Object Detection in Deep learning

Assignment: Final Project- Deep Learning Models

Student Name: Aksa Taniya

Kent ID: 811244499

Instructor: CJ Wu

Table of Contents

List of figures with page references

1. Introduction

The goal of object detection in computer vision is to find and categorize items inside digital videos or image frames. It is critical in a variety of applications like automatic driving, monitoring systems, robotics, and image interpretation. Deep learning, specifically neural networks using convolution, has transformed the area of object identification, allowing for improved precision and effectiveness in detection systems.

Traditional object detection methods relied on handcrafted features and complex algorithms to identify objects. However, deep learning models have shown remarkable performance by automatically learning discriminative features directly from raw pixel data. This ability to learn hierarchical representations has propelled the advancement of object detection techniques.

Deep learning-based detection of objects usually consists of two parts: area suggestion and object categorization. The region proposal modules create a collection of candidate regions that are likely to include objects, and the classification module identifies the existence and kind of items in these regions.

The region-based convolutional deep neural network (R-CNN) family, including the R-CNN Fast R-CNN, and Faster R-CNN, is one of the most prominent object identification designs. These models function in two stages: first, they create a collection of region suggestions via selective search and region proposal network (RPN). The proposed areas are then categorized and refined with CNNs. One-stage detection of objects approaches includes the Single Shot MultiBox Detector and You Only Look Once.

They accomplish real-time performance by predicting item box boundaries and class likelihoods from a single network run. These kinds of models are popular for their identification accuracy and efficiency. Furthermore, the introduction of algorithms for deep learning like PyTorch, TensorFlow, and Keras has made the creation and implementation of object identification models easier. These frameworks offer pre-trained models, allowing academics and practitioners to take use of cutting-edge architecture and fine-tune them to suit specific applications or datasets.

In recent years, object detection has also witnessed advancements through the integration of other techniques. For instance, feature pyramid networks (FPN) improve the detection of objects at different scales, while attention mechanisms enhance the model's ability to focus on informative regions. Additionally, models like Efficient Diet and Detectron2 have pushed the boundaries of object detection by achieving better accuracy and efficiency through novel architectural designs.

To summarize, deep learning object detection has considerably enhanced the science of computer vision through allowing accurate and efficient recognition of things in pictures and movies. Object detection algorithms continue to improve as a result of the strength of neural networks that are deep and creative architectural designs, leading to an extensive number of practical uses and supporting additional study in the field.

2. Types of object detection methods:

Object detection in deep learning encompasses various methods and architectures that differ in their approach to identifying and localizing objects within images or video frames. Here, we will discuss some of the prominent types of object detection methods in detail:

Two-Stage Methods:

Two-stage methods follow a two-step process: region proposal and object classification. These strategies influenced the advancement of object detecting techniques. Here are a few examples.

A. R-CNN: R-CNN was one of the first deep learning-based object identification algorithms. It generates region suggestions using selective search followed by implements a CNN to extract characteristics from each proposal. These characteristics are then supplied to support vector machine models (SVMs) during categorization. R-CNN uses a two-stage technique that includes region proposal and object categorization.

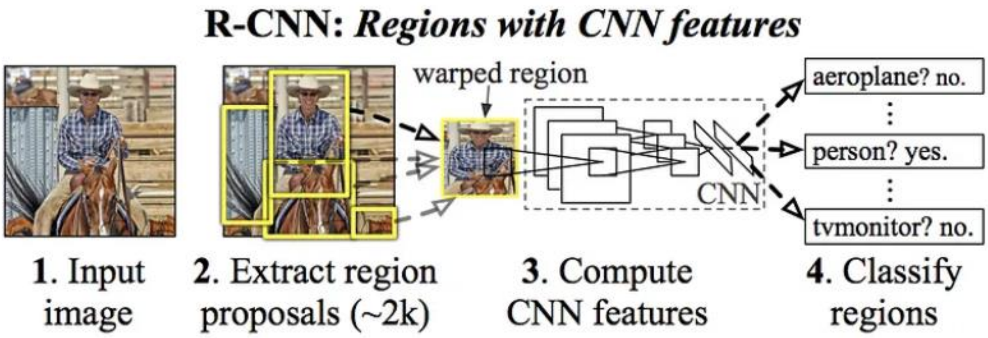

Figure 01: Architecture of RCNN

B. Fast R-CNN: Fast R-CNN enhances efficiency by sharing convolutional features between proposals, building on R-CNN. A region of interest layer of pooling is introduced for extracting fixed-sized features from convolutional feature maps. These characteristics are then applied to categorization and the bounding box regression. Fast R-CNN is an R-CNN extension that tackles some of its shortcomings, such as slow processing time and poor training convergence. Fast R-CNN integrates the region proposal and feature extraction stages into a single step, leading to a faster and more efficient detection pipeline. Here are the main features of Fast R-CNN:

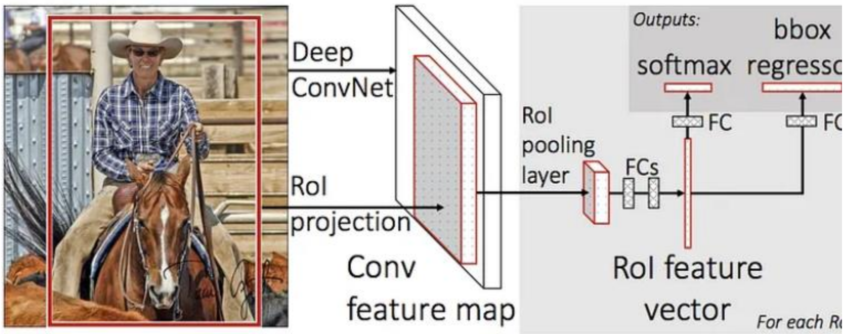

Figure 02: Architecture of Fast RCNN

C. Faster R-CNN: Faster R-CNN improves speed and accuracy even more by using a region-based proposal network (RPN). The RPN shares the detection network's convolutional characteristics and predicts object suggestions directly. The detection network then categorizes and refines these ideas. Faster R-CNN is an expansion of the Fast R-CNN object identification approach that adds a combined region proposal network (RPN) to increase speed and accuracy. Faster R-CNN substitutes the Fast R-CNN's selective search algorithm with an RPN that shares convolution attributes with the detection networks, allowing for end-to-end training and quicker inference. Let's have a look at the primary parts and method of Faster R-CNN.

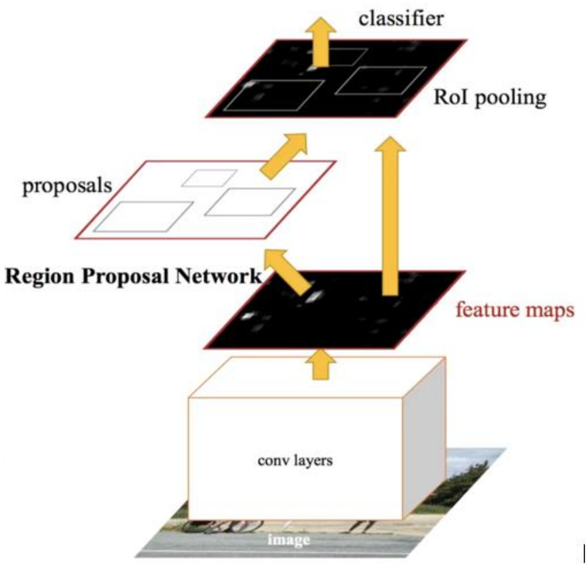

Figure 03: Architecture of Faster RCNN

One-Stage Methods:

One-stage methods aim to detect objects directly without an explicit region proposal step. These methods are known for their real-time performance and simplicity. Here are two popular examples:

A. Single Shot MultiBox Detector (SSD): SSD is a widely used one-stage object detection method. It divides the input image into a grid of cells and predicts multiple bounding boxes and class probabilities at each cell for various object scales and aspect ratios. The predictions are refined using convolutional layers at different scales to detect objects. Single Shot MultiBox Detector (SSD) is a popular object detection method that combines high accuracy with real-time inference speed. SSD addresses the challenge of detecting objects at different scales by utilizing a set of default bounding boxes, known as anchor boxes, at multiple aspect ratios and scales. It performs simultaneous object classification and bounding box regression on these anchor boxes.

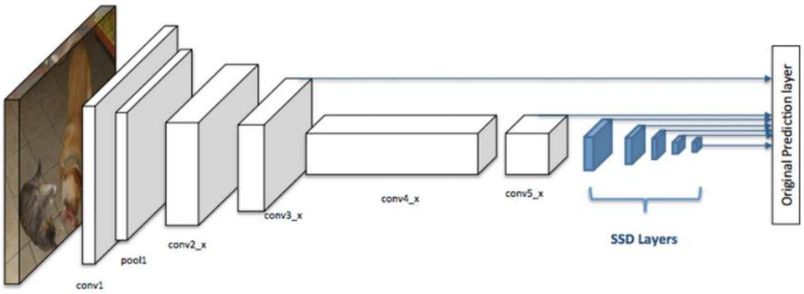

Figure 04: Architecture of SSD

B. You Only Look Once (YOLO): YOLO is another effective one-stage object detection approach. It splits the input picture into grids and forecasts box boundaries and class probabilities using just one neural network. The YOLO delivers real-time performance by doing detection at several scales and employing anchor boxes for precise localization. YOLO is a popular object identification technique noted for its speed and efficiency in real-time. YOLO provides a novel approach to object identification by framing it as a regression issue, allowing it to estimate the coordinates of the bounding box and class probabilities from a single trip through the network.

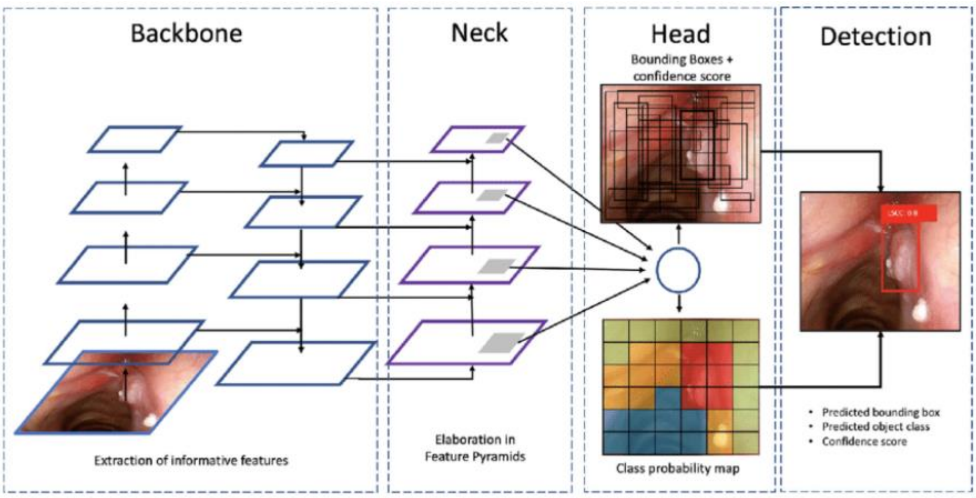

Figure 05: Architecture of YOLO

C. Efficient Object Detection Methods:

Efficient object detection methods focus on **improving** the accuracy and efficiency of detection models. They leverage innovative architectural designs and optimization techniques. Notable examples include:

a. EfficientDet: Efficient Set introduces a compound scaling method to optimize both accuracy and efficiency. It scales the backbone, feature network, and box/class prediction network simultaneously to achieve better trade-offs between these factors. EfficientDet employs an efficient object detection head that predicts class probabilities and bounding box coordinates. The detecting head is made up of convolutional layers, which are connected by anchor boxes at various sizes and aspect ratios. EfficientDet utilizes a modified version of the anchor box assignment mechanism called the Auto Assign algorithm to assign anchor boxes to ground-truth objects efficiently.

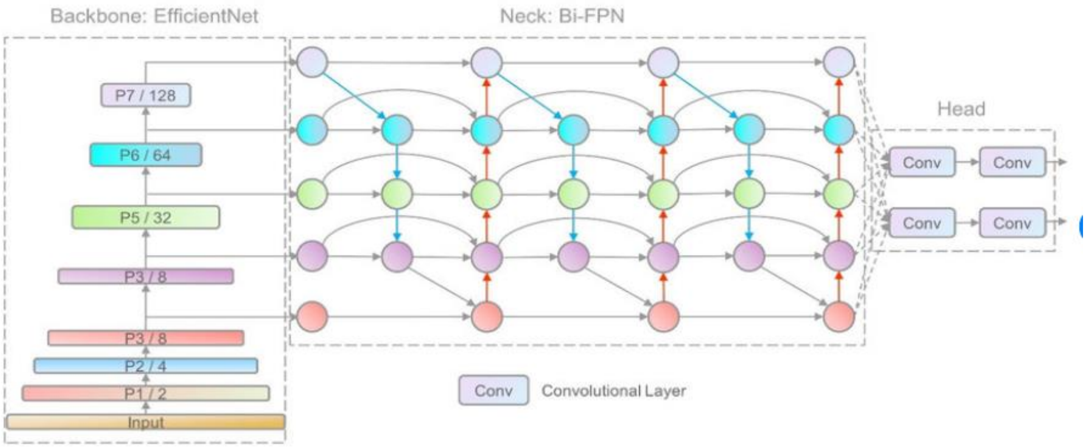

Figure 06: Architecture of EfficientNet

b. Detectron2: Detectron2 is a powerful object detection framework that provides a collection of state-of-the-art models and tools. It combines efficient backbones, feature pyramid networks, and other advanced techniques to achieve high-performance object detection. Facebook AI Research (FAIR) created Detectron2, a powerful and adaptable open-source software framework for developing cutting-edge object identification and segmentation models. It is an enhancement to the fundamental Detectron system giving enhanced speed, versatility, and usability.

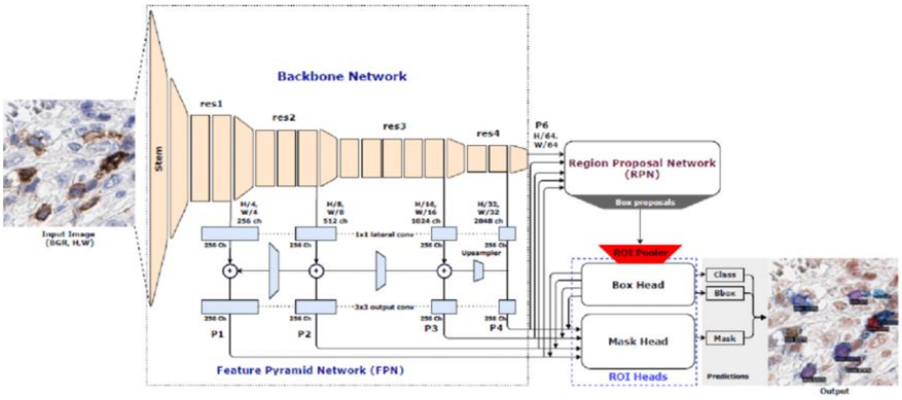

Figure 07: Architecture of Detectron2

3.Literature Review:

Mingxing Tan et.al [1] introduces EfficientDet, a highly efficient and accurate object detection framework that achieves state-of-the-art performance through compound scaling and the BiFPN module. May still be computationally intensive for deployment on resource-constrained devices. Yuxin Wu [2] presents Detectron2, an advanced object detection framework that offers modularity, flexibility, and efficient training and inference pipelines, serving as an upgrade to the original Detectron framework. Requires some familiarity with the underlying architecture and implementation details for efficient usage. Alexey Bochkovskiy [3] introduces YOLOv4, an upgraded version of the popular YOLO (You Only Look Once) object detection algorithm, achieving significant improvements in both accuracy and speed. It requires powerful hardware resources due to high computational requirements, limiting usage in resource-constrained environments. Hei Law [4] proposes CornerNet-Lite, an efficient keypoint-based object detection algorithm that achieves high accuracy while maintaining computational efficiency through keypoint estimation and regression. It mainly focuses on corner keypoints, limiting performance in scenarios where object corners are not clearly defined or visible. Kehan Zhang [5] presents an efficient object detection method for large images using deep reinforcement learning, which reduces the search space for object candidates, leading to improved efficiency. Limited to object detection in large images and may require fine-tuning for different application domains. Glenn Jocher [6] presents YOLOv5, a fast and accurate object detector that outperforms previous versions of YOLO and other state-of-the-art detectors on various benchmark datasets. YOLOv5 achieves this by introducing a new architecture that is based on anchor-free detection and focal loss. The detector is also lightweight, making it easy to deploy on edge devices. Although YOLOv5 achieves state-of-the-art performance, it still struggles with detecting small objects in cluttered scenes. Mingxing Tan [7] presents EfficientDet, a family of object detectors that achieve state-of-the-art performance on various benchmark datasets while being more efficient than previous state-of-the-art detectors. EfficientDet achieves this by introducing a new scaling method that improves the balance between model size and accuracy. The detectors are also efficient and can be deployed on resource-constrained devices. EfficientDet requires a large amount of training data to achieve state-of-the-art performance, which can be a challenge for some applications. Xizhou Zhu [8] presents Deformable DETR, an end-to-end object detection framework that integrates deformable convolutional networks and transformers. Deformable DETR achieves state-of-the-art performance on various benchmark datasets while being efficient and easy to train. The detector is also robust to occlusions and deformations. Deformable DETR requires a large amount of memory to train, which can be a challenge for some applications. Rodrigo Benenson [9] presents an empirical analysis of various object detection architectures and their performance on various benchmark datasets. The authors show that there is no single best architecture and that the performance of different architectures depends on the specific dataset and evaluation metric. The paper also provides recommendations for selecting an architecture based on the specific requirements of an application. The paper does not introduce any new object detection architectures, but rather provides a comprehensive analysis of existing ones.

4. Industrial Application of Deep Learning:

Because of its capacity to recognize and track things in real-time, object detection utilizing deep learning offers an extensive variety of industrial applications. Object detection has present and future uses in areas such as medical care, public transit, and security:

1. Healthcare: Object identification algorithms can detect and monitor numerous abnormalities in medical imaging, such as tumors, fractures, and blood vessel irregularities. It may be used to diagnose heart illness, arrhythmia, and various other heart-related problems in heart care by analyzing pictures and ECG data. Object identification techniques may additionally be employed to monitor patients' vital signs in real time.
2. Transportation: Object detection has applications in transportation such as tracking traffic, pedestrian identification, and vehicle recognition. It is additionally applicable in self-driving cars to identify and track road items including people, automobiles, and traffic signs.
3. Security: Object detection has applications in security such as surveillance, facial recognition, and anomaly detection. It is useful for monitoring public spaces, detecting intruders, and tracking suspicious behavior. It may also be used to allow or restrict access based on face recognition in access control systems.

Some particular industrial uses of deep learning for object identification include:

1. Self-autonomous driving: Object detection is a key aspect of autonomous driving. Object detection algorithms are used by companies like Tesla, Waymo, and Uber to recognize and track things on the road.
2. Quality control: Object detection can be used to detect defects in manufacturing products. Companies like Foxconn and Flextronics are using object detection algorithms to inspect their products in real-time.

3. Retail: Object detection can be used to track customer behavior in retail stores. Object detection algorithms are used by companies such as Walmart and Amazon to analyze user behavior, optimize shop designs, and improve customer experience.
4. Agriculture: Crop growth may be monitored, pests and illnesses identified, and crop production optimized using object detection. Companies like Blue River Technology and John Deere are using object detection algorithms to monitor and optimize agricultural processes.

In conclusion, the industrial applications of object detection using deep learning are vast and varied. We should anticipate more broad application of object identification algorithms in numerous sectors as technology evolves and improves.

5.Potential Development in Object Detection in Future:

Object detection has made great advances in the past few decades, yet there is still much space for improvement in the discipline. Here are some prospective advances in object detection which could occur:

1. More accurate and efficient algorithms: Deep learning techniques have considerably increased object detection accuracy, yet there is always potential for improvement. Future methods may be able to recognize things more accurately and quickly while utilizing fewer resources than existing techniques.
2. Improved object tracking: Unlike object detection, which can identify things in just one frame, object tracking may monitor objects across numerous frames. Future advancements in object tracking may improve its accuracy, dependability, and efficiency.
3. Robust detection in challenging environments: Object detection can be challenging in environments with low light, clutter, or occlusions. Future developments may improve the ability of algorithms to detect objects in these challenging environments.
4. Live video analysis: For image analysis, object detection is usually utilized, but future improvements may enable live video analysis. This would enable real-time monitoring of objects in film, which might be useful in a number of applications.
5. Integration with technologies: Object detection algorithms can be used with additional technologies like augmented reality, self-driving cars, or robots. This would allow for more complicated applications requiring object detection as a component of a broader system.
6. Development of object detection systems: Object detection has a wide range of applications, and future improvements may result in the creation of specialized object detection systems. For example, medical imaging object detecting systems, autonomous cars, or security systems.

To summarize, object detection has made great strides in the past few years, but is still a significant opportunity for improvement. As technology progresses, we may anticipate increasingly precise, effective, and adaptable item identification techniques that can be used across several sectors.

6. 1 Current Limitations:

1. Limited Generalization: The performance of object detection models can be limited when the objects in the test data differ significantly from the objects in the training data. This is known as limited generalization, and it can be due to factors such as changes in lighting, background, or object appearance.
2. Limited Data Availability: Deep learning object detection models typically require large amounts of annotated training data to achieve good performance. However, in many application domains, such as medical imaging or satellite imagery, annotated data may be limited, making it challenging to train accurate models.
3. Computational Cost: Deep learning detection of objects algorithms can have a high computational cost, making them challenging to deploy on resource-constrained devices like cell phones or embedded systems.
4. Limited Interpretability: Models based on deep learning are sometimes referred to as "black boxes" since it can be hard to comprehend how they make predictions. This is a serious constraint in applications requiring interpretability, such as medical evaluation or legal decision-making.

6.2 Solutions:

1. Data Augmentation: The use of methods of data augmentation to provide extra training data that captures differences in object physical appearance, lighting, and backdrop is one solution to the restricted generalization problem.
2. Transfer Learning: Transfer learning is an approach in which a model is trained on a big dataset and then fine-tuned on a smaller dataset. This strategy may assist in alleviating the problem of restricted data availability.
3. Model Compression: Model compression refers to a collection of strategies for lowering the computational expense of deep neural network models, thereby rendering them more suitable for deployment on devices with limited resources. Pruning, quantization, and reduction are examples of these procedures.
4. Interpretable Models: Researchers are developing new techniques for making deep learning models more interpretable, such as attention mechanisms and visualization techniques. Additionally, some researchers are exploring the use of alternative machine learning models, such as decision trees or rule-based systems, that are more transparent and easier to interpret.

References:

[1] Tan, Mingxing, Ruoming Pang, and Quoc V. Le. "Efficientdet: Scalable and efficient object detection." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.

[2] Pham, Vung, Chau Pham, and Tommy Dang. "Road damage detection and classification with detectron2 and faster r-cnn." 2020 IEEE International Conference on Big Data (Big Data). IEEE, 2020.

[3] Bochkovskiy, Alexey, Chien-Yao Wang, and Hong-Yuan Mark Liao. "Yolov4: Optimal speed and accuracy of object detection." arXiv preprint arXiv:2004.10934 (2020).

[4] Law, Hei, et al. "Cornernet-lite: Efficient keypoint based object detection." arXiv preprint arXiv:1904.08900 (2019).

[5] Uzkent, Burak, Christopher Yeh, and Stefano Ermon. "Efficient object detection in large images using deep reinforcement learning." Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2020.

[6] Hein, Dennis. "Social Distancing AI: Using super-resolution to train an object detection model on low resolution images." LUNFMS-4047-2020 (2020).

[7] Tan, Mingxing, Ruoming Pang, and Quoc V. Le. "Efficientdet: Scalable and efficient object detection." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.

[8] Zhu, Xizhou, et al. "Deformable detr: Deformable transformers for end-to-end object detection." arXiv preprint arXiv:2010.04159 (2020).

[9] Fan, Qi, et al. "Few-shot object detection with attention-RPN and multi-relation detector." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.