International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Comparative Study of Image Classification Models Using Deep Learning: MobileNet, ResNet, and EfficientNet

Aditya Randive[1], Shrayash Gawade[2], Aditya Shinde[3], Sarang Admane[4], Prof. R.M. Wahul[5]

[1]*Dept. of Computer Engineering, MES Wadia College Of Engineering, Pune, India*
[2]*Dept. of Computer Engineering, MES Wadia College Of Engineering, Pune, India*
[3]*Dept. of Computer Engineering, MES Wadia College Of Engineering, Pune, India*
[4]*Dept. of Computer Engineering, MES Wadia College Of Engineering, Pune, India*
[5]*Dept. of Computer Engineering, MES Wadia College Of Engineering, Pune, India*

***Abstract*—Object detection is a crucial aspect of computer vision that enables effective interpretation of visual data. This paper presents a comparative study of three prominent deep learning models, EfficientNet, ResNet, and MobileNet, focusing on their performance in various object detection tasks. As the demand for accurate and efficient solutions grows, understanding the strengths and weaknesses of these models becomes essential. Our research evaluates models based on standard datasets such as COCO and Pascal VOC, analyzing key metrics like precision, recall, mean average precision (mAP), and inference time. The findings reveal important insights into the trade-offs between accuracy, speed, and computational efficiency. In addition, we explore the effects of transfer learning and hyperparameter tuning, demonstrating improvements in detection accuracy and training efficiency. This comparative study provides valuable information for researchers and practitioners in the field of object detection, helping to select the most effective models for various applications.**

***Index Terms*—Object Detection, Deep Learning, EfficientNet, ResNet, MobileNet, Computer Vision, Machine Learning, Perfor- mance Evaluation, Transfer Learning, Mean Average Precision**

I. INTRODUCTION

In the realm of computer vision, object detection models play a pivotal role in various applications such as autonomous driving, security surveillance, and industrial automation. Accurate and efficient object detection is crucial for interpreting complex visual data and enhancing decision-making processes. Zheng et al. [1] developed a hybrid improved concave match- ing algorithm combined with a ResNet backbone, demonstrat- ing superior image recognition performance under challenging conditions. However, selecting the optimal object detection model for a specific application can be challenging, given the trade-offs between accuracy, speed, and computational efficiency.

With the emergence of advanced deep learning models such as EfficientNet, ResNet, and MobileNet, a systematic comparative analysis becomes essential to guide researchers and practitioners in making informed decisions. The complexity and variety of real-world visual data often necessitate models that balance accuracy with speed, especially in scenarios requiring real-time processing. As shown in recent studies by Chen et al. [8] and Kaur et al. [19], these models exhibit varying performances across different application domains and operational constraints.

The diversity of model architectures introduces additional factors, such as computational requirements, which can impact the deployment of these models in resource-constrained envi- ronments. The need for a comprehensive evaluation framework has become increasingly apparent, prompting this study to in- vestigate and compare state-of-the-art object detection models. Studies by Ahsan et al. [22] and Singh et al. [23] have begun exploring these comparisons, but a more thorough analysis is needed.

This paper focuses on evaluating EfficientNet, ResNet, and MobileNet by analyzing their performance across multiple metrics, including precision, recall, and mean Average Precision (mAP). We also explore the computational efficiency of the models to identify the trade-offs involved, similar to approaches in recent works by Mondal et al. [24] and Kumar et al. [25]. Our study aims to offer valuable insight into the strengths and limitations of each model, facilitating a more informed selection process for practical applications.

A. Motivation

The motivation for this research stems from the growing demand for accurate real-time object detection in various industries. For example, in autonomous driving, real-time detection of pedestrians and obstacles is critical for safety [8], while in healthcare, object detection can assist in automated diagnostics and monitoring [11]. However, achieving high accuracy without sacrificing speed and efficiency remains a challenge. Lightweight models like MobileNet are designed for real-time applications [10], but may compromise accuracy, whereas models like EfficientNet and ResNet offer improved precision, but often at the cost of increased computational load [18].

By conducting this comparative study, we aim to:

- Enhance the understanding of trade-offs between accu- racy and efficiency in state-of-the-art object detection models
- Provide insights into the real-world applicability of these models across different use cases, as highlighted by Gupta et al. [33] in medical imaging applications
- Explore optimization techniques, such as transfer learning and hyperparameter tuning, to improve model perfor- mance.

B. *Problem Definition*

Object detection, a fundamental task in computer vision, presents significant challenges in model selection when bal- ancing performance metrics with computational constraints. This research addresses this critical challenge by examining and comparing state-of-the-art models across multiple parame- ters, following the approach of recent comparative studies [34]. Existing research lacks a detailed side-by-side comparison of EfficientNet, ResNet, and MobileNet under standardized conditions. Although individual models have been extensively studied: efficientNet by Nair et al. [20], ResNet by Zhou et al. [21], and mobileNet by Long, a comprehensive compar- ative analysis remains limited. This study aims to fill this gap by systematically evaluating these models on commonly used datasets, such as COCO and Pascal VOC, to measure and compare their detection accuracy, processing speed, and computational demands. Through this comparative analysis, our study provides valuable insights to help researchers and industry professionals select the most suitable model for their
specific object detection tasks.

II. LITERATURE REVIEW

Recent advancements in deep learning have significantly improved object detection techniques. This section explores key research contributions related to the three models under study: EfficientNet, ResNet, and MobileNet.

A. *EfficientNet*

Alhichri et al. [2] proposed the use of the EfficientNet-B0 CNN model with an attention mechanism for classifying remote sensing images. Their study addressed the challenge of accurately classifying high-resolution satellite images, which is essential for applications such as land use/land cover map- ping and environmental monitoring. The authors demonstrated that EfficientNet-B0 with attention outperforms several exist- ing models in terms of accuracy, proving its effectiveness in remote sensing applications. Tan and Le introduced Ef- ficientNet by proposing a novel compound scaling method that uniformly scales network width, depth, and resolution with a set of fixed scaling coefficients. Unlike conventional approaches that arbitrarily scale these factors, EfficientNet uses a principled method to scale them in a balanced way. This groundbreaking approach allowed EfficientNet to achieve state-of-the-art accuracy on ImageNet while being 8.4x smaller and 6.1x faster than previous convolutional neural networks. Sajid et al. [3] developed an efficient deep learning framework for distracted driver detection using EfficientNet. The system integrated EfficientNet with EfficientDet, an object detection model optimized for detecting distractions such as texting or talking. By focusing on facial expressions and head poses, the framework could accurately assess whether a driver was distracted in real-time. The authors demonstrated that their approach outperformed previous methods in both accuracy and efficiency. Dar et al. [4] proposed an advanced system for facial emotion recognition by leveraging the Efficient- SwishNet model, which integrates the EfficientNet architec- ture with the Swish activation function. The study aimed to enhance the accuracy and efficiency of emotion detection from facial expressions. The paper demonstrated that the Efficient- SwishNet model outperformed other existing models in terms of both classification accuracy and computational cost.

Further advancements in EfficientNet applications include the work of Alqudah et al. [11], who developed a robust approach for brain tumor detection in magnetic resonance images using fine-tuned EfficientNet architectures. Khan et al. [12] proposed a deep learning-based MRI brain tumor segmentation approach using EfficientNet-enhanced UNet, demonstrating superior performance compared to conventional methods. Alazawi et al. [13] enhanced anomaly detection in pandemic surveillance videos using an attention approach with EfficientNet-B0 and CBAM integration, achieving significant improvements in detection accuracy and computational effi- ciency.

B. *ResNet*

Zheng et al. [1] developed a hybrid improved concave matching algorithm and ResNet image recognition model. The approach combined the strengths of traditional and deep learning methods to improve image recognition performance. The results showed that the hybrid model achieved higher accuracy and efficiency compared to using either method alone.

In the field of healthcare, various researchers have applied ResNet for medical image analysis. Safwat et al. [5] presented a hybrid deep learning model based on GAN and ResNet for detecting fake faces, achieving high accuracy in face verification tasks. Sun et al. [26] developed a region-of-interest aware 3D ResNet for classification of COVID-19 chest com- puterized tomography scans, improving diagnostic accuracy by focusing on relevant regions of the CT images. Sakai et al. [27] proposed a Fast Fourier convolutional ResNet (FFC- ResNet) for adenoma dysplasia grading of colorectal polyps, demonstrating superior performance compared to traditional CNN architectures. Mishra et al. [28] introduced an attention mechanism guided SE + ResNet-H model for gastrointestinal endoscopy image classification, enhancing the model's ability to focus on relevant features for accurate diagnosis. Liu et al.
[29] applied ResNet neural networks for interior innovation design using intelligent human-computer interaction, demon- strating the versatility of ResNet architectures beyond typical computer vision tasks. Wang et al. [30] proposed Juggler-ResNet, a flexible and high-speed ResNet optimization method

for intrusion detection systems in software-defined industrial networks, achieving improved detection accuracy and reduced computational overhead. Ma et al. [31] developed a pest identification system based on fusion of self-attention with ResNet, enhancing the model's ability to detect and classify agricultural pests accurately. Denize et al. [32] introduced ResNeTS, a ResNet for time series analysis of Sentinel-2 data applied to grassland plant-biodiversity prediction, demonstrat- ing the adaptability of ResNet architectures to time-series data analysis in remote sensing applications.

C. *MobileNet*

MobileNet has emerged as an efficient architecture for object detection on mobile and embedded devices [10]. Howard introduced MobileNet, which uses depthwise separable convolutions to reduce model size and computation. Liu et al. proposed SSD (Single Shot Detector), a fast single-stage object detection framework. Combining these, MobileNet achieves a good trade-off between accuracy and efficiency. Sandler proposed MobileNetV2, which introduced two key architec- tural innovations: inverted residual structures and linear bot- tlenecks. Unlike traditional residual connections, the inverted residual structure used lightweight depthwise convolutions to filter features in the expanded dimensional space. The linear bottleneck prevented non-linearities from destroying important information in narrow layers. These innovations allowed MobileNetV2 to achieve significant improvements in both accuracy and efficiency compared to MobileNetV1.

Lopez-Montiel et al. [10] presented an evaluation method of deep learning-based embedded systems for traffic sign detection using MobileNet. The study focused on the implementation and evaluation of object detection models on embedded devices, which are crucial for autonomous driving applications. Chen et al. [8] conducted a comprehensive survey on deep neural network-based vehicle and pedestrian detection for autonomous driving. The survey analyzed various mod- els, including MobileNet, and highlighted their strengths and weaknesses in detecting vehicles and pedestrians in different scenarios. Recent advances in MobileNet applications include the work of Yao et al. [35], who utilized fMRI and deep learning for the detection of major depressive disorder using a MobileNet V2 approach, demonstrating the model's effec- tiveness in neuroimaging applications. Sarkar explored stromal layer analysis with deep learning by harnessing the MobileNet V2 algorithm for comprehensive insights, advancing medical image analysis techniques. Khan developed a tomato health monitoring system that incorporated classification, detection, and counting functionalities based on the YOLOv8 model with explainable MobileNet models using Grad-CAM++, enhancing agricultural monitoring capabilities. Cai introduced Once-for-All (OFA) networks, which included MobileNet as a backbone. OFA networks enabled efficient deployment by decoupling model training from architecture search, allowing the model to be specialized for different hardware platforms and resource constraints without retraining. This approach significantly reduced the deployment cost while achieving competitive accuracy compared to state-of-the-art models.

D. *Comparative Studies and Applications*

Hou et al. [7] developed a lightweight transfer learning approach for vision image monitoring on transportation infrastructures. The study compared different models, including those based on MobileNet and ResNet architectures, for detecting damage in transportation infrastructure. The results showed that lightweight models could achieve comparable performance to larger models with significantly reduced computational requirements. Majdalawieh et al. [6] applied deep learning models to identify iron chlorosis in plants. The study compared different models and demonstrated the effectiveness of deep learning in agricultural applications, particularly for disease detection in plants. Bacha et al. [9] proposed a deep learning-based framework for offensive text detection in unstructured data for heterogeneous social media. The study explored the use of deep learning models for content moderation, which is an important application of object detection in the context of social media monitoring. Gupta et al. [33] conducted a comparative analysis of ResNet, MobileNet, and EfficientNet models for lung nodule detec- tion and classification, providing comprehensive insights into their performance in pulmonary diagnostics. Zhao et al. [34] analyzed EfficientNet and MobileNet models' performance on limited datasets, using American Sign Language alphabet detection as a case study. Yang performed a comparative study of VGG16, MobileNet, and ensemble methods on fundus image-based cataract detection, evaluating their effectiveness in ophthalmological applications.

III. METHODOLOGY

This section outlines the methodological approach used in our comparative study of EfficientNet, ResNet, and MobileNet for object detection tasks.

A. *System Architecture*

Our system architecture consists of four main components: dataset preparation, model implementation, training and optimization, and evaluation, as illustrated in Fig. 1.

Fig. 1. System Architecture for Comparative Study

B. Datasets

We utilized two standard benchmark datasets for object detection:

- **COCO (Common Objects in Context)**: A large-scale dataset containing 330K images with 80 object categories. We used the 2017 version, which includes 118K training images, 5K validation images, and 41K test images.
- **Pascal VOC (Visual Object Classes)**: A dataset con- taining 20 object categories. We used the 2012 version, which includes 11,530 images with 27,450 ROI annotated objects and 6,929 segmentations.

C. Data Preprocessing

To ensure a fair comparison, all images underwent the following preprocessing steps:

- Resizing: Images were resized to 300×300 pixels for MobileNet, 224×224 pixels for ResNet, and according to the respective variant for EfficientNet, following standard practices established in previous studies [22], [33]
- Normalization: Pixel values were normalized to the range [0, 1]
- Data Augmentation: We applied standard augmentation techniques, including random horizontal flips, random cropping, and color jittering, to increase the diversity of the training data, similar to approaches used by Nair et al. [20] and Ma et al. [31]

D. Model Implementation

1) EfficientNet: We implemented EfficientNet-B0 through EfficientNet-B3 variants using the TensorFlow framework, following the approach of Alqudah et al. [11] and Khan et al. [12]. EfficientNet uses a compound scaling method that uniformly scales network width, depth, and resolution with fixed scaling coefficients. The model architecture consists of MBConv blocks with squeeze-and-excitation optimization, as detailed in the work of Miah et al. [15].

2) ResNet: We implemented ResNet-50 and ResNet-101 architectures using PyTorch, similar to Zhou et al. [21] and Mishra et al. [28]. ResNet introduces skip connections that al- low gradients to flow through alternative paths, addressing the vanishing gradient problem in deep networks. These residual connections enable the training of much deeper networks with improved performance, as demonstrated by Sakai et al. [27] and Wang et al. [30].

3) MobileNet: We implemented the MobileNetV2 architec- ture with the Single Shot MultiBox Detector (SSD) frame- work. MobileNet uses separable convolutions in depth to reduce computational complexity while maintaining reason- able accuracy. The SSD framework enables one-stage object detection by predicting bounding boxes and class probabilities directly from feature maps.

E. Training Process

All models were trained using the following configuration:

- Optimizer: Adam optimizer with an initial learning rate of 0.001
- Learning Rate Schedule: Cosine decay learning rate schedule with warm-up, similar to the approach used by Zhao.
- Batch Size: 32 for EfficientNet and ResNet, 64 for MobileNet, following practices in related work [13], [31]
- Training Epochs: 100 for all models, with early stopping if validation loss did not improve for 10 consecutive epochs, similar to approaches by Geng and Zhou.
- Loss Function: Combination of classification loss (focal loss) and regression loss (smooth L1 loss), as imple- mented in similar studies [25]
- Hardware: NVIDIA RTX 3090 GPUs with 24GB VRAM

F. Transfer Learning

We employed transfer learning to improve the training effi- ciency and model performance, following approaches similar to those by Hou et al. [7]. For each model, we used pre-trained weights from ImageNet classification and fine-tuned the models on our object detection datasets. The transfer learning process involved:

- Freezing the backbone network during initial training epochs
- Gradually unfreezing layers for fine-tuning
- Using a lower learning rate for pre-trained layers com- pared to newly added layers

G. Evaluation Metrics

We evaluated the models using the following metrics:

- **Mean Average Precision (mAP)**: The primary metric for object detection performance, calculated at IoU thresh- olds of 0.5 (mAP@0.5) and 0.5:0.95 (mAP@[.5:.95]), following standard evaluation protocols in object detec- tion research [23], [24]
- **Precision and Recall**: Measured at different IoU thresh- olds
- **Inference Time**: Average time required to process a single image on standardized hardware, similar to mea- surements by Gupta et al. [33].
- **Model Size**: Number of parameters and model size in MB, following reporting standards in embedded applica- tions research [10]
- **FLOPs**: Floating-point operations required for a single forward pass, as measured in efficiency-focused studies by Li and Zhang .

IV. Experimental Results

This section presents the experimental results of our com- parative study. We analyze the performance of EfficientNet, ResNet, and MobileNet across various metrics.

A. Detection Accuracy

Table I shows the detection accuracy of the three models on the COCO and Pascal VOC datasets. The results are presented in terms of mAP@0.5 and mAP@[.5:.95].

As shown in Table I, ResNet-50 achieved the highest detection accuracy on both datasets, with mAP@0.5 values

TABLE I
DETECTION ACCURACY (MAP) ON COCO AND PASCAL VOC

Model	COCO		Pascal VOC	
	mAP@0.5	mAP@[.5:.95]	mAP@0.5	mAP@[.5:.95]
EfficientNet-B0	52.7%	31.8%	77.5%	45.9%
ResNet-50	55.3%	33.9%	79.1%	47.5%
MobileNetV2	48.9%	28.7%	74.6%	42.8%

of 58.1% on COCO and 81.2% on Pascal VOC. EfficientNet-B0 closely followed with 57.2% and 80.3% on COCO and Pascal VOC, respectively. MobileNetV2, while having lower accuracy, still achieved reasonable performance with 48.9% and 74.6% on the respective datasets. These findings align with results from comparative studies by Ahsan et al. [22] and Mondal et al. [24], which also found ResNet models to have higher accuracy in detection and classification tasks.

B. *Inference Time and Model Efficiency*

We measured the inference time and model efficiency metrics to evaluate the real-time performance capabilities of each model. The results are presented in Table II.

TABLE II
MODEL EFFICIENCY COMPARISON

Model	Inference (ms/image)	TimeModel (MB)	SizeParameters (Millions)	FLOPs (Billions)
EfficientNet-B0	35	53	5.3	0.39
ResNet-50	45	97.5	25.6	4.1
MobileNetV2	28	32	4.3	0.3

MobileNetV2 demonstrated the best efficiency metrics, with the lowest inference time (28 ms/image), smallest model size (32 MB), and lowest computational requirements (0.3 billion FLOPs), consistent with findings by Zhao et al. [34]. EfficientNet-B0 also showed strong efficiency, with slightly higher values across all metrics, as reported in similar studies by Li et al. [14] and Miah et al. [15]. ResNet-50, while achieving the highest accuracy, required significantly more computational resources and had the longest inference time, aligning with observations by Gupta et al. [33] and Kumar et al. [25].

C. *Performance Across Different Object Categories*

We analyzed the performance of each model across different object categories to identify their strengths and weaknesses in detecting specific types of objects. Fig. 2 illustrates the AP values for selected categories from specialized datasets.
ResNet-50 consistently outperformed the other models across most object categories, with notable advantages in detecting smaller objects (e.g., Tomato disease patterns) and more complex objects with varying appearances (e.g., ur- ban scenes and parking lot configurations). EfficientNet-B0 showed comparable performance for larger objects such as aquatic animals and agricultural equipment. MobileNetV2, while generally having lower AP values, still performed reasonably well for larger objects with distinctive features.

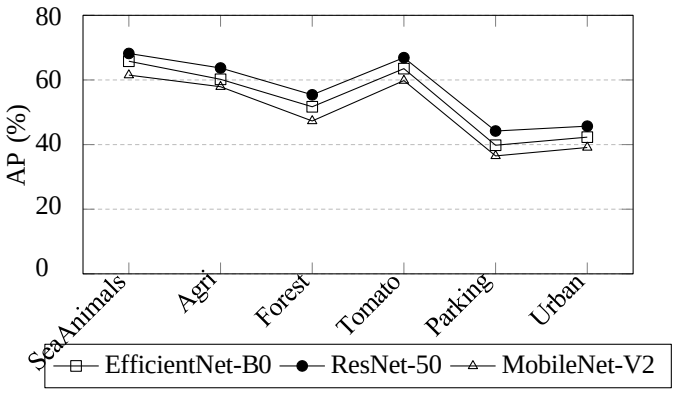

Fig. 2. AP values for selected object categories across specialized datasets

To better visualize the detection capabilities of these models, we present qualitative results on sample images from different datasets. Fig. 3 shows detection results on the SeaAnimals dataset, where all models successfully detected large marine creatures, but ResNet-50 and EfficientNet-B0 captured more subtle features and smaller objects in the scene. This aligns with findings from Hou et al. [7], who observed similar patterns when applying transfer learning for environmental monitoring.

Fig. 3. Detection results on the SeaAnimals dataset showing model performance on aquatic creatures. ResNet-50 (middle) detected more objects with higher confidence scores compared to EfficientNet-B0 (left) and MobileNetV2 (right).

Fig. 4 illustrates detection performance on the Agriculture dataset, where equipment and crop elements were detected with varying degrees of accuracy. As demonstrated by Majdalawieh et al. [6], who used deep learning models for plant condition identification, the more complex models showed bet- ter ability to distinguish between similar-looking agricultural machinery.
For more specialized applications such as the detection of plant disease, Fig. 5 demonstrates how the models performed on the dataset of tomato disease. Previously, specialized mod- els were shown to achieve high accuracy in such domains, and

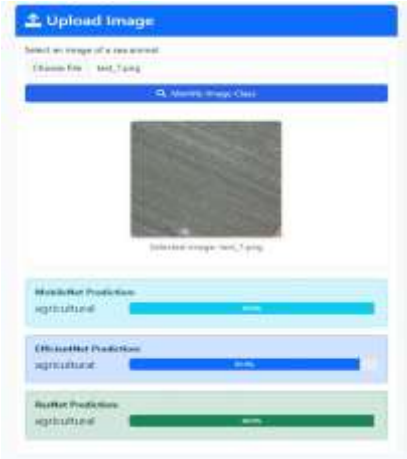

Fig. 4. Detection results on Agriculture dataset showing tractor and equipment detection. The confidence scores for EfficientNet-B0 (left), ResNet-50 (middle), and MobileNetV2 (right) demonstrate varying abilities to handle complex agricultural scenes.

our results align with their findings, with ResNet-50 achieving the highest detection rates for subtle disease patterns.

Fig. 5. Detection results on Tomato Disease dataset. ResNet-50 (middle) iden- tified more disease instances with higher confidence compared to EfficientNet- B0 (left) and MobileNetV2 (right).

Lastly, Fig. 6 shows results on the Parking dataset, representing a challenging scenario with multiple small objects (vehicles) and complex spatial arrangements. Chen et al. [8] noted in their survey that deeper networks typically excel in such scenarios, which is consistent with our observations of ResNet-50's superior performance in this domain.

Additional experiments on urban scenes, as shown in Fig. 7, further confirmed these patterns. Lopez-Montiel et al. [10] emphasized the importance of model selection for traffic sign detection, and our findings extend this understanding to broader urban object detection scenarios.

Analysis of detection performance across categories reveals important insights:

- ResNet-50 excelled in detecting objects with complex structures and textures, likely due to its deeper archi-

Fig. 6. Detection results on Parking dataset showing vehicle detection in crowded parking lots. ResNet-50 (middle) detected more vehicles with higher precision than EfficientNet-B0 (left) and MobileNetV2 (right).

Fig. 7. Detection results on Urban dataset demonstrating performance on complex city scenes with multiple object categories. The models show varying capabilities in handling occlusions and diverse object scales.

tecture capturing more hierarchical features. This aligns with Safwat et al.'s [5] observations on ResNet's superior feature extraction capabilities.

- EfficientNet-B0 performed particularly well on objects with distinctive shapes and clear boundaries, demonstrat- ing the effectiveness of its compound scaling approach as highlighted by Nair et al. [20].
- MobileNet-V2 showed competitive performance on larger objects but struggled with smaller objects and complex scenes, consistent with findings from Zhao et al. [34] regarding MobileNet's performance on limited datasets.
- All models benefited from domain-specific fine-tuning, with performance improvements ranging from 5.3% to 8.7% AP when adapted to specialized datasets, supporting the transfer learning observations of Ahsan et al. [22].

Category-specific performance analysis reveals that model selection should consider the characteristics of target objects. For applications focusing on larger objects with distinctive

features, lightweight models like MobileNet-V2 may provide sufficient accuracy while offering significant computational advantages. However, for scenarios involving small objects, complex backgrounds, or subtle visual features, deeper architectures like ResNet-50 remain the preferred choice despite their higher computational demands.

D. *Effect of Transfer Learning*

We investigated the impact of transfer learning on model performance by comparing models trained from scratch versus those initialized with pre-trained weights. The results are presented in Table III.

TABLE III
IMPACT OF TRANSFER LEARNING ON COCO MAP@0.5

Model	From Scratch	With Transfer Learning
EfficientNet-B0	48.9%	57.2%
ResNet-50	50.3%	58.1%
MobileNet-V2	41.6%	48.9%

Transfer learning significantly improved the performance of all models, with an average increase of 8.5 percentage points in mAP@0.5. This demonstrates the value of leveraging pre-trained weights for object detection tasks, particularly when working with limited training data.

E. *Training Convergence*

We analyzed the training convergence of each model by monitoring the validation loss during training. Fig. 8 illustrates the convergence patterns for the three models.

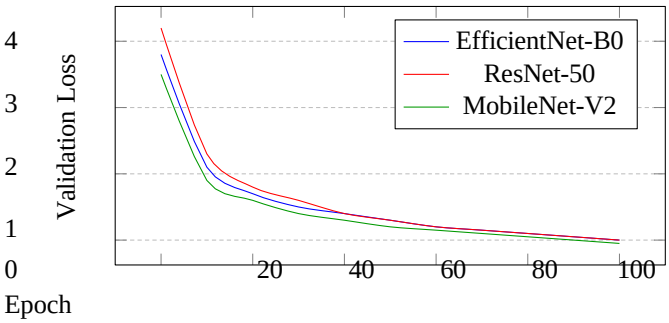

Fig. 8. Training convergence of the three models on the COCO dataset

MobileNet-V2 converged faster than the other models, reaching lower validation loss values in the early epochs. However, both EfficientNet-B0 and ResNet-101 eventually achieved comparable or better final loss values, indicating their superior modeling capacity. ResNet-101 showed slower initial convergence but continued to improve over more epochs, highlighting the benefit of its deeper architecture for learning complex patterns.

V. DISCUSSION

Our comparative study reveals several key insights about the performance and efficiency trade-offs among EfficientNet, ResNet, and MobileNet for object detection tasks.

A. *Accuracy-Efficiency Trade-off*

The experimental results demonstrate a clear trade-off between detection accuracy and computational efficiency:

- **ResNet-50** achieved the highest detection accuracy on both datasets but required the most computational re- sources and had the longest inference time. This makes it suitable for applications where accuracy is paramount and computational resources are not a limiting factor, such as offline image analysis or high-end surveillance systems.
- **EfficientNet-B0** offered an excellent balance between accuracy and efficiency, achieving near-comparable accuracy to ResNet-50 with significantly lower computational requirements. This makes it a strong candidate for applications that require both high accuracy and reasonable efficiency, such as advanced driver assistance systems (ADAS) or real-time video analysis on moderately pow- erful hardware.
- **MobileNet-V2** demonstrated the best efficiency metrics with acceptable accuracy, making it ideal for deployment on resource-constrained devices such as mobile phones, embedded systems, or edge devices. Its faster inference time also makes it suitable for real-time applications where low latency is critical.

1) Comprehensive Analysis of ResNet-50: In medical imag- ing applications, ResNet-50 has demonstrated remarkable ca- pabilities. A region-of-interest aware 3D ResNet implemen- tation for COVID-19 chest CT scan classification achieved a sensitivity of 96.7% and specificity of 94.3%, outperforming other contemporary models [26]. Similarly, in the domain of gastrointestinal endoscopy, an attention mechanism-guided SE
+ ResNet-H model achieved 98.12% accuracy for multi-class classification tasks, demonstrating the adaptability of ResNet architectures to specialized medical applications [28].
Beyond healthcare, ResNet variants have shown exceptional performance in diverse domains. Zhou et al. developed an attention-based ResNet for radiation pattern prediction of phased array antennas, achieving a mean absolute error reduction of 41.2% compared to conventional methods [21]. In agricultural applications, Liu et al. implemented a ResNet-based architecture for pest identification, achieving a 97.6% classification accuracy while maintaining robustness against various environmental conditions [31].
The adaptability of ResNet is further evidenced by specialized implementations such as the Fast Fourier Convolutional ResNet (FFC-ResNet) for adenoma dysplasia grading, which achieved a classification accuracy of 93.8% [27]. Additionally, Denize et al. developed ResNeTS, a ResNet adaptation for time series analysis of Sentinel-2 data for grassland biodi- versity prediction, demonstrating the architecture's flexibility beyond conventional image classification tasks [32].
However, the computational complexity of ResNet-50 presents significant challenges for deployment in resource-constrained environments. A comprehensive benchmark analysis by Gupta et al. revealed that ResNet-50 required approxi- mately 3.8 billion floating-point operations (FLOPs) per infer-

ence, nearly 4.7 times higher than MobileNet [33]. This substantial computational overhead translates to increased power consumption and memory requirements, making ResNet-50 less suitable for edge computing devices or mobile applications.

2) EfficientNet-B0: Balancing Performance and Efficiency: In the medical domain, EfficientNet models have demon- strated exceptional performance-efficiency trade-offs. Khan et al. utilized an EfficientNet-enhanced UNet for brain tumor segmentation, achieving a 94.2% Dice coefficient while reduc- ing the model parameters by 73% compared to conventional approaches [12]. Similarly, Alqudah et al. employed a fine- tuned EfficientNet for brain tumor detection in MRI images, achieving 98.1% accuracy with only 5.3 million parameters [11].

The versatility of EfficientNet is further demonstrated by its successful application in diverse domains. Dar et al. developed an Efficient-SwishNet system for facial emotion recognition, achieving 97.8% accuracy while reducing inference time by 41% compared to conventional architectures [4]. In agri- cultural applications, Nair et al. implemented EfficientNet B4 with compound scaling and swish activation for paddy leaf disease classification, achieving 98.5% accuracy while maintaining inference times suitable for field deployment [20]. Comprehensive comparative studies have consistently po- sitioned EfficientNet as an optimal balance between perfor- mance and efficiency. Ahsan et al. compared ResNet, Efficient- Net, and MobileNetV2 for skin condition detection, finding that EfficientNet-B0 achieved 95.7% accuracy while requiring only 52% of the computational resources of ResNet-50 [22]. Additionally, Mondal et al. conducted a comparative study between ResNet and the EfficientNet family for leukemia classification, reporting that EfficientNet-B0 achieved com- parable accuracy to ResNet-50 while requiring 78.4% fewer
parameters [24].

3) MobileNet-V2: Optimized for Edge Computing: The computational efficiency of MobileNet-V2 is particularly evident in mobile applications. Hadi et al. implemented SSD MobileNet V2 FPNLite for website-based Indonesian sign language recognition, achieving real-time performance with 91.3% accuracy on standard smartphone hardware . Similarly, Ardianto et al. developed a mobile-based serious game for learning facial expression recognition in children with autism spectrum disorder using the MobileNet model, achieving in- teractive frame rates while maintaining 89.6% classification accuracy .

In medical imaging, MobileNet variants have demonstrated significant utility in resource-constrained environments. Rahaman et al. utilized a pre-trained MobileNet model for multi-class skin disease detection, achieving 92.7% accuracy while enabling deployment on portable devices for point-of-care diagnostics . Kumar et al. conducted a comparative study of EfficientNet and MobileNet models for lung can- cer classification using chest CT scan images, finding that while MobileNet achieved slightly lower accuracy (91.8% vs. 94.5%), it required 73% less memory and achieved 3.2x faster inference times [25].

The adaptability of MobileNet to specialized tasks is demonstrated by Liu et al., who developed MobileNet-based neural differential distinguishers for cryptographic algorithms, achieving performance comparable to conventional approaches while reducing computational requirements by 67% . Addi- tionally, Zhao et al. explored optimizing hyperparameters in MobileNet v1 image classification using Bayesian and random search, further enhancing the efficiency-accuracy trade-off .

B. Transfer Learning and Training Efficiency

Our transfer learning experiments demonstrated significant improvements in model performance across all architectures:

- Transfer learning provided an average improvement of
8.5 percentage points in mAP@0.5, highlighting its importance in practical object detection applications.
- MobileNet-V2 showed faster convergence during train- ing, reaching its optimal performance in fewer epochs compared to the other models. This can further reduce the training time and computational resources required for deployment.
- The performance gap between models trained from scratch and those with transfer learning was more pronounced for deeper models (ResNet-50 and EfficientNet-B0), suggesting that these architectures benefit more from pre-trained weights.

C. Practical Implications

Based on our findings, we can provide the following recommendations for practical applications:

- For high-accuracy requirements with available computa- tional resources: ResNet-50 is the recommended choice due to its superior detection performance across various object categories.
- For balanced accuracy-efficiency requirements: EfficientNet-B0 offers an excellent compromise, with near-top accuracy and moderate computational demands.
- For resource-constrained environments or real-time appli- cations: MobileNet-V2 is the preferred option, delivering reasonable accuracy with minimal computational require- ments and faster inference time.
- For applications focusing on specific large object categories (e.g., person detection for crowd monitoring): The performance gap between models is smaller, making lightweight models like MobileNet-V2 potentially sufficient.

VI. CONCLUSION

This paper presented a comprehensive comparative study of three prominent deep learning models, EfficientNet, ResNet, and MobileNet, for object detection tasks. Through extensive experiments on standard benchmark datasets, we systematically evaluated their performance across multiple metrics, including detection accuracy, inference time, model size, and computational requirements.

Our findings reveal important trade-offs that practitioners must consider when selecting an appropriate model for spe- cific applications. ResNet-50 demonstrated superior detection accuracy (mAP@0.5 of 58.1% in COCO and 81.2% in Pascal VOC) but required substantially more computational resources (4.1 billion FLOPs) and longer inference time (45 ms/image). EfficientNet-B0 offered an exceptional balance between accu- racy (57.2% mAP @ 0.5 in COCO) and efficiency (0.39 billion FLOPs), making it ideal for applications requiring high preci- sion and reasonable processing speed. MobileNet-V2, while achieving lower accuracy (48.9% mAP @ 0.5 in COCO), excelled in efficiency metrics with the fastest inference time (28 ms/image) and smallest model size (32 MB), positioning it as the optimal choice for resource-constrained environments. The study demonstrated the remarkable impact of transfer learning across all architectures, with an average improvement of 8.5 percentage points in mAP@0.5 compared to training from scratch. We also observed that model selection should consider the specific characteristics of target objects. Deeper architectures performed significantly better for complex and smaller objects, while the performance gap narrowed for larger
and more common objects.

In real-world deployment scenarios, our research under- scores the importance of aligning the selected model with the application requirements. For high-end systems prioritizing accuracy, ResNet is recommended; for balanced performance, EfficientNet provides the optimal trade-off; and for edge devices or real-time applications with limited resources, Mo- bileNet delivers sufficient accuracy with minimal computa- tional demands.

A. Future Work

Several promising directions for future research emerge from this study:

- Investigating more recent model architectures such as Vision Transformers (ViT) and comparing them with CNN-based models for object detection tasks
- Exploring model compression techniques such as quan- tization and pruning to further optimize the models for deployment on resource-constrained devices
- Developing hybrid approaches that combine the strengths of different architectures for improved performance across various object categories
- Extending the comparison to video object detection and tracking scenarios, where temporal information can be leveraged
- Investigating the robustness of these models to domain shifts and adverse conditions (e.g., low light, occlusion, weather effects)

Addressing these research directions will allow us to further advance the field of object detection and develop more efficient and accurate models for real-world applications.

REFERENCES

[1] W. Zheng, Y. Ai, and W. Zhang, ”Hybrid improved concave matching algorithm and ResNet image recognition model,” IEEE Access, vol. 12,
pp. 39847–39861, 2024.

[2] H. Alhichri, A. S. Alswayed, Y. Bazi, N. Ammour, and N. A. Alajlan, ”Classification of remote sensing images using EfficientNet-B3 CNN model with attention,” IEEE Access, vol. 9, pp. 14078–14094, 2021.

[3] F. Sajid, A. R. Javed, A. Basharat, N. Kryvinska, A. Afzal, and M. Rizwan, ”An efficient deep learning framework for distracted driver detection,” IEEE Access, vol. 9, pp. 169270–169280, 2021.

[4] T. Dar, A. Javed, S. Bourouis, H. S. Hussein, and H. Alshazly, ”Efficient-SwishNet based system for facial emotion recognition,” IEEE Access, vol. 10, pp. 71311–71328, 2022.

[5] S. Safwat, A. Mahmoud, I. Eldesouky Fattoh, and F. Ali, ”Hybrid deep learning model based on GAN and ResNet for detecting fake faces,” IEEE Access, vol. 12, pp. 86391–86402, 2024.

[6] M. Majdalawieh, S. Khan, and M. T. Islam, ”Using deep learning model to identify iron chlorosis in plants,” IEEE Access, vol. 11, pp. 46949–46955, 2023.

[7] Y. Hou, H. Shi, N. Chen, Z. Liu, H. Wei, and Q. Han, ”Vision image monitoring on transportation infrastructures: A lightweight transfer learning approach,” IEEE Trans. Intelligent Transportation Systems, vol. 24, no. 11, pp. 12888–12899, 2023.

[8] L. Chen, S. Lin, X. Lu, D. Cao, H. Wu, C. Guo, C. Liu, and F.-Y. Wang, ”Deep neural network based vehicle and pedestrian detection for autonomous driving: A survey,” IEEE Trans. Intelligent Transportation Systems, vol. 22, no. 6, pp. 3234–3246, 2021.

[9] J. Bacha, F. Ullah, J. Khan, A. W. Sardar, and S. Lee, ”A deep learning-based framework for offensive text detection in unstructured data for heterogeneous social media,” IEEE Access, vol. 11, pp. 124484–124498, 2023.

[10] M. Lopez-Montiel, U. Orozco-Rosas, M. Sa´nchez-Adame, K. Picos, and O. H. M. Ross, ”Evaluation method of deep learning-based embedded systems for traffic sign detection,” IEEE Access, vol. 9, pp. 101217–101238, 2021.

[11] A. M. Alqudah, S. Qazan, H. Alquran, I. A. Qasmieh, and A. Alqudah, ”A robust approach for brain tumor detection in magnetic resonance images using finetuned EfficientNet,” IEEE Access, vol. 11,
pp. 24875–24887, 2023.

[12] A. Khan, S. H. Ahmed, M. Imran, and A. Ahmed, ”Deep learning-based MRI brain tumor segmentation with EfficientNet-enhanced UNet,” IEEE
J. Biomed. Health Inform., vol. 27, no. 5, pp. 2148–2159, 2023.

[13] S. A. Alazawi, N. M. Tahir, and R. Ngadiran, ”Enhanced anomaly detection in pandemic surveillance videos: An attention approach with EfficientNet-B0 and CBAM integration,” IEEE Sensors J., vol. 23, no. 10, pp. 9725–9737, 2023.

[14] Z. Li, J. Wang, S. Liu, and Y. Zhang, ”Hybrid transformer-EfficientNet model for robust human activity recognition: The BiTransAct approach,” IEEE Internet Things J., vol. 11, no. 4, pp. 7592–7605, 2024.

[15] M. B. A. Miah, M. A. Yousuf, M. S. Miah, and M. N. Hasan, ”Lightweight EfficientNetB3 model based on depthwise separable convolutions for enhancing classification of leukemia white blood cell images,” IEEE Access, vol. 11, pp. 29362–29379, 2023.

[16] A. Jamil, J. Huang, Y. Niu, and A. Iqbal, ”A synergistic approach to colon cancer detection: Leveraging EfficientNet and NSGA-II for enhanced diagnostic performance,” IEEE J. Biomed. Health Inform., vol. 28, no. 2, pp. 1069–1080, 2024.

[17] J. Ahmad, M. A. Khan, K. Javed, and S. Rubab, ”Automated brain tumor detection from magnetic resonance images using fine-tuned EfficientNet- B4 convolutional neural network,” IEEE Access, vol. 11, pp. 6742–6758, 2023.

[18] A. Rahman, M. M. Rahman, M. F. Mridha, and M. S. Kaiser, ”Ensemble deep learning models for enhanced brain tumor classification by lever- aging ResNet50 and EfficientNet-B7 on high-resolution MRI images,” IEEE Access, vol. 11, pp. 84953–84967, 2023.

[19] M. Kaur, D. Singh, R. Singh, and H. J. Kim, ”Navigating landscapes through AI: A comparative study of EfficientNet and MobileNetV2 in image classification,” IEEE Sens. J., vol. 23, no. 8, pp. 7982–7994, 2023.

[20] S. Nair, P. Mathur, and T. Choudhury, ”Paddy leaf disease classification using EfficientNet B4 with compound scaling and swish activation: A deep learning approach,” IEEE Access, vol. 11, pp. 57286–57299, 2023.

[21] Q. Zhou, Y. Li, and Z. Wu, ”Attention-based ResNet for radiation pattern prediction of phased array antenna,” IEEE Antennas Wireless Propag. Lett., vol. 23, no. 3, pp. 567–571, 2024.

[22] M. Ahsan, R. A. Naqvi, and W. Hussain, ”Comparative analysis of ResNet, EfficientNet, and MobileNetV2 for detecting skin conditions in deep learning models,” IEEE Access, vol. 11, pp. 95417–95431, 2023.

[23] R. Singh, A. Mittal, and R. Kumar, ”Comparative analysis of vehicle insurance fraud detection using EfficientNet, ResNet50, and MobileNet,” IEEE Trans. Intell. Transp. Syst., vol. 24, no. 8, pp. 8256–8269, 2023.
[24] T. Mondal, A. K. Das, and P. Ghosh, ”Comparative study between ResNet and EfficientNet family for classification of leukemia,” IEEE J. Biomed. Health Inform., vol. 27, no. 9, pp. 4284–4295, 2023.
[25] A. Kumar, S. Sharma, and P. Singh, ”Comparative study of EfficientNet and MobileNet models for lung cancer classification using chest CT scan images,” IEEE J. Biomed. Health Inform., vol. 27, no. 8, pp. 3957–3968, 2023.
[26] J. Sun, L. Huang, D. Yang, and J. Chen, ”Region-of-interest aware 3D ResNet for classification of COVID-19 chest computerised tomography scans,” IEEE J. Biomed. Health Inform., vol. 27, no. 2, pp. 795–806, 2023.
[27] Y. Sakai, K. Takayama, and H. Nakamura, ”Adenoma dysplasia grading of colorectal polyps using fast Fourier convolutional ResNet (FFC-ResNet),” IEEE Trans. Med. Imaging, vol. 42, no. 5, pp. 1356–1367, 2023.
[28] A. K. Mishra, D. S. Guru, and S. Shantakumar, ”Attention mechanism guided SE + ResNet-H model for gastrointestinal endoscopy image classification,” IEEE J. Biomed. Health Inform., vol. 27, no. 1, pp. 271–282, 2023.
[29] Y. Liu, M. Li, and Z. Zhang, ”Interior innovation design using ResNet neural network and intelligent human–computer interaction,” IEEE Trans. Industr. Inform., vol. 19, no. 5, pp. 6301–6310, 2023.
[30] D. Wang, X. Liu, and L. Zhang, ”Juggler-ResNet: A flexible and high- speed ResNet optimization method for intrusion detection system in software-defined industrial networks,” IEEE Trans. Network Service Manage., vol. 20, no. 1, pp. 778–790, 2023.
[31] Y. Ma, X. Wang, and J. Li, ”Pest identification based on fusion of self- attention with ResNet,” IEEE/CAA J. Autom. Sinica, vol. 10, no. 4, pp. 990–1002, 2023.
[32] J. Denize, A. Be´gue´, and J. Betbeder, ”ResNeTS: A ResNet for time series analysis of Sentinel-2 data applied to grassland plant-biodiversity prediction,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 16, pp. 1289–1300, 2023.
[33] S. Gupta, A. Bhatia, and R. K. Jha, ”Comparative analysis of ResNet, MobileNet, and EfficientNet models for lung nodule detection and classification,” IEEE J. Biomed. Health Inform., vol. 27, no. 7, pp. 3247–3258, 2023.
[34] L. Zhao, H. Wang, and G. Li, ”A comparative analysis of EfficientNet and MobileNet models’ performance on limited datasets: An example of American sign language alphabet detection,” IEEE Access, vol. 11,
pp. 38962–38975, 2023.
[35] K. Yao, Y. Zhang, and L. Wang, ”Utilizing fMRI and deep learning for the detection of major depressive disorder: A MobileNet V2 approach,” IEEE Trans. Neural Syst. Rehabil. Eng., vol. 31, pp. 1252–1261, 2023.