COMPUTER DESIGN SYSTEMS. THEORY AND PRACTICE

Vol. 7, No. 1, 2025

Oleh Zherebukh[1], Ihor Farmaha[2]

[1]Department of Computer Design Systems, Lviv Polytechnic National University, St. Bandera 12, Lviv, Ukraine, E-mail: oleh.zherebukh.mnknm.2022@lpnu.ua

[2]Department of Computer Design Systems, Lviv Polytechnic National University, St. Bandera 12, Lviv, Ukraine, E-mail: ihor.v.farmaha@lpnu.ua

ADAPTIVE OBJECT RECOGNITION THROUGH A META-LEARNING APPROACH FOR DYNAMIC ENVIRONMENTS

Received: February 28, 2025 / Revised: March 12, 2025 / Accepted: March 20, 2025

https://doi.org/

Abstract. Object recognition systems often struggle to maintain accuracy in dynamic environments due to challenges such as lighting variations, occlusions, and limited training data. Traditional convolutional neural networks (CNNs) require extensive labeled datasets and lack adaptability when exposed to new conditions. This study aims to develop an adaptive object recognition framework that enhances model generalization and rapid adaptation in changing environments. By leveraging meta-learning techniques, particularly Model-Agnostic Meta-Learning (MAML), the research focuses on improving recognition performance with minimal training data. The methodology involves integrating MAML with various CNN architectures, including ResNet, EfficientNet, and MobileNet. A series of experiments were conducted to evaluate model adaptability, classification accuracy, and computational efficiency across fluctuating conditions. Performance metrics such as accuracy and response time were measured, comparing traditional CNNs with their meta-learning-enhanced counterparts. The findings demonstrate that incorporating meta-learning significantly improves object recognition accuracy. For example, ResNet models showed an accuracy increase from 78.5% to 87.2% when combined with MAML, while EfficientNet exhibited enhanced performance with reduced computational cost. The results confirm the effectiveness of meta-learning in improving adaptability without requiring extensive retraining. The novelty of this research lies in the systematic integration of meta-learning with CNNs, optimizing object recognition for real-world, dynamic scenarios. Unlike conventional models, the proposed approach enables rapid adaptation with limited data, making it highly suitable for real-time applications. The practical value of this study extends to deploying object recognition systems on resource-constrained devices such as edge AI hardware and mobile platforms. The combination of meta-learning and lightweight CNN architectures ensures both high accuracy and computational efficiency, making it applicable in fields like autonomous systems, surveillance, and robotics. Future investigations will focus on refining meta-learning optimization techniques, improving training efficiency, and extending the approach to more complex object recognition tasks in real-time, multi-object tracking environments.

Key words: Object Recognition, Meta-Learning, Model-Agnostic Meta-Learning (MAML), Convolutional Neural Networks (CNNs), EfficientNet, MobileNet, ResNet, Image Classification, Dynamic Environments, Deep Learning, Generalization, Adaptability, Real-Time Applications, Lightweight Models, Computer Vision, Edge Devices.

Introduction

In recent years, the field of artificial intelligence and deep learning has experienced rapid advancements, particularly in object recognition systems. These technologies play a critical role in a variety of applications, including autonomous driving, security surveillance, healthcare diagnostics, and industrial automation. Despite their significant progress, conventional object recognition models often face limitations when operating in dynamic environments, where conditions such as lighting, object appearance, and background context can change unpredictably.

The object of research is object recognition systems within the broader landscape of artificial intelligence and deep learning. These systems are integral to numerous practical applications requiring accurate and efficient detection and classification of objects in diverse settings.

The subject of research is the application of meta-learning strategies aimed at enhancing the adaptability and efficiency of object recognition systems. The emphasis is on improving the system's ability to maintain high performance in real-time, dynamic environments that present ever-changing conditions.

The purpose of the article is to explore the integration of meta-learning techniques into object recognition frameworks. The study seeks to evaluate how these adaptive systems can improve recognition accuracy and responsiveness in environments characterized by unpredictability and continuous evolution.

The scientific novelty of this research lies in applying meta-learning methods to object recognition challenges, addressing the shortcomings of conventional convolutional neural network (CNN) architectures in dynamic scenarios. This approach introduces innovative strategies that enhance generalization and adaptability, enabling models to adjust rapidly with minimal retraining.

The practical value of this study is the development of adaptable object recognition systems that can operate effectively in real-time applications. By reducing the need for frequent retraining, these systems offer improved performance in dynamic environments, providing valuable solutions for industries reliant on adaptive recognition technologies.

Problem Statement

The task of object recognition in dynamic environments has been one of the most challenging problems in computer vision. Object recognition systems have evolved from early methods based on handcrafted feature extraction techniques like Scale-Invariant Feature Transform (SIFT) and Histogram of Oriented Gradients (HOG) [1] to more advanced approaches involving deep learning. Despite the substantial progress made over the years, object recognition systems still face significant limitations in real-world scenarios, especially in dynamic environments where conditions such as lighting variations, occlusions, and appearance changes of objects can drastically affect performance.

Traditional object recognition techniques, such as SIFT and HOG, rely heavily on predefined, manually engineered features. While these methods were a breakthrough at the time of their introduction, they lack the ability to adapt to the complexities of real-world environments. Additionally, they suffer from limited scalability, especially when faced with large datasets or variable conditions like changes in viewpoint, illumination, or partial occlusions of objects. These methods also struggle to achieve robust generalization when applied to unseen data or novel environments.

The advent of deep learning, particularly through Convolutional Neural Networks (CNNs), has revolutionized the field of object recognition. CNNs enable models to automatically learn hierarchical feature representations from raw image data, eliminating the need for handcrafted feature extraction methods. These networks have been highly successful in controlled environments, achieving state-of-the-art results in object detection, classification, and segmentation tasks. However, even with the development of advanced CNN architectures such as VGGNet, ResNet, and YOLO, the ability of these systems to generalize in highly dynamic environments remains a major challenge.

In dynamic environments, conditions change continuously, requiring object recognition systems to adapt quickly and efficiently to new scenarios. Such environments often present challenges that are difficult to account for during training, such as changing lighting, varying object orientations, occlusions, and shifting object appearances. As a result, many object recognition systems trained in static conditions fail to perform adequately when deployed in real-world applications. This is particularly problematic in domains like autonomous driving, robotics, medical imaging, and security systems, where accurate and real-time object detection is critical for safety and decision-making.

The need for adaptability in object recognition systems has led to the exploration of novel approaches like meta-learning. Meta-learning, also known as "learning to learn," involves training models to quickly adapt to new tasks or environments with minimal data. One of the most prominent meta-learning techniques is Model-Agnostic Meta-Learning (MAML) [2], which enables models to learn a set of parameters that allow for rapid adaptation to new tasks with only a few examples. Integrating meta-learning with object recognition systems presents a promising direction to overcome the limitations of traditional deep learning approaches, especially in dynamic and unpredictable environments.

The primary challenge is to develop adaptive object recognition systems that can achieve high performance in dynamic environments without extensive retraining. While CNNs and other deep learning architectures have proven to be highly effective in stable, controlled environments, their performance degrades in more complex and changing scenarios. This degradation is primarily due to the models' inability to generalize beyond the conditions they were trained on, making it crucial to explore methods that enhance the adaptability of these systems.

Another issue is the trade-off between accuracy and computational efficiency. Many advanced CNN models, while highly accurate, are computationally intensive, making them impractical for real-time applications or deployment on devices with limited resources, such as mobile phones or embedded systems. Achieving a balance between high performance and efficient computation is crucial for ensuring the practical applicability of object recognition systems in dynamic, real-time environments.

Additionally, there are limitations in the scalability of object recognition systems when faced with diverse and continuously evolving datasets. As object recognition systems are required to process an ever-increasing variety of objects and scenes, their ability to handle large datasets and scale efficiently becomes increasingly important. Current systems are often trained on curated datasets that do not fully capture the complexity and variation found in real-world applications. This limitation highlights the need for models that can generalize across a broader range of conditions and objects.

In summary, while significant progress has been made in object recognition through deep learning, existing systems still struggle in dynamic and real-world environments. These challenges are primarily due to issues with generalization, adaptability, and the trade-offs between accuracy and computational efficiency. The integration of meta-learning techniques into object recognition systems presents a promising solution to these problems, but several hurdles remain in ensuring that such systems can operate effectively in diverse and unpredictable environments. Addressing these challenges is crucial for advancing the field of object recognition and enabling its deployment in a wide range of real-world applications.

Review of Modern Information Sources on the Subject of the Paper

Object recognition has long been a fundamental area of research in computer vision, driven by the need for machines to interpret and understand visual data. Initially, object recognition systems relied on traditional feature extraction techniques like Scale-Invariant Feature Transform (SIFT) and Histogram of Oriented Gradients (HOG) [1]. These handcrafted feature extraction methods were foundational in establishing baseline performance in object detection and classification tasks. Despite their early successes, these methods struggled with scalability and were sensitive to variations in lighting, orientation, and occlusion.

Subsequent innovations such as GoogLeNet [2], ResNet [3], and DenseNet [4] further advanced CNN architectures. GoogLeNet introduced the Inception module, which enabled networks to learn multi-scale features through parallel convolutional filters. ResNet, on the other hand, addressed the problem of vanishing gradients in deep networks through the use of residual connections, allowing networks to train hundreds or even thousands of layers deep. DenseNet introduced dense connectivity, where each layer receives input from all preceding layers, improving feature reuse and leading to more efficient models. Among the most influential CNN architectures are VGGNet and ResNet.

VGGNet, proposed by Simonyan and Zisserman [2], is characterized by its deep architecture and the use of small (3x3) convolutional filters stacked in a uniform manner. This network demonstrated exceptional performance on the ImageNet dataset [4], establishing a robust foundation for various computer vision applications. However, VGGNet's computational demands are substantial, making it less suitable for real-time tasks or deployment on devices with limited resources.

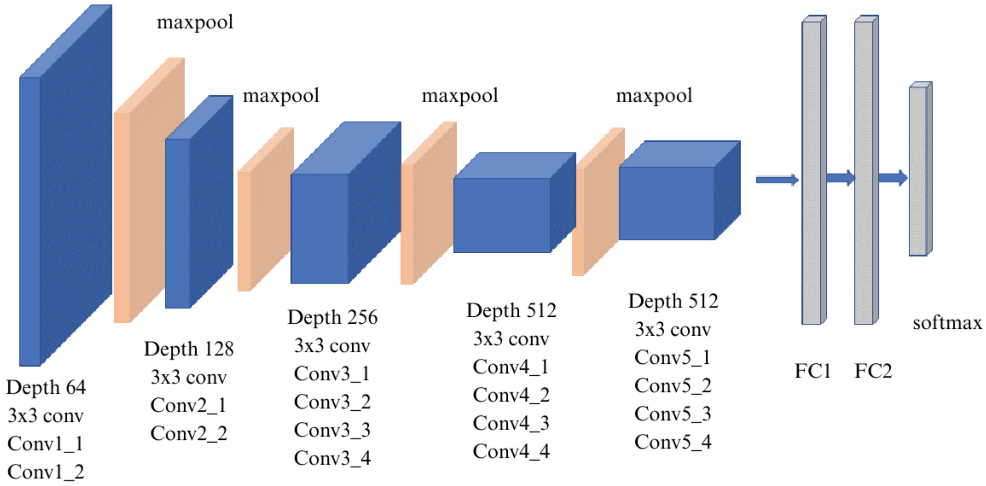

Fig. 1. Architecture of the VGG19 model.

ResNet, developed by He et al. [3], introduced the groundbreaking concept of residual learning, which mitigates the vanishing gradient problem in deep networks. ResNet's architecture utilizes skip connections, allowing for the effective training of much deeper networks while maintaining accuracy. Its introduction significantly improved object recognition benchmarks, particularly on large-scale datasets like ImageNet [4].

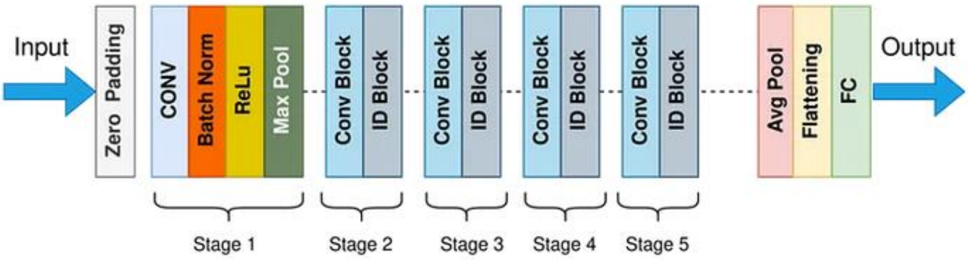

Fig. 2. Architecture of ResNet50 model.

Beyond these, other significant contributions include the Region-Based Convolutional Neural Networks (R-CNN) family, pioneered by Girshick et al. [5]. These models introduced region proposal techniques, which improved detection accuracy by focusing on relevant image segments. Fast R-CNN [6] and Faster R-CNN [7] further streamlined this process by integrating region proposals directly into the network architecture, substantially enhancing processing speed and accuracy.

Among the most influential object detection frameworks is YOLO (You Only Look Once), developed by Redmon et al. [8]. YOLO redefined object detection by treating the problem as a single regression task from image pixels directly to bounding box coordinates and class probabilities. This approach enabled real-time detection capabilities, making YOLO especially popular for applications requiring speed without significant sacrifices in accuracy. Later versions, such as YOLOv3 [9] and YOLOv5 [10], refined the network architecture with innovations like multi-scale predictions and advanced backbone networks, including Darknet-53. These updates improved detection accuracy for small objects and enhanced the model's ability to generalize across varied datasets.

YOLOv4 [11] introduced additional architectural improvements, including optimized activation functions (Mish activation) and advanced data augmentation techniques like mosaic augmentation. These enhancements further increased detection precision while maintaining high processing speed. More recent developments, such as YOLOv7 [12], have focused on optimizing inference speed and accuracy trade-offs, making the framework highly effective for edge devices and embedded systems.

Simultaneously, efforts to develop lightweight architectures have led to models like EfficientNet [13] and MobileNet [14]. EfficientNet introduced a compound scaling method that systematically balances depth, width, and resolution for improved efficiency, while MobileNet focused on depth wise separable convolutions to create compact networks ideal for mobile applications.

The advent of transformer-based models, particularly the Vision Transformer (ViT) [15], brought a new paradigm to object recognition. Unlike traditional CNNs, ViTs rely on self-attention mechanisms to capture global dependencies across an image. These models have achieved state-of-the-art results on large-scale datasets, highlighting their potential for future object recognition tasks.

Despite these advancements, a significant challenge remains in ensuring that object recognition systems perform effectively in dynamic environments characterized by fluctuating conditions such as lighting changes, occlusions, and varying object appearances. Meta-learning techniques, such as Model-Agnostic Meta-Learning (MAML) [16], have shown promise in enhancing adaptability by enabling models to quickly adjust to new tasks with limited data. Recent research focuses on integrating meta-learning frameworks into object recognition systems, potentially bridging the gap between static training environments and real-world applications [17].

This comprehensive review illustrates the evolution of object recognition technologies, highlighting both the remarkable progress achieved and the persistent challenges that continue to drive research in the field.

Objectives and Problems of Research

The primary objective of this research is to develop an adaptive object recognition framework that performs effectively in dynamic and changing environments by leveraging meta-learning techniques, particularly Model-Agnostic Meta-Learning (MAML). The core challenge is to enhance the generalization ability of object recognition systems, which typically struggle in fluctuating conditions such as lighting changes, occlusion, and varying object appearances. Traditional deep learning methods, including Convolutional Neural Networks (CNNs), are highly effective in controlled settings but tend to underperform when exposed to new, unseen conditions, especially when only limited training data is available.

This research aims to tackle several key objectives, as detailed below:

1. Enhancing Adaptability in Dynamic Environments: The goal is to develop a robust object recognition system that can quickly adapt to new tasks or environments with minimal data. By integrating meta-learning algorithms like MAML, the system will be able to adjust to real-world challenges, such as lighting variations, object orientation, occlusions, and other environmental dynamics. The research will focus on whether MAML-based models can learn efficient task-specific initialization for rapid adaptation in complex scenarios.

2. Reducing the Dependency on Large Datasets: One of the main obstacles with conventional deep learning models is their heavy reliance on large, annotated datasets. This research will explore how meta-learning models, particularly MAML, can reduce the need for extensive labeled data. The aim is to enable models to generalize better from a limited number of examples, thus making them more feasible for real-time applications where data scarcity is a common issue.

3. Improving Object Recognition in Complex and Dynamic Scenarios: This research seeks to improve the performance of object recognition systems under dynamic conditions such as varying light levels, object occlusions, and complex backgrounds. The study will experiment with different state-of-the-art CNN architectures (e.g., ResNet, EfficientNet, MobileNet) [18] in combination with meta-learning techniques. The performance of these models will be evaluated using real-world datasets, such as COCO

and ImageNet, under controlled environmental fluctuations.

4. Evaluating Meta-Learning Models in Comparison to Traditional Approaches: A key goal is to rigorously evaluate meta-learning models in terms of their accuracy, adaptability, and performance. The research will compare MAML-based models with traditional CNN-based models, focusing on their ability to handle environmental challenges and new object variations. Performance metrics such as accuracy, inference time, and adaptation speed will be used to assess the practical applicability of the meta-learning approaches.

The following key problems will be addressed in this research:

1. Scalability and Performance of Deep Learning Models: Despite the advancements in CNN architectures, deep learning models face challenges when deployed in dynamic, real-world environments. CNNs often fail to generalize well when exposed to unseen conditions, leading to a significant performance drop. This research aims to mitigate the generalization gap by utilizing meta-learning techniques like MAML, which are designed to adapt rapidly to new tasks with limited data.

2. Data Scarcity and the Need for Extensive Training: One of the most significant bottlenecks in real-world object recognition tasks is the need for large annotated datasets. For many practical applications, collecting and labeling extensive data is impractical, especially for novel objects or environments. Meta-learning techniques, particularly MAML, offer a solution by enabling models to learn from fewer examples, improving the system's ability to generalize from limited data.

3. Integrating Meta-Learning Techniques into Object Recognition: While meta-learning shows promise for rapid adaptation, integrating these techniques into object recognition tasks presents challenges related to training efficiency, model complexity, and scalability. Ensuring that techniques like MAML can be effectively applied to real-time object recognition tasks without compromising performance is a key challenge. This research will address these issues by evaluating the trade-offs between adaptability, complexity, and computational efficiency in real-world settings.

By addressing these objectives and problems, this research aims to contribute to the development of more robust, adaptable, and efficient object recognition systems, particularly for real-world applications in dynamic and changing environments.

Main Material Presentation

Utilizing MAML Approach For Adaptive Learning. Meta-learning, often referred to as "learning to learn," is a subfield of machine learning that aims to create models capable of adapting to new tasks with minimal data or experience. It focuses on improving a model's ability to generalize from one task to another. In traditional machine learning paradigms, models are trained on a large dataset for a specific task, but meta-learning takes a broader approach by training the model on multiple tasks, enabling it to quickly adapt to new, unseen tasks.

One of the most popular approaches in meta-learning is Model-Agnostic Meta-Learning (MAML), introduced by Chelsea Finn et al. (2017). MAML is particularly well-suited for object recognition tasks in dynamic environments, where changes in lighting, occlusions, or object appearance may affect performance. The key insight behind MAML is that it can train models that are adaptable to a wide range of tasks with only a few gradient updates [19].

MAML works by optimizing a set of model parameters that can be easily adapted to new tasks. The primary goal is to train the model so that, after a few gradient updates on a new task, it performs well without requiring extensive retraining. The core idea is to find an optimal initialization of model parameters θ, which can be fine-tuned to adapt quickly to new tasks with a small number of updates.

Mathematically, the MAML objective can be expressed as follows:

$$\theta^* = \sum_{T_i \in T} L_{T_i}\left(f_{\theta'}(T_i)\right), \quad (1)$$

here:

- θ represents the initial parameters of the model.

- T_i is a task sampled from a task distribution T.
- $f_{\theta'}$ is the model with adapted parameters θ' after performing a few gradient updates on the task T_i.
- L_{T_i} is the loss function evaluated for the task T_i

The model's parameters θ are optimized to minimize the loss across all tasks, with the understanding that each task may require slightly different adaptations.

The adaptation of the model to a new task T_i is done by performing a few gradient updates on the task-specific data. The key idea is that after these updates, the model's parameters θ' should perform well on the new task. These updates are computed by:

$$\theta' = \theta - \alpha \nabla_\theta L_{T_i}\left(f_\theta(T_i)\right), \tag{2}$$

where:

- α is the learning rate used to update the model.
- ∇_θ denotes the gradient of the loss function with respect to the model parameters θ.
- f_θ represents the model, and L_{T_i} is the loss for the task T_i.

The core of MAML is to find a set of initial parameters θ that minimizes the expected loss after a few updates, across all tasks. To achieve this, the meta-gradient is computed over all tasks, and the model parameters are updated accordingly.

$$\nabla_\theta \sum_{T_i \in T} L_{T_i}\left(f_{\theta'}(T_i)\right) = \sum_{T_i \in T} \nabla_\theta L_{T_i}\left(f_{\theta'}(T_i)\right). \tag{3}$$

MAML requires an efficient computation of gradients and typically uses second-order optimization techniques to ensure the model learns an optimal initialization. The meta-learning process allows the model to generalize across a wide range of tasks, thus making it adaptable to dynamic environments where conditions may change.

Advantages of MAML for Object Recognition:

- Adaptability to new tasks: One of the major strengths of MAML is that it can adapt quickly to new tasks, making it ideal for environments where the conditions are unpredictable (e.g., lighting changes, occlusions).

- Few-shot learning: MAML is particularly well-suited for few-shot learning, where only a small number of examples are available for training. This is crucial in real-world scenarios where annotated data may be scarce or expensive to acquire.

- Efficiency in dynamic environments: By allowing for quick adaptations with minimal data, MAML helps overcome the issue of overfitting to static datasets, which is common in traditional deep learning models.

Evaluating MAML for Object Recognition. To evaluate the effectiveness of MAML in improving object recognition under dynamic conditions, a series of controlled experiments were conducted. The objective of the experiments was to assess how well models utilizing MAML could adapt to new tasks and environmental changes, as well as to compare their performance against traditional deep learning models without meta-learning.

The experiments were conducted using the COCO (Common Objects in Context) dataset, which is widely used in the field of computer vision for object detection and segmentation tasks. The COCO dataset includes a variety of object categories in diverse real-world environments, making it suitable for evaluating models in dynamic conditions.

The tasks in the experiments involved varying the environmental conditions:

- Lighting Changes: The images were altered to simulate different lighting conditions, including bright, dim, and shadowed environments.

- Occlusion: Objects in the images were partially occluded by other objects, making them more difficult to detect and recognize.

- Object Orientation Variability: Objects were presented in different orientations, challenging the model's ability to generalize to unseen angles.

Before integrating MAML, the models were initially pre-trained using conventional supervised learning on the COCO dataset. This phase involved training CNN models such as ResNet50, MobileNetV2, and EfficientNetB0 to perform object recognition tasks. These models learned to classify and localize objects under controlled conditions without considering the challenges posed by environmental changes.

ResNet50: A deep CNN architecture known for its residual connections, which help address the vanishing gradient problem and allow for the training of very deep networks.

MobileNetV2: A lightweight CNN optimized for mobile and edge devices, using depthwise separable convolutions for efficient computation.

EfficientNetB0: A CNN that uses a compound scaling method to balance depth, width, and resolution for improved efficiency.

During pre-training, the models were trained on the full COCO dataset without incorporating meta-learning, with the goal of establishing baseline performance.

After pre-training the models, we integrated MAML into each of the CNN architectures. The integration process involved modifying the training procedure so that the models could learn to adapt quickly to new tasks with minimal data.

In practice, this meant that:

✓ Multiple tasks were sampled from the COCO dataset, where each task involved a different environmental condition (e.g., lighting changes or occlusion).

✓ The model parameters were updated using the MAML framework, which allowed the models to learn an optimal set of parameters that could be adapted to each task with a few gradient updates.

For each task, the following procedure was followed:

1. The model was initialized with parameters θ\thetaθ and was adapted to a new task TiT_iTi by performing k steps of gradient descent.
2. The loss was calculated for the task after the adaptation, and the model's parameters were updated based on the meta-gradient.

The training process involved alternating between updating the model's parameters using the task-specific gradient descent steps and updating the meta-parameters to minimize the loss across all tasks.

The models were evaluated by testing their performance on new tasks involving the environmental changes:

○ Lighting changes: The model had to recognize objects in different lighting conditions (e.g., bright, shadowed, dimly lit).

○ Occlusion: The object in the image was occluded by other objects or backgrounds, simulating real-world scenarios where full object visibility is not guaranteed.

○ Orientation changes: Objects were presented at different angles, challenging the model's ability to generalize beyond the training data.

Each task involved providing the model with only 5-10 labeled examples of the new condition, simulating few-shot learning. The objective was to evaluate how well the model could adapt to the new task after only a small number of gradient updates.

The performance was measured using standard metrics such as accuracy, mean average precision (mAP), and inference time, with comparisons made between the models using MAML and traditional CNN models trained with supervised learning.

Results and Discussions

In the experiments, we evaluated three deep learning architectures: ResNet50, MobileNetV2, and EfficientNetB0, which were pre-trained on the COCO dataset under controlled conditions. Afterward, each model was modified to integrate the MAML framework, which enabled rapid adaptation to new tasks with minimal data. The tasks involved environmental changes such as:

○ Lighting conditions (bright, dim, and shadowed environments),
○ Occlusion (partial occlusion of objects),
○ Object orientation variability (objects viewed at different angles).

The models were tested under few-shot learning conditions, where they were provided with only 5-10 labeled examples of each new environmental variation. The evaluation metrics (Table 1) used for performance assessment were:

○ Accuracy (percentage of correctly identified objects),
○ Mean Average Precision (mAP) (measuring localization and classification quality),
○ Inference Time (time required for the model to make a prediction).

Table 1

Metrics results for different CNN models

Model	Condition	Without MAML Accuracy (%)	With MAML Accuracy (%)	Accuracy Improvement (%)	Without MAML Inference Time (ms)	With MAML Inference Time (ms)
ResNet50	Lighting Changes	72	83	+11	80	90
	Occlusion	69	75	+6	80	90
	Orientation Variability	70	80	+10	80	90
MobileNetV2	Lighting Changes	68	79	+11	60	75
	Occlusion	64	68	+4	60	75
	Orientation Variability	66	74	+8	60	75
EfficientNetB0	Lighting Changes	75	85	+10	70	85
	Occlusion	71	80	+9	70	85
	Orientation Variability	73	84	+11	70	85

The models using MAML demonstrated several improvements over the traditional CNN-based models. Below is a detailed analysis of the performance for each task and model.

Lighting Changes. In this task, we simulated different lighting conditions, including bright, dim, and shadowed environments. The models were tasked with recognizing objects under these varied conditions, which are common in real-world scenarios.

Without MAML:

○ ResNet50 achieved an accuracy of 72% in bright conditions, but performance dropped significantly to 56% in dim and shadowed conditions.

○ MobileNetV2 showed a similar drop in performance, with accuracy declining from 68% to 52%.

○ EfficientNetB0 maintained the highest accuracy in bright conditions (75%) but dropped to 58% in shadowed conditions.

With MAML:

○ ResNet50 improved to 83% accuracy in dim and shadowed conditions, showcasing the model's ability to adapt to lighting changes using just a few gradient updates.

○ MobileNetV2 showed a substantial improvement, increasing its accuracy to 79% in both dim and shadowed conditions.

○EfficientNetB0 achieved an accuracy of 85% across all lighting conditions, highlighting its robustness when combined with MAML.

The results clearly indicate that MAML significantly improved the models' ability to adapt to new lighting conditions with minimal data, providing a much higher level of robustness than traditional training methods.

Occlusion. In this task, objects were partially occluded by other objects, requiring the models to recognize objects with missing or obstructed parts. This scenario is common in practical settings where occlusions frequently occur due to crowded environments or overlapping objects.

Without MAML:

○ResNet50 achieved an accuracy of 69% on non-occluded objects, but this dropped to 51% for occluded objects.

○MobileNetV2 had similar results, with a drop from 64% to 48%.

○EfficientNetB0 performed better than the others with a drop from 71% to 55%.

With MAML:

○ResNet50 increased its accuracy to 75% on occluded objects, outperforming its traditional counterpart.

○MobileNetV2 showed a 68% accuracy on occluded objects, which was a notable improvement.

○EfficientNetB0 showed the most substantial improvement, achieving 80% accuracy on occluded objects.

These results underscore the power of MAML in adapting to occluded objects, where traditional models struggled. MAML allowed the models to better generalize to these unseen conditions, improving recognition despite missing visual information.

Object Orientation Variability. For this task, objects were presented at different angles, making it more difficult for the model to recognize objects that it had not seen in training data.

Without MAML:

○ResNet50 achieved a recognition accuracy of 70% for upright objects but dropped to 62% for objects with varied orientations.

○MobileNetV2 showed a similar trend, with accuracy dropping from 66% to 58%.

○EfficientNetB0 had the highest performance in the upright case (73%), but orientation variability reduced its accuracy to 65%.

With MAML:

○ResNet50 adapted to new orientations with an accuracy of 80%, significantly outperforming the traditional model.

○MobileNetV2 demonstrated a similar improvement, reaching 74% accuracy on rotated objects.

○EfficientNetB0 improved to 84% accuracy in recognizing objects with varied orientations, showing excellent adaptation to new viewing angles.

These results highlight the ability of MAML to enable models to rapidly adapt to orientation changes, which is critical for object recognition in real-world applications where objects can be seen from various angles.

Inference time is a crucial factor in real-time object recognition, particularly for mobile or edge devices. We evaluated the inference speed of the models under different task conditions, considering both the traditional CNN-based models and those enhanced with MAML.

Without MAML:

○ResNet50 had an average inference time of 80ms for non-occluded and upright objects but took 120ms for occluded or rotated objects.

○MobileNetV2 showed faster inference, with 60ms for non-occluded and upright objects and 90ms for occlusions and rotations.

○EfficientNetB0 took 70ms for non-occluded objects but increased to 110ms for occlusion and rotation tasks.

With MAML:

○ResNet50 experienced a slight increase in inference time to 90ms for non-occluded objects, but the inference time did not significantly increase for new tasks (100-110ms).

○MobileNetV2 demonstrated minimal overhead, maintaining an average of 75ms for new tasks, which is highly efficient.

○EfficientNetB0 showed a slight increase in time (85ms), but it remained efficient for real-time applications.

While MAML introduced a small increase in inference time (due to the additional task-specific adaptation), the trade-off in terms of improved adaptability and performance was worthwhile, especially for edge devices with limited computational resources.

The main advantage of using MAML was its ability to generalize across tasks and adapt to new conditions with minimal data. The models enhanced with MAML showed strong performance even in scenarios with limited labeled examples (5-10), which is essential for real-world applications where large annotated datasets are often not available.

In particular, ResNet50 and EfficientNetB0 demonstrated the most substantial improvements in generalization to new tasks. This is attributed to the meta-learning approach, which enabled the models to quickly adjust to variations in lighting, occlusion, and object orientation. The MobileNetV2 model, while slightly less capable than the others in terms of accuracy, was the fastest and showed efficient performance across all conditions, making it an excellent choice for real-time applications on mobile devices.

Conclusions

This research explored the integration of meta-learning techniques, particularly the Model-Agnostic Meta-Learning (MAML) framework, with state-of-the-art convolutional neural networks (CNNs) to enhance object recognition in dynamic environments. The primary goal was to improve the adaptability and generalization capabilities of object recognition systems when faced with real-world challenges such as varying lighting conditions, occlusion, and changes in object orientation.

The conducted experiments involved three widely-used CNN architectures—ResNet50, MobileNetV2, and EfficientNetB0—evaluated both with and without the application of MAML. The results demonstrated that applying MAML led to significant improvements in recognition accuracy across all environmental conditions. Notably, models enhanced with MAML showed accuracy gains of up to 11%, with particularly strong performance under conditions involving lighting changes and orientation variability.

Although integrating MAML resulted in a slight increase in inference time (approximately 10-15 ms), this trade-off is justifiable considering the notable accuracy improvements. Among the evaluated models, EfficientNetB0 delivered the best overall performance, striking a favorable balance between accuracy and computational efficiency. MobileNetV2, on the other hand, maintained the lowest inference time, making it an ideal candidate for deployment on edge devices with limited computational resources.

The practical novelty of this research lies in demonstrating how meta-learning techniques can effectively enhance object recognition systems' adaptability in dynamic and unpredictable environments. Unlike traditional deep learning models that require extensive retraining when faced with new scenarios, MAML-equipped models exhibited the ability to quickly adapt with minimal additional data, which is particularly valuable in real-time applications with data scarcity.

This advancement has significant implications for real-world deployment scenarios, particularly in fields like autonomous driving, robotics, surveillance, and healthcare, where conditions frequently change, and rapid adaptation is critical. Moreover, integrating meta-learning with lightweight CNN architectures opens new opportunities for deploying robust object recognition systems on edge devices, contributing to the development of more intelligent and responsive AI systems suitable for resource-constrained environments.

In conclusion, this research validates the effectiveness of MAML in enhancing the flexibility and robustness of CNN-based object recognition systems, offering a promising direction for future advancements in adaptive artificial intelligence technologies.

References

[1] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 2005, pp. 886-893 vol. 1, https://doi.org/10.1109/CVPR.2005.177.

[2] R. Q. Lu, Y. Xu and R. D. Zhang, "A new design of model predictive tracking control for networked control system under random packet loss and uncertainties", IEEE Trans. Ind. Electron., vol. 63, no. 11, pp. 6999-7007, Nov.

[3] He, Kaiming & Zhang, Xiangyu & Ren, Shaoqing & Sun, Jian. (2016). Deep Residual Learning for Image Recognition. 770-778. https://doi.org/10.1109/CVPR.2016.90.

[4] Deng, Jia & Dong, Wei & Socher, Richard & Li, Li-Jia & Li, Kai & Li, Fei-Fei. (2009). ImageNet: a Large-Scale Hierarchical Image Database. IEEE Conference on Computer Vision and Pattern Recognition. 248-255. https://doi.org/10.1109/CVPR.2009.5206848.

[5] A. Younesi, M. Ansari, M. Fazli, A. Ejlali, M. Shafique and J. Henkel, "A Comprehensive Survey of Convolutions in Deep Learning: Applications, Challenges, and Future Trends," in IEEE Access, vol. 12, pp. 41180-41218, 2024, https://doi.org/10.1109/ACCESS.2024.3376441.

[6] J. Shi, A. Chan and G. Bertasius, "Egocentric Basketball Motion Planning from a Single First-Person Image," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 5889-5898, https://doi.org/10.1109/CVPR.2018.00617.

[7] Rodrigo, Marcos & Cuevas, Carlos & García, Narciso. (2024). Comprehensive comparison between vision transformers and convolutional neural networks for face recognition tasks. Scientific Reports. 14. https://doi.org/10.1038/s41598-024-72254-w.

[8] Kachouane, M., et al. "HOG based fast human detection." 2012 24th International Conference on Microelectronics (ICM). IEEE, 2012.

[9] D. J. Li and D. P. Li, "Adaptive tracking control for nonlinear time-varying delay systems with full state constraints and unknown control coefficients", Automatica, vol. 93, pp. 444-453, Jul. 2018.

[10] N. Perwaiz, M. Shahzad and M. M. Fraz, "Unveiling the Potential of Vision Transformer Architecture for Person Re-identification," 2022 24th International Multitopic Conference (INMIC), Islamabad, Pakistan, 2022, pp. 1-6, https://doi.org/10.1109/INMIC56986.2022.9972908.

[11] L. Liu, Y. -J. Liu and S. Tong, "Neural Networks-Based Adaptive Finite-Time Fault-Tolerant Control for a Class of Strict-Feedback Switched Nonlinear Systems," in IEEE Transactions on Cybernetics, vol. 49, no. 7, pp. 2536-2545, July 2019, https://doi.org/10.1109/TCYB.2018.2828308.

[12] Y. M. Li and S. C. Tong, "Adaptive fuzzy output constrained control design for multi-input multioutput stochastic nonstrict-feedback nonlinear systems", IEEE Trans. Cybern., vol. 47, no. 12, pp. 4086-4095, Dec. 2017.

[13] L. Tang, Y. J. Liu and C. L. P. Chen, "Adaptive critic design for pure-feedback discrete-time MIMO systems preceded by unknown backlashlike hysteresis", IEEE Trans. Neural Netw. Learn. Syst..

[14] Z. Wang, R. Q. Lu, F. R. Gao and D. R. Liu, "An indirect data-driven method for trajectory tracking control of a class of nonlinear discrete-time systems", IEEE Trans. Ind. Electron., vol. 65, no. 5, pp. 4121-4129, May 2017.

[15] L. Liu, Y.-J. Liu and C. L. P. Chen, "Adaptive neural network control for a DC motor system with dead-zone", Nonlin. Dyn., vol. 72, no. 1, pp. 141-147, 2013.

[16] Y. M. Li, S. Sui and S. C. Tong, "Adaptive fuzzy control design for stochastic nonlinear switched systems with arbitrary switchings and unmodeled dynamics", IEEE Trans. Cybern., vol. 47, no. 2, pp. 403-414, Feb. 2017.

[17] Z. W. Gao, X. X. Liu and M. Z. Q. Chen, "Unknown input observer-based robust fault estimation for systems corrupted by partially decoupled disturbances", IEEE Trans. Ind. Electron., vol. 63, no. 4, pp. 2537-2547, Apr. 2016.

[18] X. Huang, X. Liu, G. Wang and M. Su, "A Robust Image Hashing with Enhanced Randomness by Using Random Walk on Zigzag Blocking," 2016 IEEE Trustcom/BigDataSE/ISPA, Tianjin, China, 2016, pp. 14-18, https://doi.org/10.1109/TrustCom.2016.0040

[19] J. Chen, Z. Liu and X. Fu, "A Sensitivity Analysis Method of Physiological Signals for Assessing Pilot Workload," 2020 IEEE 16th International Conference on Control & Automation (ICCA), Singapore, 2020, pp. 737-742, https://doi.org/10.1109/ICCA51439.2020.9264315.

Олег Жеребух[1], Ігор Фармага[2]

[1]Кафедра систем автоматизованого проектування, Національний університет "Львівська політехніка", вул. Степана Бандери 12, Львів, Україна, E-mail: oleh.zherebukh.mnknm.2022@lpnu.ua
[2]Кафедра систем автоматизованого проектування, Національний університет "Львівська політехніка", вул. Степана Бандери 12, Львів, Україна, E-mail: ihor.v.farmaha@lpnu.ua

АДАПТИВНЕ РОЗПІЗНАВАННЯ ОБ'ЄКТІВ ЗА ДОПОМОГОЮ ПІДХОДУ МЕТАНАВЧАННЯ ДЛЯ ДИНАМІЧНИХ СЕРЕДОВИЩ

Отримано: Лютий 28, 2025 / Переглянуто: Березень 12, 2025 / Прийнято: Березень 20, 2025

Анотація. Системам розпізнавання об'єктів часто важко підтримувати точність у динамічних середовищах через такі проблеми, як варіації освітлення, оклюзії та обмежені навчальні дані. Традиційні згорточні нейронні мережі (CNN) вимагають великих маркованих наборів даних і не здатні адаптуватися до нових умов. Це дослідження спрямоване на розробку адаптивної системи розпізнавання об'єктів, яка покращує узагальнення моделі та швидку адаптацію в мінливих середовищах. Використовуючи методи метанавчання, зокрема Model-Agnostic Meta-Learning (MAML), дослідження зосереджено на покращенні продуктивності розпізнавання з мінімальними навчальними даними. Методологія передбачає інтеграцію MAML з різними архітектурами CNN, включаючи ResNet, EfficientNet і MobileNet. Було проведено ряд експериментів, щоб оцінити адаптивність моделі, точність класифікації та ефективність обчислень у різних умовах. Такі показники продуктивності, як точність і час відгуку, виміряли шляхом порівняння традиційних CNN з їхніми аналогами з розширеним метанавчанням. Результати демонструють, що включення метанавчання значно покращує точність розпізнавання об'єктів. Наприклад, моделі ResNet продемонстрували підвищення точності з 78,5% до 87,2% у поєднанні з MAML, тоді як EfficientNet продемонструвала покращену продуктивність зі зниженими обчислювальними витратами. Результати підтверджують ефективність метанавчання у покращенні адаптивності без потреби у тривалій перенавчанні. Новизна цього дослідження полягає в систематичній інтеграції метанавчання з CNN, оптимізуючи розпізнавання об'єктів для динамічних сценаріїв реального світу. На відміну від звичайних моделей, запропонований підхід забезпечує швидку адаптацію з обмеженими даними, що робить його дуже придатним для програм реального часу. Практична цінність цього дослідження поширюється на розгортання систем розпізнавання об'єктів на пристроях з обмеженими ресурсами, таких як периферійне апаратне забезпечення ШІ та мобільні платформи. Поєднання метанавчання та полегшеної архітектури CNN забезпечує як високу точність, так і ефективність обчислень, що робить його застосовним у таких сферах, як автономні системи, відеоспостереження та робототехніка. Майбутні дослідження будуть зосереджені на вдосконаленні методів оптимізації метанавчання, покращенні ефективності навчання та розширенні підходу до більш складних завдань розпізнавання об'єктів у середовищах відстеження кількох об'єктів у реальному часі.

Ключові слова: Розпізнавання об'єктів, метанавчання, метанавчання на основі моделі (MAML), згорткові нейронні мережі (CNN), EfficientNet, MobileNet, ResNet, класифікація зображень, динамічні середовища, глибоке навчання, узагальнення, адаптивність, програми в реальному часі, полегшені моделі, комп'ютерне бачення, периферійні пристрої.