Proceedings of CONF-SEML 2026 Symposium: Learning and Decision Making in Multi Agent Software Systems
DOI: 10.54254/2755-2721/2026.BA33074

Transfer Learning-Enhanced Lightweight CNN for Edge Computing: Real-Time Recognition of Daily Objects on Mobile Devices

Chenxi Hao

Faculty of Innovation Engineering, Macau University of Science And Technology, Macau, China
1240014126@student.must.edu.mo

Abstract. Amid the rapid advancement of edge machine learning, real-time, low-latency image recognition on resource-constrained mobile devices is increasingly in demand for intelligent daily scenarios yet existing research lacks focus on lightweight models for common small daily objects. This study addresses the research gap by exploring efficient recognition of 6 daily objects (Airpods, bottles, lipsticks, etc.) based on edge computing. A 220-image dataset (80% training, 20% testing) was built with preprocessing and augmentation. Lightweight CNN models (MobileNetV1/V2, EfficientNet) were trained via transfer learning on Edge Impulse with a control variable method. Results show that EfficientNet achieved 95.8% test accuracy without overfitting, while MobileNetV2 (160×160) achieved an optimal balance of accuracy (87.5%), speed, and size. Transfer learning resolves small-sample issues, and data augmentation boosts generalization. The optimized model was successfully deployed for real-time mobile recognition.

Keywords: Edge Machine Learning, Transfer Learning, Lightweight CNN, Image Recognition, Mobile Deployment

1. Introduction

Edge computing has emerged as a pivotal trend in mobile intelligent applications, enabling low-latency and energy-efficient image recognition via neural networks [1]. Current research in edge-based image recognition mainly focuses on large-scale object categories or high-performance hardware, while studies on lightweight neural network-driven real-time recognition of daily small objects remain scarce—especially for scenarios requiring deployment on ordinary mobile devices with limited computing resources [2]. This study centers on edge-based real-time recognition of 6 common daily objects. This study sets out to investigate three core research objectives: the construction of a small-sample dataset tailored for edge neural network training, the identification of an optimal lightweight convolutional neural network (CNN) architecture that achieves a balanced trade-off among accuracy, inference speed, and model size for mobile deployment, and the verification of the effectiveness of transfer learning in enhancing neural network recognition performance under small-sample constraints. The research adopts dataset construction, data augmentation, and transfer learning-based neural network training on the Edge Impulse platform,

DOI: 10.54254/2755-2721/2026.BA33074

with a control variable method for performance comparison. This work enriches the application of edge computing and neural networks in daily object recognition, provides a reference for small-sample edge intelligence development, and offers technical support for the future expansion of mobile intelligent scenarios.

2. Literature review

2.1. MobileNetV1

This study adopts MobileNetV1 as the baseline control model to test the baseline performance of lightweight CNNs for edge-based daily object recognition [3]. Its pioneering depthwise separable convolution drastically cuts model parameters and computational cost by decoupling spatial and channel-wise convolution, a core design for resource-constrained edge devices [4]. As a classic lightweight architecture for mobile vision, it sets a reliable performance threshold for evaluating subsequent lightweight models in small-sample recognition tasks.

2.2. MobileNetV2

MobileNetV2 is selected as the core candidate model, optimized for MobileNetV1's flaws [5]. Its inverted residual structure and linear bottleneck layers resolve feature transfer loss and gradient vanishing in shallow networks, enabling better extraction of fine-grained features of daily small objects (e.g., Airpods, lipsticks) and achieving a balance among accuracy, speed, and model size—factors critical for edge mobile deployment [6].

2.3. EfficientNet

EfficientNet serves as the high-performance control model, exploring the upper limit of accuracy of lightweight CNNs [7]. Its compound scaling strategy uniformly scales network width, depth and resolution, achieving a more efficient accuracy-computation trade-off than single-dimension scaling [8]. It avoids overfitting in small-sample datasets and delivers superior feature fitting ability for daily object recognition.

2.4. Transfer learning

This study adopts the "domain adaptation + knowledge transfer" framework to mitigate the issue of poor generalization on small-sample data [9]. By fine-tuning models pre-trained on ImageNet on the self-constructed dataset, transfer learning transfers pre-trained feature extraction knowledge, compensates for insufficient small-sample feature learning and reduces training costs, ensuring experimental validity under small-sample constraints [10].

3. Methodology

3.1. Research problem definition and experimental design logic

The core research problem addressed in this study is how to select and optimize a lightweight neural network architecture that balances recognition accuracy, inference speed, and model size for edge-based daily small-object recognition under small-sample constraints.

DOI: 10.54254/2755-2721/2026.BA33074

To solve this problem, the control variable method was adopted as the core experimental design principle—fixing all training parameters and environmental conditions except the neural network architecture (the independent variable), thereby accurately comparing the performance differences (dependent variables: accuracy, inference time, model size) of different models. Three mainstream lightweight convolutional neural network (CNN) models were selected as the research objects: MobileNetV1, MobileNetV2, and EfficientNet. These models were chosen for their wide application in edge computing, characterized by shallow network layers, fewer parameters, and low computational costs, which align with the limited storage and computing resources of ordinary mobile devices.

The experimental framework was divided into three stages: According to figure 1 and 2, firstly constructing a targeted small-sample dataset to address data scarcity. As shown in Figure 3, the second stage involved training the three models under consistent parameters using transfer learning and data augmentation; According to figure 4, thirdly, verifying and comparing model performance on edge devices to determine the optimal architecture.

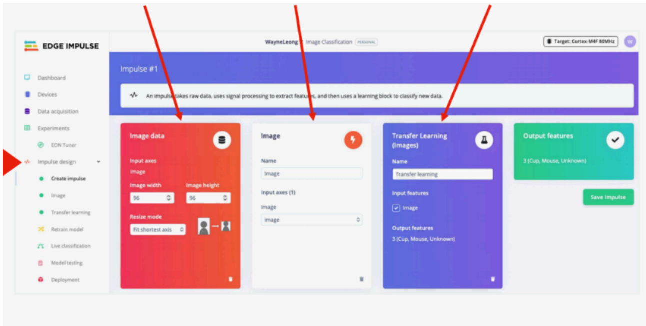

Figure 1. Step1.a

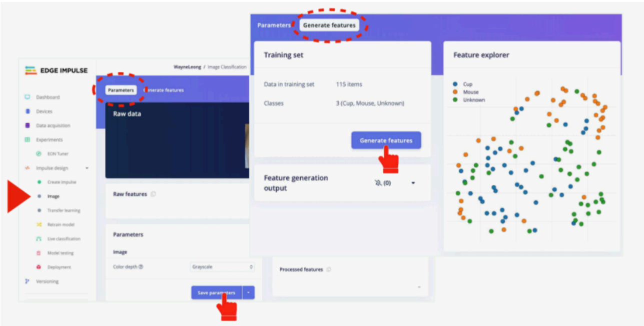

Figure 2. Step1.b

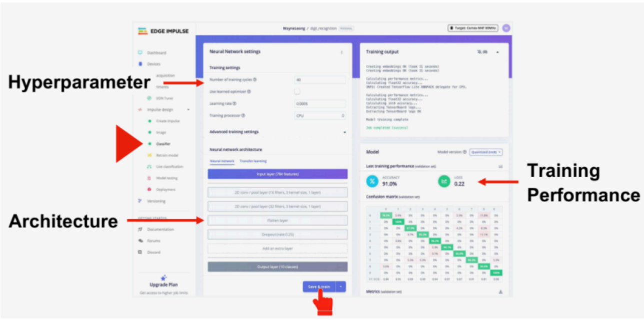

Figure 3. Step 2

DOI: 10.54254/2755-2721/2026.BA33074

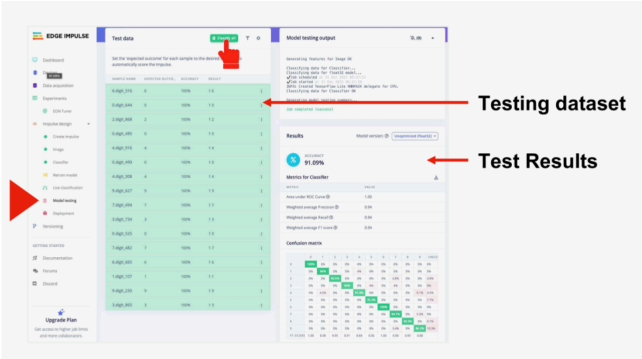

Figure 4. Step 3

3.2. Dataset construction and preprocessing: foundational support for small-sample training

A critical prerequisite for valid experimental comparison is a standardized dataset. To avoid data bias affecting model performance evaluation, a dedicated dataset was constructed comprising six daily small objects.

Data collection: Images were captured using an iPhone 13 (12MP rear camera) under three controlled lighting conditions and three shooting distances, resulting in 220 valid images after excluding blurred samples.

Dataset splitting: Stratified sampling was used to split the data into an 80% training set (176 images) and 20% test set (44 images), ensuring each object accounted for the same proportion in both sets.

Data augmentation and preprocessing: To mitigate small-sample overfitting, the training set was augmented (horizontal flipping, ±15° rotation, ±20% brightness adjustment, 0.8-1.2x zooming) to 528 images. All images were resized to 160×160 pixels, converted to RGB format, and normalized to [0,1] to match model input requirements.

This standardized dataset construction eliminated data-related confounding variables, ensuring that subsequent performance differences between models were attributable solely to architectural differences—thus laying the foundation for the control variable method.

3.3. Model training with consistent parameters: implementing the control variable method

All three models were trained on the Edge Impulse platform with strictly consistent parameters to ensure a fair comparison:

During training, model-specific challenges were addressed without changing core parameters:

first problem:Severe Overfitting: Validation set accuracy (91.7%) vs. test set accuracy (66.7%), leading to extremely poor generalization.

Solutions: Switch to EfficientNet architecture (accuracy increased to 95.8%); raise Dropout rate to 0.3;

enable random flip/rotation data augmentation.

second problem:Class Confusion: High misclassification rate between shape-similar objects

Solutions: The solution was to supplement 20 representative images for confusing categories, after which the misclassification rate decreased significantly.

DOI: 10.54254/2755-2721/2026.BA33074

4. Result

According to the comprehensive analysis presented in Tables 1–3, the performance of different model architectures, three-dimensional trade-offs, and the impact of data quality are systematically evaluated as follows.

First, regarding model architecture, EfficientNet, leveraging its compound scaling strategy, achieved an outstanding test accuracy of 95.8% without any signs of overfitting, demonstrating its superior capability in feature extraction and generalization. This leads to the conclusion that, within the allowable range of computing power, adopting an advanced architecture is the optimal choice for pursuing a high-precision approach to edge AI tasks.

Second, in terms of the three-dimensional trade-off between accuracy, speed, and memory usage, MobileNetV2 stands out by achieving the best balance across these three critical metrics, making it well-suited for deployment on resource-constrained edge devices.This result confirms that there is no absolute optimal solution for edge AI model selection; instead, the choice must be comprehensively weighed according to specific hardware constraints and application scenarios.

Finally, with respect to data quality, the implementation of data augmentation and targeted sampling for confusing categories significantly enhanced the model's generalization ability, while also validating the critical role of color features in the task. This leads to the conclusion that high-quality, diverse datasets deliver far greater performance value than simple parameter tuning does alone, thereby emphasizing the fundamental importance of data preprocessing and curation in model development.

Table 1. Impulse models: architecture and performance summary

NAME	INPUT	DSP BLOCKS	LEARN BLOCKS
Impulse #1	96×96	Image (RGB)	MobileNetV2 96x96 0.35
Impulse#2 ○	96×96	Image (RGB)	EfficientNet
Impulse#3 ○	96×96	Image (RGB)	MobileNetV1 96x96 0.1
Impulse#4 ○	160×160	Image (RGB)	MobileNetV2 96x96 0.35
Impulse#5 ○	96×96	Image (Grayscale)	MobileNetV2 96x96 0.35
Impulse #9	96×96	Image (RGB)	Classifier

Table 2. Network architecture selection vs. performance metrics

LEARN BLOCKS	F32_V_ACC	F32_T_ACC	I8_V_ACC	I8_T_ACC	F3
MobileNetV2 96x96 0.35	91.7%	=	66.7%	=	3
EfficientNet	95.0%	95.8%	95.0%	=	94
MobileNetV1 96x96 0.1	50.0%	12.5%	20.0%	=	2
MobileNetV2 96x96 0.35	95.0%	87.5%	80.0%	=	5
MobileNetV2 96x96 0.35	80.0%	83.3%	80.0%	=	3
Classifier	86.1%	38.6%	86.1%	=	2

DOI: 10.54254/2755-2721/2026.BA33074

Table 3. Multi-dimensional performance overview of impulse series experiments

F32_V_ACC	F32_T_ACC	I8_V_ACC	I8_T_ACC	F32_LATENCY	F32_FLASH	I8_LATENCY
91.7%	=	66.7%	=	3 ms.	1.6M	5 ms.
95.0%	95.8%	95.0%	=	94 ms.	15.4M	143 ms.
50.0%	12.5%	20.0%	=	2 ms.	178.5K	3 ms.
95.0%	87.5%	80.0%	=	5 ms.	1.6M	11 ms.
80.0%	83.3%	80.0%	=	3 ms.	1.6M	5 ms.
86.1%	38.6%	86.1%	=	2 ms.	473.4K	4 ms.

5. Conclusion

Six object recognition systems were successfully constructed, with the EfficientNet model achieving a test accuracy of 95.8%, validating its high precision for the target task. Transfer learning effectively mitigates small-sample edge image recognition challenges and reduces training costs, and MobileNetV2 emerges as the optimal universal choice due to its favorable trade-off between accuracy, speed, and model size, while data augmentation and hyperparameter tuning are critical for enhancing edge model robustness and performance. Future work will include expanding the dataset with more diverse and complex samples to improve generalization, exploring additional lightweight models (e.g., MobileNetV3, ResNet-18) , conducting comparative experiments with advanced deep networks, and deploying optimized models to edge devices such as ESP to validate cross-platform compatibility.

References

[1] Yang, C., Zhang, J., & Li, M. (2022). Edge computing-enabled real-time image recognition for mobile devices: A case study. Journal of Real-Time Image Processing, *19*(4), 897-908.

[2] Kwak J, Jin J, Hu X. Challenges and solutions for small-sample object recognition in mobile edge computing [J]. IEEE Internet of Things Journal, 2023, 10(8): 6689-6698.

[3] Howard, A. G., et al. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv: 1704.04861.

[4] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 1251-1258.

[5] Sandler, M., et al. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. Proceedings of the CVPR, 4510-4520.

[6] Zhang, X., et al. (2020). Lightweight convolutional neural networks for image recognition: A review. Neurocomputing, 395, 32-42.

[7] Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. Proceedings of the ICML, 6105-6114.

[8] Liu, W., et al. (2021). Scaling lightweight CNNs for edge intelligence: A survey. IEEE Internet of Things Journal, 8(18), 13732-13755.

[9] Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 22(10), 1345-1359.

[10] Deng, J., et al. (2009). ImageNet: A large-scale hierarchical image database. Proceedings of the CVPR, 248-255.