{\n "SummaryDoc": "# Research Advanced in Object Detection based on Deep Learning\n\n## Abstract\nObject detection is a fundamental research area in computer vision, focusing on the accurate prediction of the category and location of objects within a scene. This paper reports recent advancements in object detection using deep learning, presenting the evolution of technology and evaluating various methodologies to highlight their strengths and limitations, ultimately suggesting future research directions.\n\n## Keywords\ncomputer vision, natural language process, Object detection, anchor-free.\n\n## 1. Introduction\nWith the rapid advance of artificial intelligence, autonomous vehicles are becoming integral to future transportation systems. Object detection, a critical component of environmental perception systems in autonomous vehicles, plays a crucial role in ensuring safe and reliable navigation by accurately identifying road elements such as pedestrians, vehicles, and traffic signs.\n\n## 2. Method\n### 2.1. Two-stage Detection Approaches\nTwo-stage methods, such as Faster R-CNN, effectively harness deep learning for object detection by first generating region proposals and then classifying these regions.\n\n### 2.2. Single-stage Detection Approaches\nSingle-stage methods like YOLO bypass the region proposal stage to directly predict object boundaries and classifications from the input image, significantly speeding up the process.\n\n### 2.3. Key point-based Detection Approaches\nKey point-based methods identify crucial points on the object to determine its position and orientation without relying on anchor boxes, thus reducing computational complexity and resource consumption.\n\n### 2.4. Transformer-based Detection Approaches\nTransformers use an attention mechanism to enhance object detection models, processing data in parallel to improve both the speed and accuracy of detection.\n\n## 3. Experiment\n### 3.1. Datasets and Metrics\nPascal VOC and COCO datasets are utilized to evaluate the performance of different object detection methods. Metrics such as average precision (AP) and mean average precision (mAP) are employed to assess accuracy and speed.\n\n### 3.2. Performance Comparison\n\*\*Table 1. Accuracy (mAP) comparison between different representative methods\*\*\n| Method | Category | VOC2007 | COCO |\n|:--------|:---------|-------------:|----:|\n| Fast RCNN | Two-stage method | 70.0 | 19.7 |\n| Faster RCNN | | 73.2 | 21.9 |\n| SSD | Single-stage method | 81.6 | 26.8 |\n| YOLOv4 | | \*\*81.8\*\* | 43.5 |\n| Retina-Net | | 79.5 | 36.2 |\n| RefineDet | | 79.2 | 39.1 |\n| CornerNet | Key point-based method | 79.5 | 42.5 |\n| CenterNet | | 78.0 | 41.8 |\n| FCOS | | 78.8 | 44.7 |\n| Deformable DETR | Transformer-based method | 79.5 | \*\*52.3\*\* |\n| Swin Transformer | | \*\*81.0\*\* | \*\*57.7\*\* |\n\n## 4. Discussion\nDiscusses current challenges and future directions in object detection, highlighting the need for more generalized models capable of operating across various scenes and conditions.\n\n## 5. Conclusion\nSummarizes the state of object detection technology, emphasizing the significant strides made in terms of accuracy and speed thanks to deep learning advancements.\n\n## References\n1. Girshick R. Fast R-CNN. Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV). Santiago: IEEE, 2015. 1440–1448.\n2. Ren SQ, He KM, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks. arXiv: 1506.01497, 2015."\n}