108/eetiot.4541.

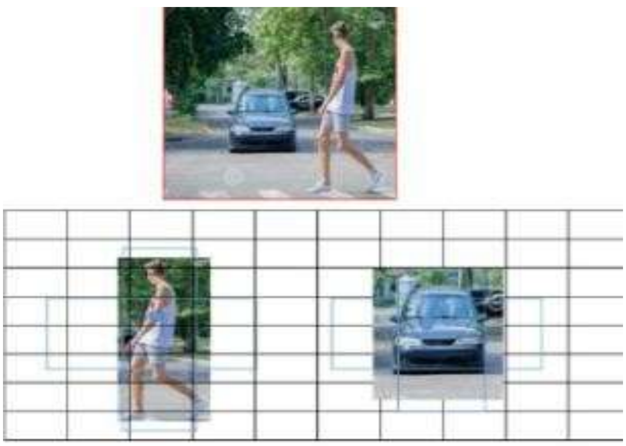


Figure2.SampleofobjectdetectionbyYOLO

DifferentversionsofYOLO:

YOLOv1:Introducedreal-timeobjectdetectionbut struggled with accuracy and small objects.

YOLOv2:Improvedspeedandaccuracyusingbatch normalization, anchor boxes, and multi-scale training.

YOLOv3:Enhancedobjectdetectionwithastronger backbone (Darknet-53) and multi-scale predictions.

YOLOv4: Optimized for performance with better speed and accuracy, especially on GPUs.

YOLOv5:Apopularversionfocusedoneaseofuse,with frequent updates and improved flexibility.

YOLOv6&YOLOv7:Furtheroptimizedforreal-time detection and edge device deployment.

YOLOv8:Thelatest,balancinghighperformanceand real-time processing for resource-limited device.

2. R-CNN(Regions with Convolutional Neural Networks) is an object detection method that combines region proposals with CNNs to identify and classify objects in images. It works by generating region proposals, extracting features using a pre-trained CNN, classifying each region with an SVM, and refining the bounding box with regression. Though accurate, it is computationally expensive. Variants like Fast R-CNN and Faster R-CNN improve speed and efficiency by optimizing the process and integrating region proposal generation within the network-CNN and Fast(er) R-CNN:Use a two-step approach: first, identify regions where objects are expected, and then detect objects in those regions. Region proposal algorithms usually have slightly better accuracy but are slower to run.

Limitations:

1. Slowtrainingandinferencetime
2. Highcomputationalcost
3. Requires external region proposal methods (Selective Search)
4. Notsuitableforreal-timeapplications
5. Highmemoryusageduetofeaturestorage

- **FasterR-CNN:**

AimstobefasterandmoreaccuratethanpreviousR- CNN versions.
Uses a region proposal network (RPN) to create bounding boxes and utilizes those boxes to classify objects. The RPN shares full-image convolutionalfeatureswiththedetectionnetwork, enabling nearly cost-free region proposals.

Region Proposal Network (RPN): The RPN is the core innovationinFasterR-CNN.Itreplacestheslowexternal region proposal methods used in earlier models like Selective Search. The RPN is a fully convolutional network that scans the feature map of the image and generatescandidateobjectproposals.Itdoessobysliding asmallwindowovertheimage,predictingthelikelihood ofanobjectbeingpresentateachlocation,andproposing bounding boxes (anchors) around potential objects.

Fast R-CNN Detection Network: After the RPN generates the object proposals, the Fast R-CNN network isusedtoclassifyandrefinetheseproposals.Itperforms:

Classification: Each proposal is assigned a class label (e.g.,car,pedestrian).
Bounding Box Regression: The coordinates of the bounding boxes are refined for more accurate object localization.

Limitations:

1. Stillslowerthanone-stagedetectors(e.g.,YOLO)

2. Complextrainingpipeline

3. Performancedropsonsmallobjects

4. Sensitivetohyperparametertuning

5. Notidealforedgeormobiledevices

HowItWorks

Step 1: The input image is passed through a ConvolutionalNeural Network(CNN)togeneratefeature maps.

Step 2: The RPN takes the feature maps from the CNN and generatesa setofobject proposals (potentialregions where objects may exist).

Step 3: The proposals are sent to the Fast R-CNN detection network for object classification and bounding box refinement.

Step 4: The network is trained end-to-end, optimizing both the RPN and detection network simultaneously.

The sample object detection figure of RCNN referred from ‘S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN:TowardsReal-TimeObjectDetectionwithRegion Proposal Networks," in *Advances in Neural Information Processing Systems*, vol. 28, 2015.’

Figure3.SampleofobjectdetectionbyRCNN

3. SSD(SingleShotMultiBoxDetector):

SingleShotMultiBoxDetectorBasedonadeep convolutional neural network (CNN) that predicts bounding boxes and class probabilities in a single pass. SSD Designed for real-time applications, like video surveillanceandautonomousdriving.ItUsesmulti-scale featuresanddefaultboxestoimproveaccuracy.SSD300 achieves74.3%mAPat59FPS,whileSSD500achieves 76.9%mAPat22FPS.SSDismuchfasterthantwo-shot RPN-based approaches like R-CNN.

Limitations:

1. Loweraccuracyonsmallobjects

2. Fixed-sizedefaultboxesmaymissvariedobjectshapes

3.Struggles with crowded scenes and occlusion

4.Sensitivity to object scale and aspect ratio

ThesamplefigureisreferredfromW.Liu,D.Anguelov, D.Erhan,C.Szegedy,S.Reed,C.-Y.Fu,andA.C.Berg, "SSD: Single Shot MultiBox Detector," in *Proc. EuropeanConf.ComputerVision(ECCV)*,2016,pp.21– 37.

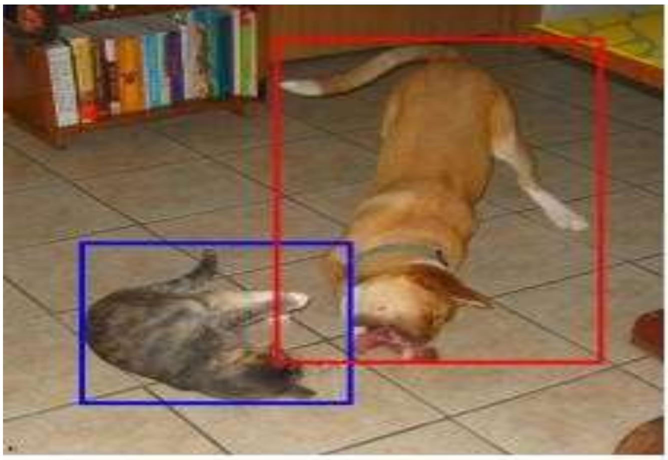


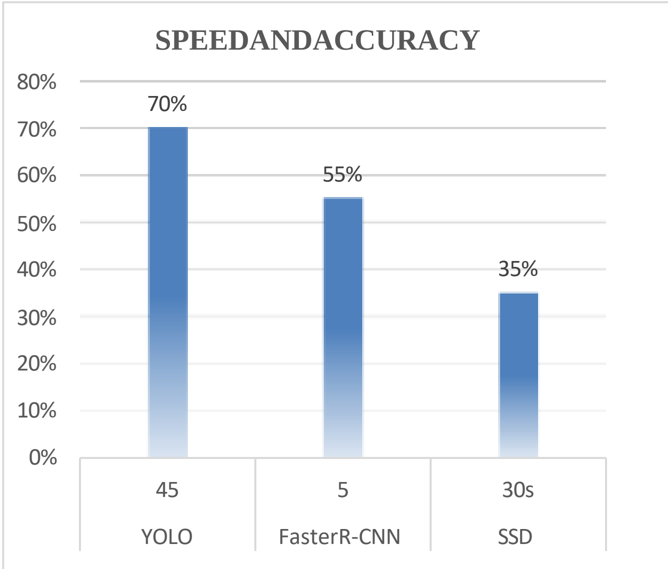
Figure4.SampleofObjectdetectionbySSD

SSDisdesignedforobjectdetectioninreal-time. Faster R-CNN uses a region proposal network to create boundary boxes and utilizes those boxes to classify objects. While it is considered the start-of-the-art in accuracy,thewholeprocessrunsat7framespersecond. Far below what real-time processing needs. SSD speeds up the process by eliminating the need for the region proposal network. To recover the drop in accuracy, SSD applies a few improvements including multi-scale features and default boxes. These improvements allow SSD to match the Faster R-CNN's accuracy usinglower resolutionimages,whichfurtherpushesthespeedhigher. According to the following comparison, it achieves the real-timeprocessingspeedandevenbeatstheaccuracyof the Faster R-CNN. (Accuracy is measured as the mean averageprecisionmAP:theprecisionofthepredictions.)

3. COMPARISONVARIOUSALGORITHM

These are three popular deep learning algorithms for object detection, each offering a different balance between **speed** and **accuracy**. Here's a detailed comparison of these three models based on speed and accuracy:

Sr.No	Algorithms	Speed(Fps)	Accuracy
1	YOLO	45-100fps	~60-70%mAP
2	FasterR-CNN	5-10fps	~50-55%mAP
3	SSD	30-50fps	~30-35%mAP



Followingcomparisonwereferredfromvariousresearch papers and study about their speed and accuracy

KeyObservations:

- YOLOv5isoneofthehighest-performingversionsin terms of accuracy, reaching up to 70% mAP.
- Faster R-CNN offers the best accuracy among these algorithms(especiallywhenusingenhancementslike Feature Pyramid Networks).
- SSD provides lower accuracy but has a faster speed and is often used for real-time applications

4. CONCLUSIONS

Inconclusion,ourimplementationoftheYOLOv8 algorithm for object detection has proven to be highly effective compared to the traditional RCNN approach. YOLOv8 demonstrated superior speed and efficiency, processing images in real time while maintaining high accuracy in object localization and classification. Unlike RCNN, which requires time-consuming region proposal networks and separate stages for classification and bounding box regression, YOLOv8 operates in a single, unified pipeline, significantly reducing computational overhead.ThismakesYOLOv8amorepracticalsolution for real-time applications where speed and performance are critical. Overall, the adoption of YOLOv8 offers a clear advantage in both performance and scalability, making it a preferred choice over RCNN for modern object detection tasks.

REFERENCES

[1] R. Girshick, J. Donahue, T. Darrell and J. Malik, "RichFeatureHierarchiesforAccurateObjectDetection and Semantic Segmentation," *2014IEEE Conference on Computer Vision and Pattern Recognition, Columbus*, OH, pp. 580-587, 2014.

[2] R.Girshick,"FastR-CNN,*"2015IEEE InternationalConferenceonComputerVision(ICCV),* Santiago,pp.1440-1448,2015.

[3] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN:TowardsReal-TimeObjectDetectionwithRegion Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137-1149, 2017.

[4] M.A.E.Muhammed,A.A.AhmedandT.A.Khalid, "Benchmark analysis of popular ImageNet classification deepCNNarchitectures,*"2017InternationalConference On Smart Technologies For Smart Nation (SmartTechCon),* Bangalore, pp. 902-907, 2017.

[5] G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger, "Densely Connected Convolutional Networks," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, pp. 2261-2269, 2017.

[6] X. Wang, A. Shrivastava and A. Gupta, "A-Fast-R-CNN:HardPositiveGenerationviaAdversaryforObject Detection," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, pp. 3039-3048, 2017.

[7] **Article:** Shafiee, Mohammad Javad, et al. "Fast YOLO: A fast you only look once system for real-time embedded object detection in arXiv:1709.05943 (2017).

[8] Anzai Y. Pattern recognition and machine learning. Elsevier; 2012.

[9] K.E.A.vandeSande,J.R.R.Uijlings,T.Geversand A. W. M. Smeulders, "Segmentation as selective search for object recognition," *2011 International Conference on Computer Vision, Barcelona*, pp. 1879-1886, 2011.

[10] F. Sultana, A. Sufian and P. Dutta, "A Review of ObjectDetectionModelsBasedonConvolutionalNeural Network", *Intelligent Computing: Image Processing Based Applications*, pp. 1–16, 2020.

[11] G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger, "Densely Connected Convolutional Networks,"*2017IEEEConferenceonComputerVision andPatternRecognition(CVPR)*,Honolulu,HI,pp. 2261-2269, 2017.

[12] X.Wang,A.ShrivastavaandA.Gupta,"A-Fast-R-CNN:HardPositiveGenerationviaAdversaryforObject Detection," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, pp. 3039-3048, 2017.

[13] J.Redmon,S.Divvala,R.GirshickandA.Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, pp. 779-788, 2016.

[14] B. Liu, W. Zhao and Q. Sun, "Study of object detection based on Faster R-CNN," *2017 Chinese AutomationCongress(CAC)*,Jinanpp.6233-6236,2017.

[15] Sumit1 , Shrishti Bisht2 , Sunita Joshi3 , Urvi Rana4" Comprehensive Review of R-CNN and its Variant Architectures," *April 2024 International Research Journal on Advanced Engineering Hub (IRJAEH)*, e ISSN: 2584-2137,2017.

[16] J. Kelly, Tuning digital pi controllers for minimal variance in manipulated input 329 moves applied to imbalanced systems with delay, Can. J. Chem. Eng. 76 (1998) 330 967–974. 331

[17] S.Narasimhan,C.Jordache,DataReconciliationand GrossErrorDetection:*An332IntelligentUseofProcess Data,GulfPublishingCompany*,Houston,TX,2000.333

[18] J. Hedengren, T. Edgar, Approximate nonlinear model predictive control with in 336 situ adaptive tabulation, *Computers and Chemical Engineering 32 (2008)* 706– 337 714. 1