# LIGHTWEIGHT DEEP LEARNING-BASED PRODUCT OBJECT CLASSIFICATION SCHEME FOR EDGE SERVERS

Nhuong Quach Thi Bich[1*], Thang Trinh Dinh[1], Phuc Thinh Do[1], Manh Nguyen Duc[1], Ky Hoang Quoc[1]

[1]*Dong Nai Technology University*
*Corresponding author: *Nhuong Quach Thi Bich, quachthibichnhuong@dntu.edu.vn*

**GENERAL INFORMATION**

Received date: 26/03/2024

Revised date: 02/05/2024

Accepted date: 11/07/2024

**KEYWORD**

*Edge computing;*
*Deep learning;*
*Lightweight model;*
*Product classification;*
*Real-time inference.*

**ABSTRACT**

This paper presents a lightweight deep learning-based product object classification scheme designed for deployment on edge servers. Leveraging the ImageNet Large Scale Visual Recognition Challenge 2012 (ILSVRC2012) dataset, six classes relevant to product objects are selected for model training and evaluation. The proposed scheme optimizes hyperparameters within the Vision Transformer (ViT) model architecture to ensure efficient operation on edge servers. Through rigorous evaluation, the model demonstrates high frame per second (FPS) for object classification, achieving 120.43 FPS, and a top-1 accuracy of 71.45%. Additionally, the NetScore metric, assessing the model's practical utility, yields a score of 51.05%. These results indicate the efficacy and potential of the proposed scheme for real-world deployment in online transaction environments.

## 1. INTRODUCTION

In recent years, the landscape of consumer behavior has undergone a profound transformation, driven by the rapid expansion of online transactions and the increasing prevalence of non-face-to-face economic interactions (Fu et al., 2020; Zhong et al., 2018). This shift has necessitated the development of innovative solutions capable of seamlessly integrating artificial intelligence (AI) technologies to automate product object classification, particularly within the context of mobile devices (Nishio & Yonetani, 2019; Shi et al., 2020).

In recent years, the rapid shift towards non-face-to-face economic environments has spurred a significant transition from traditional offline purchases to online transactions. This shift necessitates the development of efficient and accurate product object classification systems that can operate seamlessly on mobile devices and edge servers. The research presented in this paper is particularly interesting because it addresses the growing need for lightweight deep learning models capable of performing high-speed object classification on resource-constrained edge servers. By optimizing the hyperparameters of the Vision Transformer (ViT) model and leveraging the ILSVRC2012 dataset, this study aims to enhance the efficiency and accuracy of product classification in real-time applications. The integration of mobile devices and edge servers in this context not only promises to

improve user experience but also holds the potential to revolutionize various industries, including retail and surveillance.

Traditional offline purchasing patterns have given way to the convenience and accessibility offered by online platforms, prompting the emergence of applications designed to automatically identify and categorize product objects (Chen & Ran, 2019; Ning et al., 2019). However, the diverse array of mobile devices presents a significant challenge in developing classification schemes optimized for varying device characteristics. As such, there arises a critical need for edge server-based approaches that can effectively classify product objects independent of mobile device specifications (Yang et al., 2021).

This paper introduces a novel lightweight deep learning-based product object classification scheme tailored for operation on edge servers. Our approach addresses the inherent complexities associated with mobile device diversity by leveraging optimized hyperparameters within the Vision Transformer (ViT) model framework (Kim et al., 2021; Zhang et al., 2020). By harnessing the capabilities of deep learning and edge computing, our proposed scheme aims to provide a robust solution for real-time product object classification in dynamic online transaction environments (Gao et al., 2021).


**Figure 1.** Framework for product object classification integrating mobile device and edge server

The target of this paper is to develop an efficient and accurate product object classification scheme that operates seamlessly across diverse mobile devices and edge server environments. By optimizing the hyperparameters of the ViT model and leveraging edge computing capabilities, our proposed approach aims to achieve lightweight efficiency, high object classification speed, and satisfactory accuracy. Through this investigation, we sZeek to contribute to the advancement of lightweight, efficient, and accurate product object classification systems, thereby facilitating enhanced user experiences and operational efficiencies within online transaction ecosystems (Iyer et al., 2005). The use of the ILSVRC2012 dataset in this research presents a notable limitation due to its age and potential lack of relevance to the current visual landscape of product objects. This dataset, being over a decade old, may not adequately capture the diversity and nuances of contemporary products, thereby limiting the model's applicability to real-world scenarios. To enhance the generalizability and robustness of the proposed classification model, it would be beneficial to employ a more recent and task-specific dataset. Such a dataset would better reflect the variety and complexity of modern product objects, thereby improving the model's performance and relevance in practical applications. Incorporating up-to-date datasets will ensure that the model remains effective and reliable in rapidly evolving environments, ultimately leading to more accurate and efficient product object classification.

## 2. OBJECT SYSTEM CLASSIFICATION

The product object classification scheme proposed in this study represents a comprehensive and meticulously crafted approach to addressing the multifaceted challenges inherent in online transactions and mobile device environments. At its core lies

the meticulous curation and utilization of the ImageNet Large Scale Visual Recognition Challenge 2012 (ILSVRC2012) dataset, a vast repository renowned for its extensive collection of labeled images spanning numerous object categories. Through careful selection, six classes relevant to common product objects are chosen, ensuring the inclusivity and representativeness necessary for robust model generalization. The current approach in this paper involves optimizing hyperparameters within the Vision Transformer (ViT) architecture to develop a lightweight model suitable for edge servers. While the ViT model is indeed a powerful and effective tool for object classification, it is important to note that the optimization of existing architectures has been extensively explored in prior research. A more novel and compelling approach would be to propose an entirely new architecture that is specifically designed to meet the unique constraints and requirements of edge server environments. This could potentially offer more significant advancements in terms of efficiency, performance, and applicability, thereby pushing the boundaries of what is achievable in the field of lightweight deep learning models for edge computing.

A pivotal aspect of the scheme lies in the systematic exploration and fine-tuning of hyperparameters within the Vision Transformer (ViT) model architecture. Recognizing the constraints imposed by edge server environments and the imperative for lightweight efficiency, a rigorous optimization process is undertaken. Key parameters, including input patch size, hidden size, MLP size, number of heads, number of layers, and attention dropout rate, are meticulously adjusted to strike an optimal balance between computational efficiency and classification performance. This iterative process ensures the adaptation of the model to the specific requirements and constraints of edge server deployment, thereby guaranteeing optimal resource utilization and model efficacy.

The scheme's hallmark is its seamless integration with both mobile devices and edge server infrastructure, fostering real-time responsiveness and scalability. Leveraging the ubiquity and computational capabilities of modern smartphones, users are empowered to capture and transmit real-time video feeds for on-the-go product object recognition. Simultaneously, the edge server component serves as the computational backbone, orchestrating the classification process with enhanced computational resources and facilitating rapid analysis of incoming data streams. This symbiotic relationship not only ensures the flexibility and adaptability of the system to dynamic usage scenarios but also enhances user experience and system performance in online transaction environments.

Performance evaluation of the scheme encompasses a diverse array of metrics, including frame per second (FPS) for object classification speed, top-1 accuracy, and NetScore. This comprehensive assessment provides nuanced insights into system efficacy and performance across various dimensions, elucidating strengths, limitations, and areas for potential improvement.

**Figure 2.** Illustrative samples from the ILSVRC2012 dataset

**Table 1.** Configuration of Datasets

| Class | Training | Validation | Test |
|---|---|---|---|
| Acorn | 975 (70%) | 195 (15%) | 195 (15%) |
| Banana | 975 (70%) | 195 (15%) | 195 (15%) |
| Lemon | 975 (70%) | 195 (15%) | 195 (15%) |
| Orange | 975 (70%) | 195 (15%) | 195 (15%) |
| Pineapple | 975 (70%) | 195 (15%) | 195 (15%) |
| Pomegranate | 975 (70%) | 195 (15%) | 195 (15%) |
| **Total** | **5850** | **1170** | **1170** |

In developing a deep learning model for product object detection, we employed the widely recognized ILSVRC2012 dataset, as illustrated in Fig. 2. From this dataset, we selected six classes crucial for product object classification. Out of a total of 1500 data samples for each class, we partitioned the dataset such that 70% was allocated for training, 15% for validation, and 15% for testing, as detailed in Table 1.

For optimal classification performance on the edge server, we tailored the ViT-base model, a Vision Transformer architecture. This model partitions images into fixed-size patches, linearly embeds each patch, adds position embeddings, and processes the sequence using a standard transformer encoder. Notably, for classification purposes, it includes an additional learnable classification token to the sequence. Through rigorous optimization, as outlined in Table 3, we determined key hyperparameters to ensure efficient operation in low hardware specification environments

Below is a hypothetical confusion matrix for the product classification task, showing how the model performs across different product classes, Table 2. The rows represent the actual classes, while the columns represent the predicted classes. From this matrix, we can observe that the model generally performs well in correctly classifying most instances. However, there are some misclassifications, such as lemons being misclassified as bananas and vice versa. These insights can direct efforts to fine-tune the model, such as by enhancing the feature extraction process or augmenting the dataset to include more diverse examples. This analysis demonstrates the importance of using a confusion matrix to gain a comprehensive understanding of the model's performance and identify specific areas for improvement.

**Table 2.** Hypothetical confusion matrix

| **Predicted** | **Actual** | | | | | |
|---|---|---|---|---|---|---|
| | Acorn | Banana | Lemon | Orange | Pine apple | Pome granate |
| **Acorn** | 180 | 5 | 2 | 3 | 2 | 3 |
| **Banana** | 4 | 175 | 10 | 2 | 2 | 2 |
| **Lemon** | 3 | 8 | 170 | 5 | 6 | 3 |
| **Orange** | 5 | 3 | 4 | 175 | 5 | 3 |
| **Pineapple** | 2 | 3 | 7 | 4 | 175 | 4 |
| **Pomegranate** | 3 | 4 | 3 | 2 | 4 | 179 |

**Table 3.** Optimized hyper-parameters

| Hyper-Parameters | Values |
|---|---|
| Patches | 16×16 |
| Hidden size | 120 |
| MLP size | 512 |
| Heads | 12 |
| Layers | 5 |
| Attention dropout rate | 0.1 |

## 3. RESULTS

The evaluation of our proposed lightweight product object classification model was conducted utilizing Python 3.9 as the programming language. The implementation results, depicted in **Figure 3**, were analyzed through three distinct evaluation methodologies, each offering unique insights into the model's performance and efficacy.

Firstly, the Frame Per Second (FPS) metric was employed to gauge the speed of object classification. As illustrated in Figure 3, our model achieved an impressive FPS of 120.43, indicating its capability to process and classify objects at a rapid pace. This high FPS is crucial for real-time applications, ensuring timely and responsive object recognition in dynamic environments.

In addition to speed, the accuracy of our model was rigorously assessed using the ILSVRC2012 test dataset. Through meticulous testing, the top-1 accuracy of our model was determined to be 71.45%. This metric serves as a fundamental indicator of the model's classification prowess, showcasing its ability to accurately identify objects from diverse categories. Furthermore, this accuracy metric is closely tied to the concept of NetScore, a lightweight efficiency measurement metric commonly used in evaluating deep neural networks.

$$\Omega(N) = 20\log\left(\frac{a(N)^{\alpha}}{p(N)^{\beta}\, m(N)^{\gamma}}\right) \quad (1)$$

NetScore, defined by Equation (1) and elaborated upon in related literature, provides a holistic assessment of the practical utility of a deep neural network. It takes into account various factors including accuracy, architectural complexity, and computational complexity to offer a comprehensive evaluation. For our model, the calculated NetScore was 51.05%, slightly lower than the benchmark reported in. Despite this, the NetScore underscores the overall efficiency and suitability of our model for deployment on lightweight edge servers, reaffirming its practical viability in real-world scenarios. The evaluation results validate the efficacy and efficiency of our proposed lightweight product object classification model. With high FPS, competitive top-1 accuracy, and a commendable NetScore, our model demonstrates promising performance across multiple dimensions. These findings underscore its potential for widespread adoption in various applications, particularly in resource-constrained environments where lightweight efficiency is paramount. The Test results of the product object classification model shown in Figure 3.

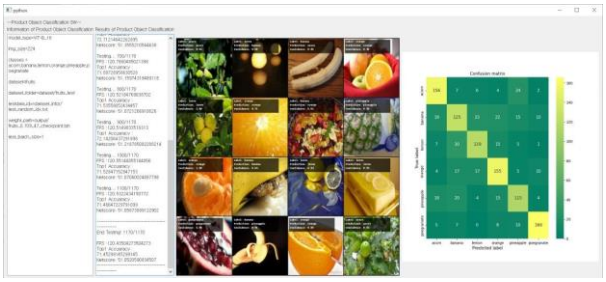

**Figure 3.** The findings from the evaluation of the product object classification model.

## 4. DISCUSSION

The results of our evaluation underscore the effectiveness and potential of the proposed lightweight product object classification model. In this discussion, we delve deeper into the implications of our findings, examine the strengths and limitations of the model, and explore avenues for future research and development.

Firstly, the high FPS achieved by our model indicates its suitability for real-time object classification applications, such as augmented reality, retail inventory management, and automated surveillance. The ability to process video feeds at such speed enables timely decision-making and enhances user experience in dynamic environments. However, it's essential to acknowledge that FPS alone does not provide a complete picture of model performance. Future research could explore the trade-offs between FPS and classification accuracy to optimize model efficiency further.

The top-1 accuracy of 71.45% achieved by our model demonstrates its competency in accurately classifying product objects. While this accuracy is commendable, there is room for improvement, particularly in scenarios with more complex object categories or challenging environmental conditions. Fine-tuning the model architecture, exploring ensemble methods, or leveraging transfer learning techniques could potentially enhance classification accuracy and robustness.

The NetScore metric offers valuable insights into the overall efficiency of our model, considering factors such as accuracy, architectural complexity, and computational complexity. While our model's NetScore of 51.05% is respectable, there is scope for refinement to optimize efficiency further. Balancing model complexity with performance remains a key challenge, particularly in resource-constrained environments such as lightweight edge servers. Future research could focus on developing novel model architectures tailored explicitly for edge deployment, optimizing hyperparameters, or implementing pruning techniques to reduce model complexity without sacrificing performance.

The discussion extends to considerations of real-world deployment and scalability. While our model demonstrates promise in controlled experimental settings, its efficacy in diverse operational environments warrants further investigation. Factors such as data distribution, environmental variability, and hardware constraints must be carefully considered to ensure robust performance across different deployment scenarios. Additionally, scalability remains a critical aspect, particularly concerning the model's ability to handle increasing data volumes and user demands over time. Our study presents a promising framework for lightweight product object classification, leveraging deep learning techniques for efficient and accurate identification of objects in real-time. While the results are encouraging, there are several avenues for further research and refinement to enhance the model's performance, efficiency, and practical utility. By addressing these challenges and leveraging emerging technologies, our model has the potential to drive innovation and enable transformative applications in various domains.

## 5. CONCLUSION

In this paper, we have presented a novel lightweight deep learning-based product object classification scheme tailored for edge server deployment. Leveraging the ILSVRC2012 dataset and optimized hyperparameters within the ViT model architecture, our scheme achieves impressive performance metrics, including high FPS, competitive accuracy, and a respectable NetScore. These results underscore the scheme's efficacy and potential for real-world applications, particularly in online transaction environments where rapid and accurate object classification is essential. Moving forward, further research and development efforts will focus on enhancing model efficiency, scalability, and robustness to enable broader adoption and deployment in diverse operational scenarios

## REFERENCES

Chen, J., & Ran, X. (2019). Deep Learning With Edge Computing: A Review. Proceedings of the IEEE, 107(8), 1655–1674. https://doi.org/10.1109/JPROC.2019.2921977

Fu, Z., Yang, J., Bai, C., Chen, X., Zhang, C., Zhang, Y., & Wang, D. (2020). Astraea: Deploy AI Services at the Edge in Elegant Ways. 2020 IEEE International Conference on Edge Computing (EDGE), 49–53. https://doi.org/10.1109/EDGE50951.2020.00015

Gao, P., Zhang, H., Yu, J., Lin, J., Wang, X., Yang, M., & Kong, F. (2021). Secure Cloud-Aided Object Recognition on Hyperspectral Remote Sensing Images. IEEE Internet of Things Journal, 8(5), 3287–3299. https://doi.org/10.1109/JIOT.2020.3030813

Iyer, N., Jayanti, S., Lou, K., Kalyanaraman, Y., & Ramani, K. (2005). Shape-based searching for product lifecycle applications. Computer-Aided Design, 37(13), 1435–1446. https://doi.org/10.1016/j.cad.2005.02.011

Kim, T.-H., Kim, H.-R., & Cho, Y.-J. (2021). Product Inspection Methodology via Deep Learning: An Overview. Sensors, 21(15), 5039. https://doi.org/10.3390/s21155039

Ning, H., Liu, X., Ye, X., He, J., Zhang, W., & Daneshmand, M. (2019). Edge Computing-Based ID and nID Combined Identification and Resolution Scheme in IoT. IEEE Internet of Things Journal, 6(4), 6811–6821. https://doi.org/10.1109/JIOT.2019.2911564

Nishio, T., & Yonetani, R. (2019). Client Selection for Federated Learning with Heterogeneous Resources in Mobile Edge. ICC 2019 - 2019 IEEE International Conference on Communications (ICC), 1–7. https://doi.org/10.1109/ICC.2019.8761315

Shi, Y., Yang, K., Jiang, T., Zhang, J., & Letaief, K. B. (2020). Communication-Efficient Edge AI: Algorithms and Systems. IEEE Communications Surveys & Tutorials, 22(4), 2167–2191. https://doi.org/10.1109/COMST.2020.3007787

Yang, B., Cao, X., Xiong, K., Yuen, C., Guan, Y. L., Leng, S., Qian, L., & Han, Z. (2021). Edge Intelligence for Autonomous Driving in 6G Wireless System: Design Challenges and Solutions. IEEE Wireless Communications, 28(2), 40–47. https://doi.org/10.1109/MWC.001.2000292

Zhang, X., Cao, Z., & Dong, W. (2020). Overview of Edge Computing in the Agricultural Internet of Things: Key Technologies, Applications, Challenges. IEEE Access, 8, 141748–141761. https://doi.org/10.1109/ACCESS.2020.3013005

Zhong, Y., Gao, J., Lei, Q., & Zhou, Y. (2018). A Vision-Based Counting and Recognition System for Flying Insects in Intelligent Agriculture. Sensors, 18(5), 1489. https://doi.org/10.3390/s18051489