Cheng-Yang Fu et al., [1], Here, the paper discusses state-of-the-art general object detection. By introducing additional large-scale context in object detection and improving accuracy, especially for small objects, they introduce their resulting system DSSD for deconvolutional single shot detector. Results are discussed by both PASCAL VOC and COCO detection. Their DSSD has 513x513 input which achieves 81.5% mAP on VOC2007 test, 80.0% mAP on VOC2012 test, and 33.2% mAP on COCO which surpasses the state-of-the-art method R-FCN on each dataset. Also, they expect many improvements in finding more efficient ways to combine the features of encoder and decoder.

Golnaz Ghiasi et al., [2], The authors discuss the architecture for object detection that is manually designed. A better architecture of feature pyramid for object detection. The NAS-FPN has a merger of top-down and bottom-up connections. It improves mobile detection accuracy by 2 AP compared to the other detector with less detection time.They have suggested using Neural Architecture search and have optimized the process of designing feature pyramid networks for object detection. On the COCO dataset, they ran trials and found that their architecture, known as NAS-FPN, is adaptable and useful for creating precise detection models.

Hei Law and Jia Deng., [3], Here, a new approach named ‘CornerNet’ has been proposed. We also see the use of Single convolutional neural network. Through this they have successfully terminated the need of designing anchor boxes which were used in preceding single-stage detectors. Along with this they have also introduced ‘corner pooling’. Here we can observe that the CornerNet accomplishes a 42.1% AP on MS COCO dataset.

Robert J. Wang et al., [4], They discuss the increasing need of Convolutional Neural Network (CNN) models on Mobile devices which have finite computing power and memory resources. A more efficient architecture named ‘PeleeNet’ has been proposed. Also, they say that the model size of Pelee Net is just 66% that of MobileNet. They also made use of the single shot multibox detector with PeeleNet and enhanced the architecture for fast speed. Also achieving 23.6 FPS on iPhone, 8 and 125 FPS on NVIDIA TX2. They are able to do real-time prediction for image classification and object recognition tasks on mobile devices by fusing efficient architecture design with mobile GPU and hardware-specified optimized runtime libraries.

Zhuang Liu et al., [5], They have extended their work on Deeply supervised object detectors (DSOD), this structure/architecture can learn object detectors from nothing. They discuss two critical problems and their solution which is DSOD. They also say that previous efforts mostly failed due to complex loss functions and finite datasets.Additional research on PASCAL VOC 2007, 2012, and MS COCO demonstrates that DSOD outperforms cutting-edge methods. DSOD offers enormous promise in a variety of scenarios, including those involving depth, medicine, multi-spectral imaging, etc.

Wei Liu et al., [6], They have proposed object detection with a single deep neural network. The network is also capable of generating scores for the categories and also constructs adjustments to the box to better match the object shape.The SSD technique's ability to do away with a second proposal generation stage and subsequent resampling stages makes it a more streamlined and effective method, which is one of its main advantages. The SSD offers a unified framework for both training and inference and achieves accuracy that is comparable to approaches that include an additional object proposal step.

Mingxing Tan et al., [7], The authors provide a compound scaling technique that simultaneously scales all backbone, feature network, and box/class prediction networks' resolution, depth, and width. This method significantly outperforms prior art in terms of efficiency and enables the efficient and effective utilization of resources. Overall, this research offers insightful advice and useful tips for enhancing the effectiveness of object detection systems based on neural networks. The proposed optimisations and the EfficientDet family of object detectors have the potential to be used in a variety of fields, such as robotics, surveillance, and autonomous driving.

Zhishuai Zhang et al., [8], The research suggests Detection with Enriched Semantics (DES), a unique single-shot object detection network. With the help of a semantic segmentation branch and a global activation module, this approach enriches the semantics of object detection characteristics within a standard deep detector. There is no need for additional annotation because the segmentation branch is supervised by weak segmentation ground-truth. Additionally, the global activation module develops a self-supervised understanding of the connection between channels and object classes. They have achieved tremendous results in their experiments on both PASCAL VOC and MS COCO datasets.

Shifeng Zhang et al., [9], The single-shot detector created by the RefineDet algorithm presented in this research is more accurate than two-stage techniques while still being efficient. The tests performed on the PASCAL VOC 2007, PASCAL VOC 2012, and MS COCO datasets show that RefineDet provides cutting-edge detection accuracy while operating at a high level of efficiency. The findings show that the suggested algorithm maintains the effectiveness of one-stage approaches while offering greater accuracy than two-stage methods.

Tsung-Yi Lin et al., [10], The paper suggests a brand-new one-stage object detector called RetinaNet that maintains the simplicity and speed benefits of one-stage detectors while achieving state-of-the-art accuracy. The Focal Loss is used to train the RetinaNet detector, which outperforms existing state-of-the-art two-stage detectors in accuracy while keeping the speed of earlier one-stage detectors.Because it offers a better balance between accuracy and speed, the RetinaNet detector makes a substantial contribution to the area of object detection and has considerable implications for practical applications.

References:

[1] Cheng-Yang Fu, Wei Liu, Ananth Ranga, Ambrish Tyagi, Alexander C. Berg
UNC Chapel Hill, Amazon Inc., "DSSD: Deconvolutional Single Shot Detector"

[2] Golnaz Ghiasi, Tsung-Yi Lin,Quoc V. Le, Google Brain., "NAS-FPN: Learning Scalable Feature Pyramid Architecture for Object Detection"

[3] Hei Law, Jia Deng., "CornerNet: Detecting Objects as Paired Keypoints"

[4] Robert J. Wang, Xiang Li, Charles X. Ling., "Pelee: A Real-Time Object Detection System on Mobile Devices"

[5] Zhiqiang Shen, Zhuang Liu, Jianguo Li, Yu-Gang Jiang, Yurong Chen, Xiangyang Xue
Fudan University, Tsinghua University, Intel Labs China., "DSOD: Learning Deeply Supervised Object Detectors from Scratch"

[6] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg., "SSD: Single Shot MultiBox Detector"

[7] Mingxing Tan, Ruoming Pang, Quoc V. Le, Google Research, Brain Team., "EfficientDet: Scalable and Efficient Object Detection"

[8] Zhishuai Zhang, Siyuan Qiao, Cihang Xie, Wei Shen, Bo Wang, Alan L. Yuille,
Johns Hopkins University, Shanghai University, Hikvision Research., "Single-Shot Object Detection with Enriched Semantics"

[9] Shifeng Zhang, Longyin Wen, Xiao Bian, Zhen Lei, Stan Z. Li, CBSR & NLPR, Institute of Automation, Chinese Academy of Sciences, Beijing, China. University of Chinese Academy of Sciences, Beijing, China. GE Global Research, Niskayuna, NY. Faculty of Information Technology, Macau University of Science and Technology, Macau, China., "Single-Shot Refinement Neural Network for Object Detection"

[10] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, Piotr Doll´ar, Facebook AI Research (FAIR)., "Focal Loss for Dense Object Detection"