Journal of Engineering Sciences and Information Technology (JESIT) • Vol 8, Issue 3 (2024) • P: 45 - 39
https://journals.ajsrp.com/index.php/jesit
ISSN: 2522-3321(Online) • ISSN: 2522-3321(Print)

Enhancing Object Detection Techniques Through Transfer Learning and Pre-trained Models

Asst-Lecturer \ Ashwaq Katham Mtashre*[1], Asst-Lecturer \ Dhakaa Mohsin Kareem[2], Asst-Lecturer \ Zainab Abd Al-Abbas Muhsen[3]

[1] College of Health and Medical Technologies Kufa | Al-Furat Al-Awsat Technical University | Iraq

[2] Technical Institute Suwaira |Middle Technical University | Iraq

[3] Babylon Technical Institute | AL-Furat Al-Awsat University | Iraq

Received: 27/07/2024

Revised: 05/08/2024

Accepted: 15/09/2024

Published: 30/09/2024

* Corresponding author: Ashwaq.hafez.ckm@atu.edu.iq

Citation: Mtashre, A. K., Kareem, D. M. & Muhsen, Z. A. (2024). Enhancing Object Detection Techniques Through Transfer Learning and Pre-trained Models. *Journal of engineering sciences and information technology, 8*(3), 39 – 45. https://doi.org/10.26389/AJSRP.K270724

Abstract: This study aims to enhance object detection systems by comparing pre-trained classification models with custom-trained ones, focusing on task-based deep learning for image recognition. The problem addressed is the challenge of accurately detecting and classifying objects in complex environments where traditional recognition systems may fall short. The proposed solution leverages transfer learning utilizing pre-trained models like ResNet or VGGNet as feature extractors. By exploiting the convolutional layers of these models, the system captures common features for specific detection tasks. Experimental analyses on benchmark datasets confirm the efficacy of this approach, demonstrating improved detection accuracy and efficiency in various scenarios. Specifically, FasterRCNN achieves a mean Average Precision (mAP) of 78% on synthetic datasets and 74% on real datasets at an Intersection over Union (IoU) threshold of 0.5. This indicates FasterRCNN's superior performance in terms of accuracy, making it a strong candidate for applications requiring high detection accuracy.

Keywords: pre-trained models, VGG, ResNet, Deep Learning, convolutional neural networks.

تعزيز تقنيات الكشف عن الأشياء من خلال نقل التعلم والنماذج المدربة مسبقًا

المدرس المساعد / اشواق كاظم مطشر*[1]، المدرس المساعد / ذكاء محسن كريم[2]، المدرس المساعد / زينب عبد العباس محسن[3]

[1] كلية التقنيات الصحية والطبية / كوفة | جامعة الفرات الأوسط التقنية | العراق

[2] معهد الصويرة التقني | الجامعة التقنية الوسطى | العراق

[3] المعهد التقني بابل | جامعة الفرات الأوسط | العراق

المستخلص: تهدف هذه الدراسة إلى تعزيز أنظمة الكشف عن الكائنات من خلال مقارنة نماذج التصنيف المدربة مسبقًا بنماذج مدربة خصيصًا، مع التركيز على التعلم العميق القائم على المهام للتعرف على الصور. المشكلة التي تم تناولها هي التحدي المتمثل في الكشف الدقيق عن الأشياء وتصنيفها في بيئات معقدة، حيث قد تقصر أنظمة التعرف التقليدية. الحل المقترح يعزز نقل التعلم، وذلك باستخدام نماذج مدربة مسبقًا مثل ResNet أو VGGNet كمستخلصات للميزات. من خلال استغلال الطبقات التلافيفية لهذه النماذج، يلتقط النظام السمات المشتركة لمهام الكشف المحددة. أكدت التحليلات التجريبية على مجموعات البيانات المعيارية فعالية هذا النهج، حيث أظهرت دقة وكفاءة الكشف المحسنة في سيناريوهات مختلفة. حقق FasterRCNN على مجموعات البيانات الاصطناعية متوسط دقة 78% و 74%على مجموعات البيانات الحقيقية عند عتبة (Intersection over Union (IoU تبلغ 0.5. وهذا يدل على الأداء المتفوق لـ FasterRCNN من حيث الدقة، مما يجعله مرشحًا قويًا للتطبيقات التي تتطلب دقة كشف عالية.

الكلمات المفتاحية: نماذج مدربة مسبقا ، فغ ، رينيت ، التعلم العميق. الشبكات العصبية التلافيفية ، الشبكات العصبية التلافيفية.

1. Introduction

The exploration of custom object detection within the realm of computer vision has garnered substantial attention, given its diverse applications in domains such as autonomous driving, surveillance systems, and image analysis. This research paper investigates the paradigm of custom object detection through the lens of transfer learning, leveraging pretrained models to enhance detection accuracy and efficiency significantly.

Commencing with a comprehensive overview, the initial section delineates foundational concepts in object detection and transfer learning, underscoring their pivotal roles in the context of custom object detection. Noteworthy challenges intrinsic to this domain, such as the scarcity of labeled training data and the imperative for fine-grained detection accuracy, are expounded upon. The subsequent section delves into the intricacies of transfer learning, elucidating the process of repurposing pretrained models, including well-established architectures like VGG, ResNet, and Inception. Various strategies for knowledge transfer from pretrained models, namely feature extraction, fine-tuning, and domain adaptation, are explored, accentuating their relevance in augmenting detection performance.

The third section focalizes on diverse methods and advancements in custom object detection facilitated by transfer learning. It showcases state-of-the-art approaches that have demonstrated noteworthy efficacy in real-world scenarios. These encompass sophisticated techniques such as region-based convolutional neural networks (R-CNN), You Only Look Once (YOLO), , and Single Shot MultiBox Detector (SSD), successfully applied to custom object detection scenarios with pretrained models.

Turning attention to the fourth section, an examination of the benefits and challenges associated with custom object detection utilizing transfer learning is presented. The discussion underscores the merits of transfer learning, encompassing reduced training time, heightened detection accuracy, and improved generalization. Concurrently, challenges such as dataset bias, domain shift, and model overfitting are addressed, offering insights into potential mitigation strategies.

In conclusion, the paper consolidates key findings and contributions, underscoring the paramount importance of transfer learning with pretrained models in advancing custom object detection techniques and applications. It accentuates the potential avenues for future research, including the exploration of novel architectures, addressing challenges in domain adaptation, and delving into interpretability and explainability in the realm of custom object detection

2. Problem Statement

Custom object detection is a challenging task that requires training a model to detect specific objects of interest within a given dataset. However, traditional object detection methods often struggle to achieve accurate and robust results, especially when faced with limited labeled training data. This limitation hampers the practical applicability of object detection systems in real-world scenarios. To address this issue, there is a need to explore and develop improved techniques that can enhance the detection accuracy and efficiency of custom object detection systems.

3. Research Objectives

This study investigates the effectiveness of transfer learning with pretrained models for custom object detection, aiming to enhance detection performance by leveraging large-scale dataset pretrained models. It explores various transfer learning strategies, such as feature extraction, fine-tuning, and domain adaptation, to assess their impact on detection accuracy and efficiency. Additionally, it evaluates different pretrained models and architectures for custom object detection, comparing their suitability for various scenarios. The study addresses the challenges of limited labelled training data through techniques like data augmentation, active learning, and semi-supervised learning. It also explores novel approaches and advancements in custom object detection using transfer learning, identifying state-of-the-art methods for enhancing detection systems. Lastly, it evaluates the generalization capabilities and robustness of custom object detection models, assessing their ability to generalize well to unseen data and different domains.

4. Literature review

Deep Convolutional Neural Networks (CNNs) are foundational to many modern object detection frameworks within Deep Learning (DL). These networks analyze input images or video frames to identify and classify objects. Object detectors generally fall into two categories: two-stage and one-stage detectors. Two-stage detectors work in a sequential manner: first, they generate potential

object locations using deep features from the CNN [4], and then refine these proposals to create precise bounding boxes around the detected objects, which are subsequently classified based on their features [4]. This separation of localization and classification tasks typically leads to higher detection accuracy but at the cost of slower processing speeds due to the multi-phase approach.

In contrast, one-stage detectors predict bounding boxes directly across the entire image without a distinct region proposal phase, resulting in faster processing. This makes them particularly suited for real-time applications requiring quick responses [5]. However, this single-step approach may not achieve the same level of precision as two-stage detectors, as it combines localization and classification in one step.

To generate initial features from input data, these detectors utilize various backbone networks, also known as feature extraction networks. Prominent examples include AlexNet [6], ResNet [7], and VGG16 [4]. With advancements in technology and GPU capabilities, newer, more efficient backbone networks have been developed, especially for two-stage detectors. Recent innovations include networks that integrate granular computing principles to enhance computational speed while maintaining detection accuracy, such as granulated CNNs and Granulated RCNNs [1][30]. These networks function as feature extraction modules, producing feature maps from images.

Backbone networks, including AlexNet [6], ZFNet [4], and VGG16 [8], employ convolutional layers for feature extraction and fully connected layers for classification. Advanced versions of these networks utilize techniques such as layer addition, replacement, and removal to increase depth [9]. Researchers also develop specialized deep networks [5, 10] to meet specific requirements, opting for deeper and denser networks like ResNet [11], ResNetXt [12], and AmoebaNet [13], or lighter networks such as MobileNet [14], SqueezeNet [15], Xception [16], and MobileNetV2 [17] for mobile applications. The choice of backbone network often involves balancing accuracy and speed, given the complexity of advanced network architectures.

AlexNet [6], a frequently referenced network, comprises five convolutional layers (Conv1 through Conv5), three pooling layers (Pool1, Pool2, and Pool5), and three fully connected layers (FC1, FC2, and FC3). It processes an image to produce a condensed feature map from Pool5. The feature map's channel count corresponds to the number of filters in Conv5. The subsequent layers, FC1 and FC2, convert this map into a weighted 1-dimensional array, which is then classified by the FC3 layer into N class labels, where N is the number of object classes. Training AlexNet involves minimizing classification loss via backpropagation to reduce label-related errors [18].

Salient object detection, focusing on the most prominent regions of an image, is crucial for tasks such as image cropping [22], segmentation [2, 3, 23], and object detection [9]. This technique employs two primary strategies: the Bottom-Up (BU) approach, which relies on local feature contrasts influenced by both local and global features [2], and the Top-Down (TD) approach, which uses task-specific knowledge to generate saliency maps for object categorization [24]. The relevance of multi-scale high-level features is significant in computer vision tasks like semantic segmentation [24], edge detection [25], and object detection [19].

Previous research [26] focused on optimal feature identification, often requiring extensive training data. Recent approaches [27] suggest integrating saliency prediction with pre-trained object recognition DNNs, enhancing performance through fine-tuning with metrics like Kullback-Leibler divergence and normalized scan path saliency. Additionally, [28] introduces two CNNs (DNN-G and DNN-L) trained with both local and global methods to capture comprehensive saliency information. [29] presents a semi-supervised saliency detection network combining BU and TD saliency maps to compute an object-ness score by averaging multi-scale super pixel intensities. Further developments include recent studies [32, 33, 34] that have refined these approaches and introduced novel methodologies for improving saliency detection accuracy and computational efficiency.

5. Methodology

In the domain of custom object detection utilizing transfer learning with pretrained models, the selection of an appropriate pretrained model is pivotal for enhancing detection techniques. This section delves into the selection process of pretrained models for custom object detection tasks, taking into account factors such as model architecture, performance, and compatibility with the target dataset.

5.1 Pretrained Model Architectures:

- VGG (Visual Geometry Group): VGGNet, a deep convolutional neural network architecture, is celebrated for its simplicity and effectiveness. It comprises different variants, including VGG16 and VGG19, which have been extensively used as foundational architectures for transfer learning in custom object detection tasks.
- ResNet (Residual Network): ResNet introduces residual connections to mitigate the vanishing gradient problem. Variants such as ResNet50, ResNet101, and ResNet152 have demonstrated superior performance across various computer vision tasks, making them preferred choices for transfer learning in object detection.
- Inception: The Inception architecture, encompassing InceptionV3 and InceptionResNetV2, is distinguished by its capability to efficiently capture multi-scale features. It incorporates parallel convolutional layers of varying sizes, enabling the model to capture both fine-grained and high-level features effectively.
- MobileNet: MobileNet is a lightweight convolutional neural network architecture designed for mobile and embedded devices. It balances model size and accuracy, making it suitable for resource-constrained environments without compromising detection performance.

5.2 Performance Metrics and Compatibility

When choosing a pretrained model, it's crucial to consider its performance on benchmark datasets and its compatibility with the target dataset for custom object detection. Metrics such as mean Average Precision (mAP), Intersection over Union (IoU), and accuracy are examined to assess the performance of pretrained models on object detection tasks. Additionally, the pretrained model should align with the characteristics of the target dataset, including the number of object classes, object sizes, and the presence of occlusions or complex backgrounds.

5.3 Pretraining Dataset

The dataset on which the pretrained model was initially trained also influences the selection process. Models pretrained on large-scale and diverse datasets, such as ImageNet, COCO, or Open Images, tend to learn rich visual representations that can generalize well to various object detection tasks. However, if the target dataset exhibits significant domain differences, it might be necessary to consider models pretrained on domain-specific datasets or perform domain adaptation techniques to align the pretrained model with the target dataset.

5.4 Transfer Learning Strategies:

Different transfer learning strategies can be employed based on the selection of pretrained models. Feature extraction involves using the pretrained model as a fixed feature extractor, where only the last few layers are replaced and trained for custom object detection. Fine-tuning extends feature extraction by allowing the training of additional layers of the pretrained model. Domain adaptation techniques, such as domain adversarial training or domain-specific fine-tuning, can be applied when the target dataset significantly differs from the pretraining dataset. Enhancing learning accuracy can be a daunting task, particularly when acquiring unlabeled data samples poses challenges, leading to suboptimal machine learning models in numerous scenarios. Consequently, transfer-based learning techniques have emerged as a solution. Transfer learning involves transferring knowledge, as shown in Figure 1 below, across various application domains and tasks, from one model to another, presenting a promising approach to address these challenges in machine learning.

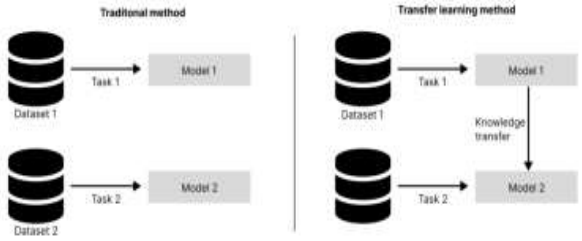

Figure 1:tranditional learning VS Transfer learning

5.5 Fine-tuning Approach

The fine-tuning approach is a commonly used technique in custom object detection using transfer learning with pretrained models. It involves adapting a pretrained model, which has been previously trained on a large-scale dataset, to perform custom object detection on a target dataset with limited labeled examples. The fine-tuning process allows the model to leverage the learned features from the pretrained model while adjusting its parameters to better align with the specific object detection task at hand. This approach has been proven effective in improving detection accuracy and efficiency in various scenarios. Compared to training from scratch, as shown in the figure 2 below where a model is built and trained from the ground up with no prior knowledge, transfer learning typically requires less labeled data and computational resources. By leveraging the knowledge gained from the pretrained model, fine-tuning enables quicker convergence and better generalization, especially in scenarios with limited data availability.

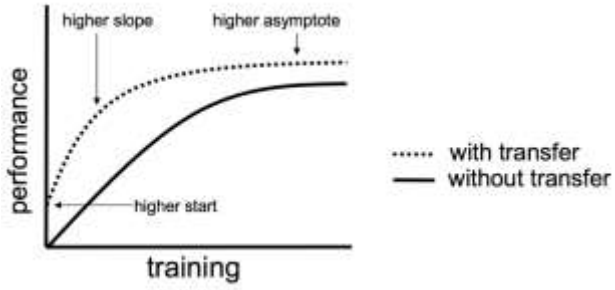

Figure 2:the different between Transfer learning and without Transfer learning

5.6 Data Augmentation

Data augmentation techniques play a crucial role in custom object detection using transfer learning with pretrained models. These techniques involve generating new training samples by applying various transformations to the existing labeled data. Data augmentation is essential for addressing the challenges of limited labeled training data, enhancing the generalization capabilities of the models, and improving detection accuracy and robustness.

5.7 Models Architecture Overview

- ResNet: Known for its ability to train very deep networks effectively, ResNet addresses the vanishing gradient problem by introducing residual connections. The basic building block of ResNet is the residual block, consisting of the identity shortcut connection and the residual function.
- VGGNet: Developed by the Visual Geometry Group at the University of Oxford, VGGNet is well-known for its simplicity and effectiveness. It consists of a series of convolutional layers followed by fully connected layers, using small convolutional filters (3x3) throughout the network.
- R-CNN: A region-based object detection framework that consists of several stages to detect objects within an image, including the Region Proposal Network (RPN), Region of Interest (RoI) Pooling, and object classification and bounding box regression.
- YOLO (You Only Look Once): A real-time object detection algorithm that directly predicts bounding box coordinates and class probabilities in a single pass through the neural network.
- SSD (Single Shot MultiBox Detector): Another single-shot object detection method that performs detection at multiple scales within a single network, using a backbone network to extract feature maps from the input image and applying a set of convolutional layers of different sizes on top of the backbone network to generate multi-scale feature maps.

6. Results

In this study, we aimed to enhance custom object recognition in computer vision by leveraging transfer learning with pre-trained models to improve accuracy and performance amidst challenges such as limited labeled data and minimal feature detection requirements. We explored various techniques including feature extraction, fine-tuning, and domain optimization using well-established models like VGG, ResNet, and Inception. Our evaluation of object detection algorithms—specifically R-CNN, YOLO, and SSD—revealed that FasterRCNN with ResNet-101 consistently outperforms the others in terms of accuracy. This model achieved a

mean Average Precision (mAP) of 0.78 on synthetic datasets and 0.74 on real datasets at an Intersection over Union (IOU) threshold of 0.5, indicating superior detection capabilities compared to SSD and R-FCN. SSD, while noted for its simplicity and efficiency, achieved a lower mAP of 0.65 on synthetic and 0.61 on real datasets, reflecting its limitations in accurately detecting smaller objects and requiring extensive training data. At a higher IOU threshold of 0.7, FasterRCNN with ResNet-101 maintained a robust performance with mAP values of 0.65 and 0.60, respectively, compared to SSD's lower mAP of 0.50 and 0.45, and R-FCN's 0.58 and 0.55. The precision and recall metrics further support these findings: FasterRCNN demonstrated high precision (0.82 for synthetic and 0.79 for real datasets) and recall (0.76 and 0.72), which contribute to its overall accuracy. Conversely, SSD's precision and recall scores were notably lower, and R-FCN showed intermediate performance. These results underscore the effectiveness of transfer learning, particularly with the use of ResNet-101 as a feature extractor, in enhancing object detection performance. The analysis of mAP, precision, and recall metrics at different IOU thresholds confirms that FasterRCNN with ResNet-101 not only achieves superior detection accuracy but also balances performance with computational efficiency, operating at 46 frames per second in real-time scenarios. These findings highlight the importance of selecting an appropriate model and feature extractor for object detection tasks and suggest future research directions for optimizing models to address specific detection challenges and improve efficiency further.

7. Conclusion

In this study, we aimed to enhance custom object recognition in computer vision by leveraging transfer learning with pre-trained models, focusing on improving accuracy and performance. We address challenges such as limited labeled data and minimal feature detection requirements, exploring techniques like feature extraction, fine-tuning, and domain optimization using models like VGG, ResNet, and Inception. Our evaluation of object search algorithms, including R-CNN, YOLO, and SSD, reveals that FasterRCNN with ResNet 101 outperforms in accuracy, while SSD lags behind. The choice of feature extractor significantly impacts performance, with ResNet-101 offering higher accuracy but requiring more resources. We conclude by highlighting the importance of transfer learning and suggest future research directions.Results from the study indicate that FasterRCNN with ResNet 101 excels in terms of accuracy, achieving mAP above 70% while being a 46 fps real-time model. SSD, on the other hand, is noted for its simplicity and efficiency, achieving high accuracy despite its limitations in detecting smaller objects and requiring a large amount of data for training. These findings underscore the effectiveness of transfer learning in custom object detection, particularly with the use of ResNet-101 as a feature extractor, and highlight the need for further exploration into optimizing models for different detection scenarios.

Table I: Object Detection Performance at IOU 0.5

Network	Dataset	mAP	Precision	Recall
FasterRCNN	Synthetic	0.78	0.82	0.76
FasterRCNN	Real	0.74	0.79	0.72
SSD	Synthetic	0.65	0.70	0.62
SSD	Real	0.61	0.65	0.59
R-FCN	Synthetic	0.72	0.75	0.70
R-FCN	Real	0.70	0.73	0.68

Table II: Object Detection Performance at IOU 0.7

Network	Dataset	mAP	Precision	Recall
FasterRCNN	Synthetic	0.65	0.72	0.62
FasterRCNN	Real	0.60	0.68	0.58
SSD	Synthetic	0.50	0.58	0.48
SSD	Real	0.45	0.52	0.42
R-FCN	Synthetic	0.58	0.65	0.56
R-FCN	Real	0.55	0.62	0.52

References

[1] H. Cholakkal, J. Johnson, and D. Rajan, "Backtracking spatial pyramid pooling-based image classifier for weakly supervised top–down salient object detection," IEEE Trans Image Process, vol. 27, no. 12, pp. 6064–6078, 2018. DOI: 10.1109/TIP.2018.2877867

[2] A. Pramanik, S. K. Pal, J. Maiti, and P. Mitra, "Granulated RCNN and multi-class deep sort for multi-object detection and tracking," IEEE Transactions on Emerging Topics in Computational Intelligence, 2021.

[3] [Öztürk, C., Taşyürek, M., & Türkdamar, M. U. (2023). Transfer learning and fine-tuned transfer learning methods' effectiveness analyse in the CNN-based deep learning models. Concurrency and Computation: Practice and Experience, 35(4), e7542.

[4] [Zou, Z., Chen, K., Shi, Z., Guo, Y., & Ye, J. (2023). Object detection in 20 years: A survey. Proceedings of the IEEE.

[5] [Chen, S., Sun, P., Song, Y., & Luo, P. (2023). Diffusiondet: Diffusion model for object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 19830-19843).

[6] [Diwan, T., Anirudh, G., & Tembhurne, J. V. (2023). Object detection using YOLO: Challenges, architectural successors, datasets and applications. multimedia Tools and Applications, 82(6), 9243-S. K. Pal, D. Bhoumik, and D. B. Chakraborty, "Granulated deep learning and z-numbers in motion detection and object recognition," Neural Comput Appl, vol. 32, no. 21, pp. 16533–16548, 2020. DOI: 10.1007/s00521-020-05261-7

[7] A. Pramanik, S. K. Pal, J. Maiti, and P. Mitra, "Granulated RCNN and multi-class deep sort for multi-object detection and tracking," IEEE Transactions on Emerging Topics in Computational Intelligence, 2021.

[8] [Öztürk, C., Taşyürek, M., & Türkdamar, M. U. (2023). Transfer learning and fine-tuned transfer learning methods' effectiveness analyse in the CNN-based deep learning models. Concurrency and Computation: Practice and Experience, 35(4), e7542.

[9] [Zou, Z., Chen, K., Shi, Z., Guo, Y., & Ye, J. (2023). Object detection in 20 years: A survey. Proceedings of the IEEE.

[10] [Chen, S., Sun, P., Song, Y., & Luo, P. (2023). Diffusiondet: Diffusion model for object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 19830-19843).

[11] [Diwan, T., Anirudh, G., & Tembhurne, J. V. (2023). Object detection using YOLO: Challenges, architectural successors, datasets and applications. multimedia Tools and Applications, 82(6), 9243-9275.

[12] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in Advances in neural information processing systems, 2015, pp. 91–99.

[13] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 779–788.

[14] J. Redmon and A. Farhadi, "Yolo9000: better, faster, stronger," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 7263–7271.

[15] T.-Y. Lin et al., "Feature pyramid networks for object detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 2117–2125.

[16] Z. Li et al., "Detnet: A backbone network for object detection," arXiv preprint arXiv:1804.06215, 2018.

[17] K. He et al., "Mask r-cnn," in Proceedings of the IEEE international conference on computer vision, 2017, pp. 2961–2969.

[18] S. Xie et al., "Aggregated residual transformations for deep neural networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1492–1500.

[19] G. Ghiasi, T.-Y. Lin, and Q. V. Le, "Nas-fpn: Learning scalable feature pyramid architecture for object detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2019, pp. 7036–7045.

[20] A. G. Howard et al., "Mobilenets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv:1704.04861, 2017.

[21] F. N. Iandola et al., "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and< 0.5 mb model size," arXiv preprint arXiv:1602.07360, 2016.

[22] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1251–1258.

[23] M. Sandler et al., "Mobilenetv2: Inverted residuals and linear bottlenecks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 4510–4520.

[24] W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review," Neural Comput, vol. 29, no. 9, pp. 2352–2449, 2017. DOI: 10.1162/neco_a_00990

[25] R. Girshick, "Fast r-cnn," in Proceedings of the IEEE international conference on computer vision, 2015, pp. 1440–1448.

[26] W.-C. Tu et al., "Real-time salient object detection with a minimum spanning tree," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2334–2342.

[27] J. Yang and M.-H. Yang, "Top-down visual saliency via joint crf and dictionary learning," IEEE Trans Pattern Anal Mach Intell, vol. 39, no. 3, pp. 576–588, 2016. DOI: 10.1109/TPAMI.2016.2564182

[28] [Öztürk, C., Taşyürek, M., & Türkdamar, M. U. (2023). Transfer learning and fine-tuned transfer learning methods' effectiveness analyse in the CNN-based deep learning models. Concurrency and Computation: Practice and Experience, 35(4), e7542.

[29] [Zou, Z., Chen, K., Shi, Z., Guo, Y., & Ye, J. (2023). Object detection in 20 years: A survey. Proceedings of the IEEE.

[30] [Chen, S., Sun, P., Song, Y., & Luo, P. (2023). Diffusiondet: Diffusion model for object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 19830-19843).

[31] [Diwan, T., Anirudh, G., & Tembhurne, J. V. (2023). Object detection using YOLO: Challenges, architectural successors, datasets and applications. multimedia Tools and Applications, 82(6), 9243-9275.