EECE 7370 Assignment 3

1. **Reviewer:** Sakshi Bhatia, Reviewed on: 4 October, 2024

2. **Paper Details:**
    a. Paper Title: YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors
    b. Authors: Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao,
    c. Published: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)
    d. Citation: C. -Y. Wang, A. Bochkovskiy and H. -Y. M. Liao, "YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors," 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 2023, pp. 7464-7475

3. **Paper Summary:** This paper introduces a new version of YOLO (You Only Look Once) family of object detection systems. They have proposed a new architecture for real-time object detector and the corresponding scaling method, and incorporated a trainable bag-of-freebies approach. This improved speed and accuracy while performing in real-time. The paper shows improvement over YOLOv4, Scaled-YOLOv4, and YOLOR on MS COCO dataset.

4. **Main Contributions:**
    a. Enhanced architecture design: The paper proposes Extended-ELAN (E-ELAN) based on Efficient Layer Aggregation Network (ELAN). This architecture continuously enhances the network's learning ability, enabling it to learn more diverse features.
    b. The paper introduces model scaling techniques for concatenation-based architectures which will allow these architectures to use hardware efficiently.
    c. Trainable bag-of-freebies: The paper introduces optimization techniques called trainable bag-of-freebies. These strengthen the training cost for improving accuracy of object detection, without increasing the inference cost. This includes methods such as model re-parametrization, dynamic labeling, and scaling techniques for concatenation-based architectures.
    d. State-of-the-art performance for object detection on COCO dataset over previous YOLO versions.

5. **Paper Strengths:**
    a. Real-time performance: The paper improves efficiency while achieving the same or better AP on various benchmarks compared to previous YOLO versions, hence making it a suitable choice for real-time applications.
    b. Innovative techniques: The paper introduces several different methods and modifications to improve the existing object detection approaches.
    c. Extensive ablation studies: The authors present extensive experimental studies that illustrate the impact of each proposed component on improving the overall performance. This also makes the method easier to understand and adapt according to your unique requirements.

6. **Paper Weaknesses:**
    a. Complexity of Implementation: The introduction of several new components might complicate the implementation of YOLOv7.
    b. Limited exploration of diverse datasets: The paper mainly focuses on experiments conducted over the COCO dataset, with no analysis of performance on other datasets.
    c. Lack of novelty: The paper introduces innovative and effective methods, but they are simply built over the existing YOLO techniques.

7. **Comments on Experiments:** The experiments conducted demonstrate the improvements in YOLOv7 over previous versions of YOLO and other real-time object detectors. Ablation studies effectively show the impact of each proposed component in the improvement of both speed and accuracy over a variety of model sizes. However, comparisons on diverse datasets would have strengthened the author's claims of improvements over existing object detection models.

8. **How could the work be extended:**
    a. Robust testing: Further research could explore robustness under adversarial conditions, such as occlusion, extreme lighting changes, or noisy environments.
    b. Domain-specific testing: Future research could focus on studying YOLOv7 on specific applications such as medical imaging to evaluate the generability of this method.
    c. This work could be expanded to study object detection in the 3D scenario.

9. **Additional Comments**
    a. **Unclear Points:**
        i. The paper needs more clarification on how much the "trainable bag-of-freebies" techniques add to the training time and complexity
        ii. The paper does not discuss how much energy efficiency is achieved in hardware-constrained models, which could be important for the deployment of large-scale models on tiny hardware devices.
    b. **Open Research Questions:**
        i. How can the work be extended to perform tasks like image segmentation and 3D object detection in real-time?
        ii. Can the proposed "trainable bag-of-freebies" approach be applied to other deep learning techniques, in general?
    c. **Applications:**
        i. Autonomous driving and Robotics: For detecting pedestrians, traffic signals, road signs, obstacles, and other vehicles in real-time.
        ii. Surveillance: This method can be used to detect real-time security threats via cameras.
        iii. AR/VR applications need real-time object detection, and the techniques proposed in this paper could prove to be quite successful for these applications.

10. **Final Recommendation: Accept**

The paper is well-written with sufficient explanations and experiments to illustrate the impact of various components, and compares performance on edge, normal, and cloud GPUs. The experiments performed show enhanced speed and accuracy without increasing inference time. Although the experiments could have been conducted on more datasets for a good comparison, the proposed techniques will find applications in many systems and provide further ideas to the Computer Vision community. Hence, this paper deserves to be accepted.