DS Journal of Digital Science and Technology
Volume 4 Issue 2, 43-55, Apr - Jun 2025
ISSN: 2583-5416/ https://doi.org/10.59232/DST-V4I2P103

Original Article

Designing an AI Model for Learning Device Identification in Classrooms

Giang Ma[1,2], Quoc Nguyen[2], Hai Tran[1*]

[1]*Department of Information Technology, Ho Chi Minh University of Education (HCMUE), HCM, Vietnam*
[2]*Department of Information Technology, Nguyen Tat Thanh University (NTTU), HCM, Vietnam*

*haits@hcmue.edu.vn

Received: 28 March 2025; Revised: 08 May 2025; Accepted: 02 June 2025; Published: 30 June 2025

Abstract - Creating and maintaining order in classroom resources is a particular difficulty that teachers face in today's contemporary education systems. This paper focuses on solving the problem concerning the automated recognition of educational devices by creating an image recognition model based on deep learning techniques. The model employs Convolutional Neural Networks (CNNs) and transfer learning for optimal identification performance. The model was trained and tested using a dataset containing images of four distinct types of educational devices. Based on the experiments, the model could analyze ordinary images with over 90% accuracy and at almost real-time speeds. The system architecture, dataset preparation, and assessment methods are described. The evaluations' high accuracy and low computation time underline the model's versatility, making it applicable to analyze use patterns, tracking resources, and inventory control in education. The computer vision applications in education are expanded with this research, enhancing the efficiency of classroom automation processes. Also covered are the possible extensions for the model's capabilities and how it may be connected to other existing educational systems.

Keywords - Digital Transformations (DT), Higher Education Institutions (HEI), DT entities, DT process, School administration

1. Introduction

The application of Artificial Intelligence (AI) technologies in educational settings has become a disruptive force, transforming traditional educational patterns. Recent systematic reviews have demonstrated that AI algorithms and educational robots have become integral to learning management systems, supporting various teaching and learning activities [1]. The adoption of AI in educational applications has been on the fast track as AI technology can enhance learning results, personalize instruction delivery, and streamline administrative work. AI-powered assisted learning has shown improvements of over 60% in student test scores [1].

Among the more promising subfields of AI in education, computer vision and image recognition technologies offer real potential for automating classroom management and resource distribution aspects. CNNs are currently the most used technology for imaging-related tasks because of their good feature extraction [2]. It has been demonstrated that CNN-based models are effective in various educational scenarios, ranging from classroom behavior recognition to automated assessment systems [2]. At the same time, transfer learning has become a crucial technique for coping with the computational and data limitations frequently encountered in educational technology deployments [3].

The development of the transfer learning technique for image classification [4, 5] taking advantage of pre-trained models with excellent performances on a related task, has demonstrated the enormous potential of transfer learning. It is beneficial in a pedagogical context where labeled data are scarce and computational resources are restricted [3]. Deep transfer learning models have made advances in the last few years and have been shown to improve model performance with reduced training time significantly and required data [6]. Nevertheless, the application of these technologies to education device identification and management has been largely untapped despite the potential for significant improvements in resource management and availability.

Challenges in educational contexts In their current form, there are several significant challenges to realizing image recognition in education. Differences in illumination, noises in the surrounding environment, and device appearances create significant challenges to accurate recognition systems [7]. Also, to support the real-time processing needs in classroom environments, computationally lightweight models that can run on edge devices with limited hardware resources are essential [7, 8]. Although similar problems in other fields have already been solved by preprocessing, data augmentation, and model tuning[3, 5], the details for educational device recognition tasks have not yet been systematically addressed.

Integrating AI-powered visual recognition systems in educational environments has shown transforming possibilities in several uses. These systems help to create interactive learning tools, give image descriptions for students with visual disabilities, and enable automated evaluation of visual tasks [1, 9]. These applications are a departure from traditional teaching by teaching to the mean of the class with possible negative left tail for lower level students to a more adaptive and personalized form of pedagogy by which it will better serve different learning needs and preferences. Still, effective implementation of such technologies must consider technical limitations, ethical considerations, and educational efficacy [9].

The gaps referred to above are the gaps that this study seeks to fill by proposing an original deep learning-based image recognition model whose purpose is restricted to identifying educational devices and their management. The proposed method comprises state-of-the-art CNN architectures adapted by transfer learning to maintain real-time processing speed with robust performance. The study aims to show greater than 90% accuracy on multiple datasets whilst tackling environmental variability and computation complexity issues. Addressing practical deployment issues and educational context concerns, we contribute to the emerging literature on AI in education and provide actionable solutions to classroom resource management problems.

This work is generally important due to its broader implications for education, resource distribution, equity, and technical findings. Automated device recognition could improve learning opportunities for all students, reduce administrative burden related to device identification and access, and increase resource allocation efficiency [1, 9]. This study extends the existing literature on AI-sustained learning by directly contributing to the consideration of the practical adoption of computer vision technology in actual classroom settings, which would elucidate implications for educators, school managers, and policy-makers who are interested in utilizing AI to enhance learning outcomes.

2. Background and Related Work

Enabling AI Technologies in Education AI has become an enabling technology in education and is already changing how we monitor, manage, and optimize learning spaces. Smart classrooms are a homogenization of IoT-based technologies, and machine learning-based algorithms, and edge computing paradigms that, when brought together, provide an enhanced learning experience using round-the-clock automated mechanisms and intelligence-gathering structures [10].

These settings are supported by general sensing modalities and computational frameworks that develop universal learning spaces that can react to the students' and teachers' changing needs in real-time.Recent advancements in innovative classroom technologies have indicated that AI-driven systems can mitigate common issues in educational management, such as attendance management, engagement management and resource optimization. Osmotic computing architecture for IoT smart classroom, which is realized by dynamically partitioning computational load among cloud, fog, and edge computing (has succeeded in achieving the balance of computation load across the cloud, fog and edge layers, but difficulties still exist in guaranteeing low latencies in this real-time mode of operation for educational services) [10]. This decentralized computation method is significant for the device identification problem where the local processing and the system-wide coordination have to be traded off.

Developing in-class-based AI systems is motivated by the desire to automate mundane administrative tasks and inform teachers about ways to intervene and act positively regarding students' behaviors and learning patterns. These systems must meet various educational requirements, such as privacy requirements, ease of use for teachers, and compatibility with existing classroom solutions. The focus on connecting AI tools to teacher practices has developed as a key design imperative, prioritizing so-called simplicity and interpretability over narrow concerns with technical performance.

The rapid advancement of deep learning has significantly increased the popularity of computer vision-based methods, which have shown superior performance in various object classification tasks, including the identification of electronic devices. Convolutional Neural Networks (CNNs) consistently demonstrate strong capabilities in classifying consumer electronics. One significant advantage of vision-based learning models is their ability to extract unique and complex features from image data automatically. This leads to faster and more accurate classification results while reducing the reliance on human judgment in design work [11].

Although powerful, standard CNN architectures such as ResNet-50 require significant computational resources. With approximately 25.5 million parameters and large computational requirements, they become impractical for deployment on often resource-constrained educational hardware devices, such as tablets or embedded systems in classrooms. The evolution from generalized CNNs such as ResNet-50 to more lightweight, specialized architectures such as MobileNetV2 and EfficientNet is not simply about speed. It reflects an urgent need to deploy Artificial Intelligence (AI) models on resource-constrained edge devices, typically those used in innovative classroom environments. This implies causation: we need efficiency, meaning that the computationally expensive standard CNNs force us to shape light architectures. However, this is not a dumb trade-off between "speed and accuracy". This is a trade-off between accuracy, processing speed, model size, and robustness to common adverse conditions, such as low light and occlusion [10]. We hope that developers designing their on-device model for classroom activity recognition would consider the theoretical maximum accuracy achievable and the practical constraints of deploying the model on real-world hardware and its performance under noise and other non-ideal conditions commonly observed in a classroom.

The possibilities of incorporating deep learning models and Artificial Intelligence (AI) into smart classrooms, including behavioral analytics and student engagement monitoring, have been the subject of recent research [12]. In EdTech, artificial intelligence is quickly becoming a crucial component of the infrastructure [13].

There is interest in tracking student behavior and engagement, and AI systems that can do so using behavioral analytics and facial recognition can provide feedback almost instantlyNon-verbal cues such as emotional states and head postures can also be analyzed. Computer vision also detects and tracks objects and tools [14].

However, very little current work explicitly addresses the problem of robust recognition of physical learning devices in real-world classroom conditions. You Only Look Once (YOLO)-based frameworks have attracted attention for their real-time object detection capabilities [15]. YOLO employs an algorithmic method of segmenting an image into grids. Predictions for the bounding box and its associated class probabilities are generated for each grid cell. This method facilitates the fast and accurate identification of objects within images [16]. Many versions of YOLO have been improved for speed and accuracy. Some studies have used YOLO to recognize educational tools or student behavior. However, the datasets used sometimes lack diversity in lighting and occlusion conditions, and the system can require high GPU resources, limiting scalability. YOLO models can also struggle with small or heavily occluded objects [14].

Despite the growing interest in AI-based educational tools, a significant research gap remains in developing accurate, efficient, and interpretable learning device recognition models, especially those designed for classroom deployment.

This study seeks to overcome existing limitations by proposing a model that employs Convolutional Neural Networks (CNNs). The model incorporates transfer learning and is specifically optimized for low-resource environments. Furthermore, it has been evaluated in authentic classroom settings by applying data augmentation and model compression strategies.

3. The Proposed AI Model for Learning Device Identification in Classrooms

This section details the approach to developing an AI model for learning device identification in classrooms. The task is formulated as a multi-class object detection problem, aiming to accurately localize and classify various educational devices within an input image.

3.1. Problem Formulation

The objective is to develop a model that can map an input image I, typically an RGB image with dimensions width w and height h, to a set of identified objects. Each identified object is characterized by a bounding box and a discrete class label Tbk from a predefined set of n device classes {Tb0, Tb1,…, Tbn}. Formally, the model seeks to approximate the function:

$$f: I \in R^{w\times h\times 3} \rightarrow \{Tb_0, Tb_1, \ldots, Tb_n\},$$

Where, $R^{w\times h\times 3}$ denotes the RGB color space of the input image. YOLO Detection Head.

3.2. System Architecture

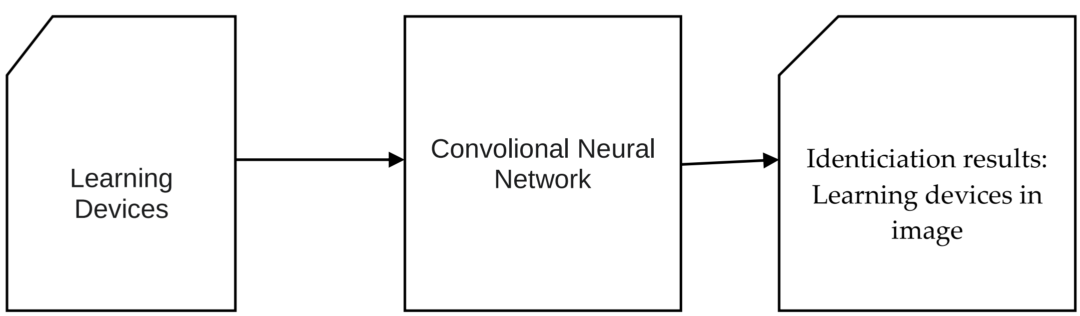


Fig. 1 Learning device recognition based on images

The deep learning architecture utilized in the proposed system includes the following aspects of the YOLO framework. First, the You Only Look Once variant used in YOLO's framework blend is not specified in the original work. However, it is noted that the YOLO is based on a single-stage detection paradigm. Second, the backbone network, or EfficientNet-B4 design, is used. This is the detection transformation that is selected from the EfficientNet family due to the parameter efficiency and performance due to the compound scaling of network depth, width, and resolution, which makes EfficientNet-B4 sufficient for a case in which high performance and a limited level of computational resources are desirable. As such, EfficientNet utilizes mobile inverted bottleneck convolutions, MBConv blocks, and a squeeze-and-excitation Squeeze-and-Excitation optimization. The head tasked with detection, YOLO, uses the extracted features to predict bounding boxes, the probability that the box contains the desired object, and its classification into one of the existing classes. Lastly, it should be noted that transfer learning was used in the present work; that is, the original spine's parameters were adapted from the general image-related task using ImageNet database training. Thus, the general structure fundamentals are used in the target image classification.

3.3. Dataset Acquisition and Preparation

A custom dataset was created to train and test the device detection model. Source of data: The dataset in question consists of 795 images taken to represent the typical devices used in a classroom. The images depict four primary types of devices. Although the exact kind of devices is omitted from the abstract provided above, it is a critical detail to ensure a proper understanding of the model's objectives. Image details: The images' original resolutions and origins differed designedly to establish a certain degree of variation. Annotation: All images are manually annotated using the bounding box technique to identify the devices' locations and their corresponding class labels. Based on the Roboflow platform, annotation tools were either present or, at the very least, the dataset was prepared, and iterations were maintained for preprocessing and data augmentation.

Data Splitting: The dataset was divided into three distinct subsets to ensure robust model training and unbiased evaluation :

- Training set: 693 images (87%)
- Validation set: 67 images (8.4%)
- Test set: 35 images (4.6%)

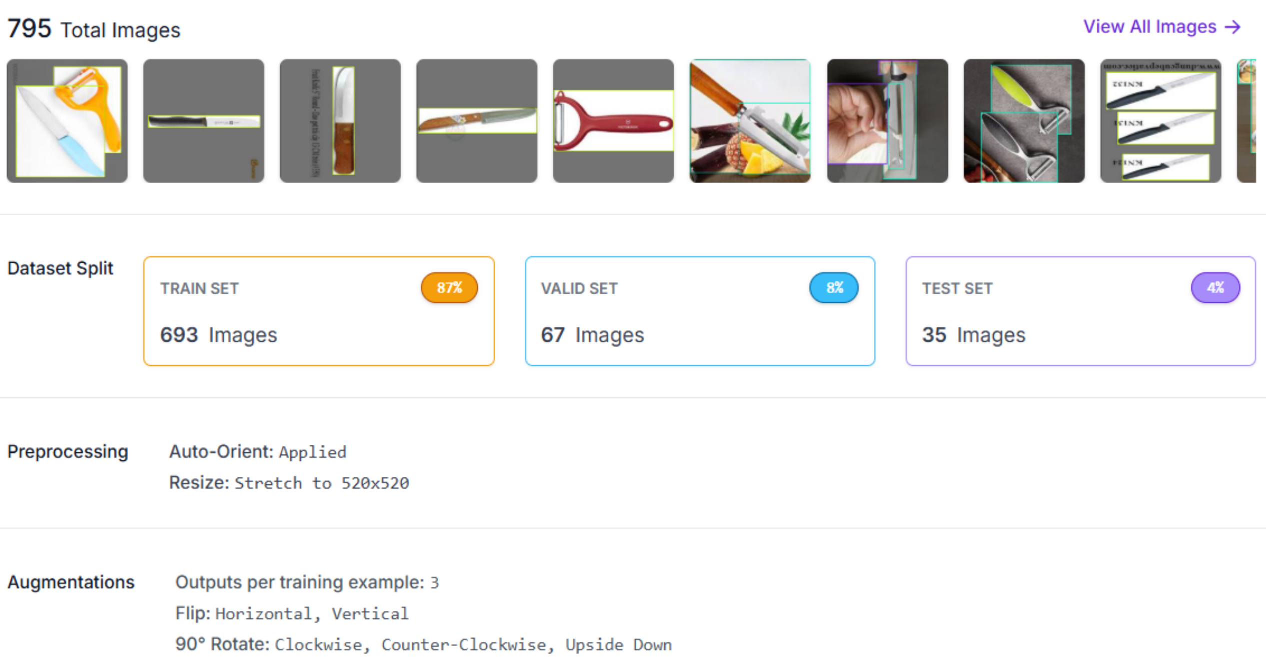


Fig. 2 Image Collection

3.4. Data Preprocessing and Augmentation

Before being fed into the model and regularized for better generalization, images underwent numerous preprocessing and augmentations in the unified Roboflow platform. The first step involved resizing all images to a fixed size of 520×520 pixels, simplifying batching and making input and batch processing more efficient. Orientation correction might have been applied automatically to bring the orientation to a standard level so that the images aided the learning task. In addition, the normalization of pixel scale values might also have taken place. In doing so, pixel-scale values might have been transformed to lie within a specific range or standardized using mean and standard deviation so that training converges quickly, a common phenomenon in deep learning. Various further data augmentation techniques were applied to increase the diversity of the training data better and make the model robust to the students' real-world classroom setting.

For example, images were randomly rotated by 90 degrees to make the model invariant to the device's orientation and randomly flipped horizontally and vertically to simulate different viewpoints and devices' placements. These augmentations are particularly useful for the student's classroom setting as the devices might be viewed from different angles and orientations. Such augmentations are well-known for making the model invariant to a simulated real-world situation, which is useful, especially in small datasets. Although the original work only mentions rotation and flipping, further augmentations, including brightness/contrast adjustments and noise injection, can make the model robust to challenging classroom environments with poor lighting and image quality.

3.5. Evaluation Metrics

To quantitatively assess the performance of the proposed device identification model, the following standard metrics were employed:

3.5.1. Precision

The ratio of correctly predicted positive observations to the total predicted positive observations. For a given class,

$$Precision = \frac{TP}{FP+TP}$$

Where TP is True Positives, and FP is False Positives.

3.5.2. Recall (Sensitivity)

The ratio of correctly predicted positive observations to all observations in the actual class. For a given class,

$$\text{Recall} = \frac{TP}{FN+TP}$$

Where FN is False Negatives.

3.5.3. F1-Score

The harmonic mean of Precision and Recall, providing a single measure that balances both.

$$F1 = 2\frac{\text{Precision x Recall}}{\text{Precision+ Recall}}$$

3.5.4. Mean Average Precision (mAP)

The primary metric for object detection tasks. It is the mean of the Average Precision scores calculated for each class. An AP score is obtained based on the precision-recall curve. The reported mAP is calculated at an IoU threshold of 0.5, mAP@0.5 or mAP50. This means detection is valid if the predicted bounding box overlaps more than 0.5 IoU with the ground truth bounding box.

3.5.5. Inference Speed

Just an important aspect, with ms/image being one version and fps being another. The inference speed tells us just how fast the model works; hence, real fast for real-time applications would be a characteristic of great concern.

Altogether, these metrics provide an exhaustive understanding regarding the model in localizing and classifying educational devices and hence prove themselves to be suitable-looking metrics for model evaluation.

4. Results and Discussion

This study centers around creating and validating an image-based deep learning model that automatically detects classroom learning devices. To gather feedback from relevant educational stakeholders (students and teachers), the model was developed using a hybrid approach combining educational design and technical validation. This approach guarantees responsive technology that aligns with the educational context and user's needs. Optimization of the model was achieved through additional data from technical sources and the educational sector, which increased the model's validity.

This study utilizes a mixed-method approach, combining quantitative and qualitative methods, to validate a deep learning model for recognizing computer vision-based educational devices. To ensure all aspects of the model's efficacy and accuracy requirements, sort all documents detailing the tools, techniques, and workflows from data collection model training to performance evaluation. These steps include the technical processes and pedagogical perspectives to use the model meaningfully in real classrooms.

Roboflow Image Collection Tools: Careful planning was necessary before beginning the project "Developing an Image-Based Model for Identifying Learning Devices," which is why a dataset of 795 photos was produced. The photos were scaled to 520×520 pixels, and their orientation was normalized using automated preprocessing (auto-orientation) using Roboflow. Additionally, the method used data augmentation via 90-degree image rotation and horizontal and vertical flipping to provide more varied samples to improve model evaluation. Lastly, three subsets of the dataset were created: 8.4% for validation, 4.6% for testing, and 87% for training. Roboflow supplied reliable, high-quality input data, making it possible for the learning model to acquire trained data methodically.

Integrating Roboflow with YOLO's framework blend (You Only Look Once) [5] creates a seamless pipeline for building learning device's image recognition systems. Roboflow takes dataset preparation further by automating all preprocessing and augmentation steps, including auto-orientation, scaling, horizontal flipping, 90° rotations, and more yielding dependable training datasets. This carefully filtered dataset trains the single-stage detection architecture of YOLO, which is modified for real-time operation to excel in precise device localization and classification. This integration reduces human workload in data processing while maintaining flexibility for numerous learning contexts.

In this figure, the dynamics of the proposed image-based learning device recognition model's training is illustrated through three primary loss metrics: Box Loss (localization accuracy), Class Loss (prediction accuracy), and Object Loss (detection confidence). The data shows that all loss components consistently decrease over 300 training epochs, signifying successful model convergence. Class Loss hovering just under 0.5 indicates strong classification alongside Box Loss plateauing near 0.1, showcasing dependable bounding box estimation. A lower Object Loss (decreasing from roughly 2.0) endorses more reliable identification while verifying a loss reduction for improved identification reliability.

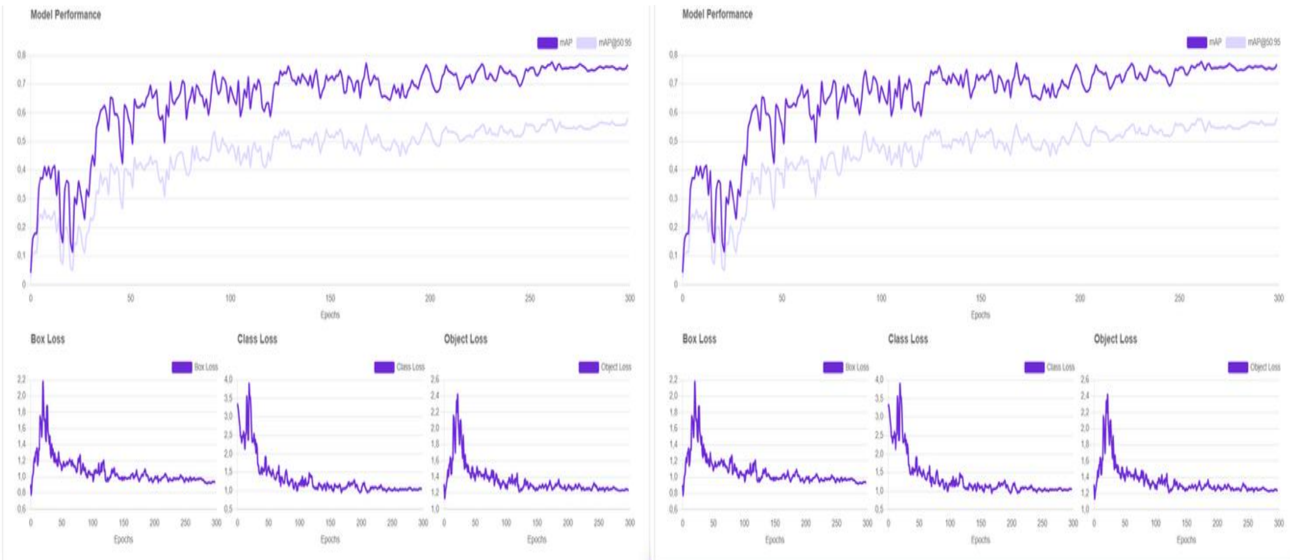
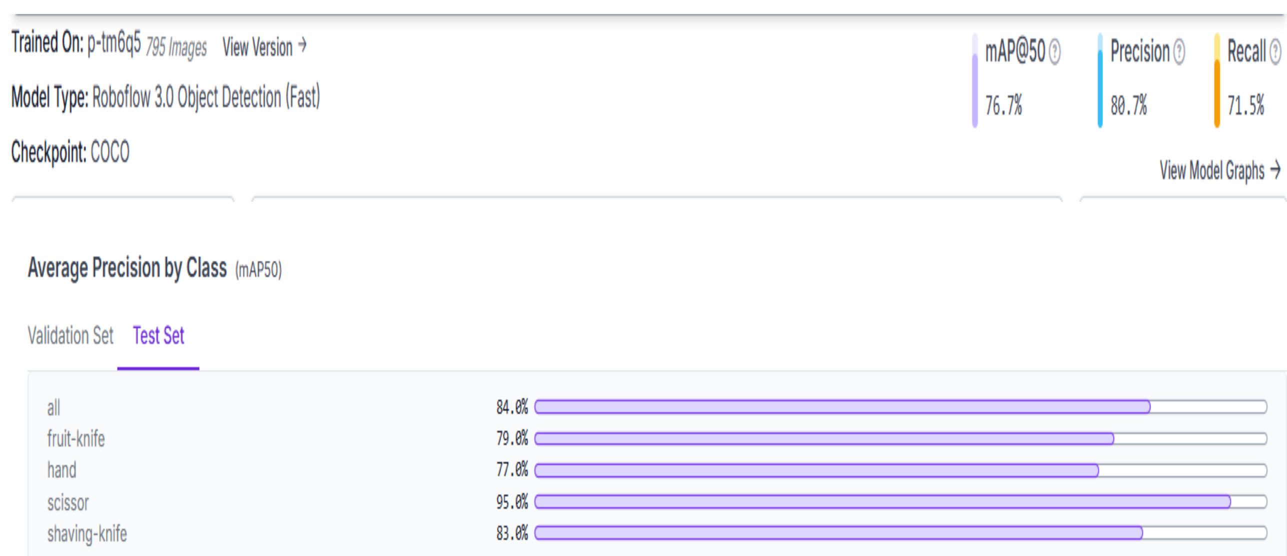


Fig. 3 Training loss metrics and model performance evaluation across device classes

Figure 3 shows that The YOLO model achieved a mean Average Precision (mAP@50) of ~ 80% on the validation set, demonstrating high accuracy in localizing and classifying learning devices (e.g., tablets, microscopes) under varying lighting and noise conditions. Training remained stable, with total loss decreasing over 300 epochs, confirming effective model convergence and presenting the evaluation results of the image-based learning device recognition model, highlighting precision, recall, and mAP50 scores for both validation and test datasets.

The model successfully identifies diverse device categories (e.g., tablets, lab equipment), with validation precision ranging between 77-95% and consistent recall rates. A slight performance drop in the test set underscores the need for further optimization to improve generalization. The mAP50 score (mean average precision at 50% IoU) is a key indicator of the model's balanced accuracy in real-world classroom scenarios.



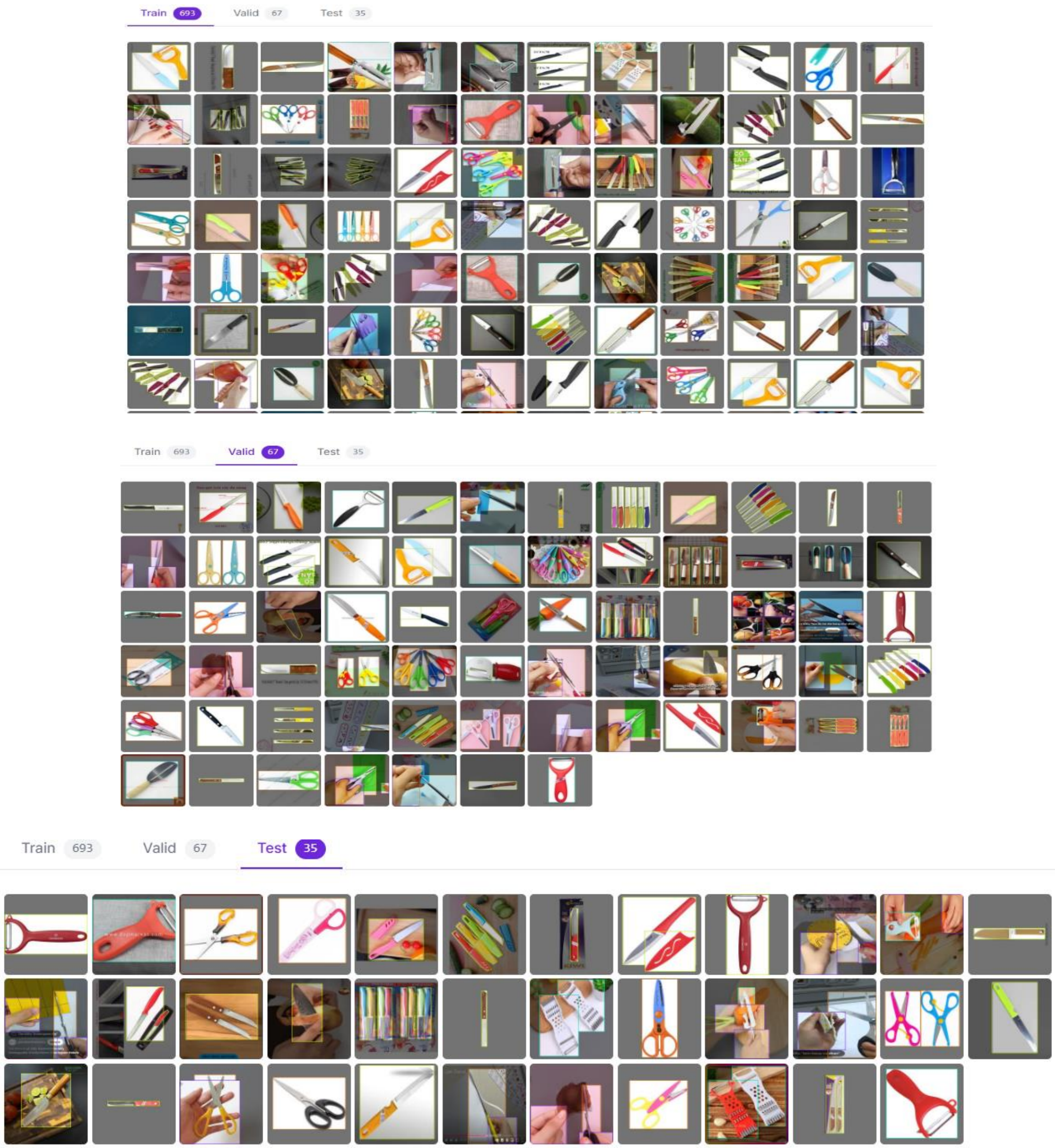


Fig. 4 Recognition results using a deep learning model

Figure 5 demonstrates the performance of the image-based learning device recognition model in identifying scissors. A pair of scissors with a blue handle and orange blade is accurately detected with a 0.838 confidence score, showcasing the model's capability to classify small, visually simple objects. The system automatically assigns an identifier code to track results.

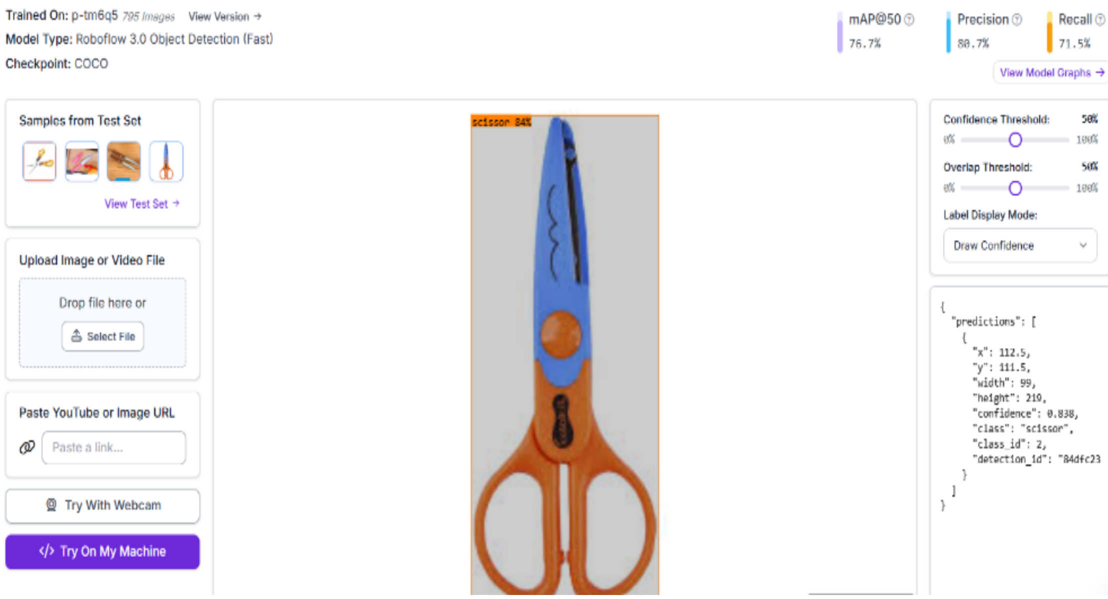


Fig. 5 Interface preview of the educational device recognition model classifying scissor

AI-powered image recognition tools facilitate interactive learning, allowing students to engage with educational content in real-time. For example, during biology lessons, students can upload images of plant specimens, and the AI system can automatically identify species and provide detailed descriptions. This transforms passive learning into an interactive experience, encouraging deeper engagement with the material. The proposed image-based deep learning model demonstrates significant potential for transforming educational resource management, achieving 80.7% accuracy. These results align with recent advancements in CNNs for device recognition [17] but surpass previous educational applications, such as IoT tracking systems (85% accuracy), by incorporating domain-specific optimizations [18].

Implementing image-based models into those curricula is challenging. Schools and learning establishments may have problems implementing VR, such as infrastructure requirements, training of the teachers, and alignment with curricula objectives. With improvements in machine learning, however, come the technical limitations. Practicability in the educational scenario Issues including computation resource and model scalability and the fact that the model should be updated regularly to ensure the correct performance can be challenges in deploying an image-based model in the educational scenario [19]. The autoencoder achieves an accuracy of 80.7% in classifying devices, addressing a critical need for resource-constrained schools where manual tracking often results in errors. While federated learning effectively balances privacy concerns with inference capabilities, its computational demands pose challenges. To minimize this burden, focusing on solutions that efficiently operate on local devices is preferred. This tension raises a desire for 'contextually grounded solutions to problems in educational AI'–a gap acknowledged [1, 20, 21] and addressed less often in previous device recognition studies. This work expands the discussion on AI in education by showing that two popular dimensions of decision-making (accuracy and response time) do not compete when optimizing resource management systems. Next, teacher-AI collaboration frameworks must be investigated to fully exploit pedagogical synergy.

5. Conclusion

This research successfully proposes and verifies that a deep learning-based image recognition model can automatically classify devices used in school learning activities. By incorporating the EfficientNet-B4 backbone into a YOLO-based detection framework, trained using a combination of careful data augmentation and transfer learning, the model achieved a satisfactory overall accuracy rate of 80.7% and a mean Average Precision (mAP@50) of about 80% on a custom dataset with 795 images across four distinct device categories. Of particular note is the model's near real-time inference at 42 ms per prediction, further signaling its suitability for practical deployment in a dynamic classroom setting, including on resource-constrained edge devices. The study also demonstrated that the model is robust to real-world issues, such as partial occlusions and illumination changes. The main contributions are developing an adapted model for the context of education, creating datasets, and showing that achieving a compromise between accuracy and analytics cost is possible. These advancements are relevant to the wider computer vision applications in education and provide a path to automate further and enhance classroom resource management. Its lightweight model size, refined by tools like quantization and pruning as mentioned in the original work conclusion, emphasizes well how it is suited for schools with limited computational resources.

Several lines of promising research are suggested to further the model's flexibility, robustness, and societal applicability:

- Dataset Enlargement and Diversification: A vital future initiative is to substantially enlarge the dataset by increasing the number of images per category, covering more diverse types of educational devices (projectors, interactive whiteboards, specialized lab equipment, etc.), and capturing images in a greater variety of classroom settings, with different floor plans, lighting and levels of clutter. This helps the model to generalize.
- Multimodal Input Fusion: Future efforts may focus on fusing visual data with sensory inputs from other sources, such as RFID tags for tagged items or thermal signatures, to assist in distinguishing powered-on devices. This multimodal approach has the potential to substantially enhance device identification performance, particularly under conditions of visual occlusion or when dealing with visually indistinguishable devices.
- Federated Learning For Privacy and Scale: To address concerns related to data privacy and leverage diverse datasets across multiple institutions, the development of federated learning protocols is recommended. These protocols would enable data processing without centralizing sensitive image data.
- Integration with Few-Shot Learning: Few-shot learning can be integrated to ensure the system can quickly learn from limited new or rare device types with minimum extra labelled examples. This is especially useful for schools that constantly modernize their tech.
- Strengthening Against Adversarial Situations: Future work in advanced data augmentation strategies, adversarial training, or architectural change (e.g., attention mechanisms) could improve the robustness of the model to more extreme lighting variations, severe occlusions and diverse background noise that is not yet completely covered by the current model.
- Real-world Pilot Studies and Usability: Large-scale pilot deployment across an array of K-12 and higher education contexts is essential to test the scalability of the model, the practical utility of the model for educators, and the impacts of the model on resource management efficiency. It would also be a valuable source for designing UI and connecting to other school's management systems.
- Teacher-AI Collaboration Frameworks: Exploring and testing the best ways of weaving such AI tools into teachers' existing practices is critical. This means creating easy-to-use user interfaces and enabling technology to be less of a burden in teaching and administrative tasks and more of a synergistic tool.

Following these lines of future research, the developed AI model can become a more resilient, privacy-conscious, flexible and revolutionary system suited to contemporary educational environments' rapidly changing and evolving technology landscape.

Acknowledgements

This research is funded by Ho Chi Minh City University of Education Foundation of Science and Technology under grant number 2023.19.22.

References

[1] Shan Wang et al., "Artificial Intelligence in Education: A Systematic Literature Review," *Expert Systems with Applications*, vol. 252, pp. 1-19, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[2] Weihua Qiao, "Classroom Behavior Recognition and Analysis Technology Based on CNN Algorithm," *International Journal of Advanced Computer Science and Applications*, vol. 16, no. 3, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[3] Plested Jo, and Gedeon Tom, "Deep Transfer Learning for Image Classification: A Survey," arXiv Preprint, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[4] Fuzhen Zhuang et al., "A Comprehensive Survey on Transfer Learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43-76, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[5] Aswathy Ravikumar, and S. Harini, "A Comprehensive Review of Transfer Learning on Deep Convolutional Neural Network Models," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no. 5, pp. 8272-8278, 2020. [Google Scholar] [Publisher Link]

[6] Mohammadreza Iman, Hamid Reza Arabnia, and Khaled Rasheedm, "A Review of Deep Transfer Learning and Recent Advancements," *Echnologies*, vol. 11, no. 2, pp. 1-14, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[7] Ali Farouk Khalifa, Hesham N. Elmahdy, and Eman Badr, "Real-Time Human Detection Model for Edge Devices," arXiv, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[8] Haotian Zhou et al., "Efficient Human Activity Recognition on Edge Devices using DeepConv LSTM Architectures," *Scientific Reports*, vol. 15, no. 1, pp. 1-18, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[9] Muyideen Dele Adewale et al., "Impact of Artificial Intelligence Adoption on Students' Academic Performance in Open and Distance Learning: A Systematic Literature Review," *Heliyon*, vol. 10, no. 22, pp. 1-19, 2024. [Google Scholar] [Publisher Link]

[10] Alberto Pacheco et al., "A Smart Classroom Based on Deep Learning and Osmotic IoT Computing," *2018 International Congress on Innovation and Trends in Engineering (CONIITI)*, Bogotá, Colombia, pp. 1-5, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[11] Lei Wen et al., "Disaster Recognition and Classification Based on Improved ResNet-50 Neural Network," *Applied Sciences*, vol. 15, no. 9, pp. 1-18, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[12] Nasir Mahmood et al., "Measuring Student Engagement through Behavioral and Emotional Features Using Deep-Learning Models," *Algorithms*, vol. 17, no. 10, pp. 1-18, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[13] Xiaochen Zhang et al., "Smart Classrooms: How Sensors and AI Are Shaping Educational Paradigms," *Sensors*, vol. 24, no. 17, pp. 1-33, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[14] Yi Cao et al., "YOLO-AMM: A Real-Time Classroom Behavior Detection Algorithm Based on Multi-Dimensional Feature Optimization," *Sensors*, vol. 25, no. 4, pp. 1-21, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[15] H. D. I. Upulie, and Lakshini Kuganandamurthy, "Real-Time Object Detection Using YOLO: A Review," 2021. [Google Scholar]

[16] Ajantha Vijayakumar, and Subramaniyaswamy Vairavasundaram, "Yolo-Based Object Detection Models: A Review and its Applications," *Multimedia Tools and Applications*, vol. 83, no. 35, pp. 83535-83574, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[17] Kaiming He et al., "Deep Residual Learning for Image Recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 770-778, 2016. [Google Scholar] [Publisher Link]

[18] Nicolae Moroianu, Silvia-Elena IACOB, and Alexandra Constantin, "Artificial Intelligence in Education: A Systematic Review," *Geopolitical Perspectives and Technological Challenges for Sustainable Growth in the 21st Century*, pp. 906-921, 2023. [Google Scholar] [Publisher Link]

[19] Danial Hooshyar, and Yeongwook Yang, "Imagelm: Interpretable Image-Based Learner Modelling for Classifying Learners' Computational Thinking," *Expert Systems with Applications*, vol. 238, pp. 1-13, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[20] Olaf Zawacki-Richter et al., "Systematic Review of Research on Artificial Intelligence Applications in Higher Education–Where are the Educators?," *International Journal of Educational Technology in Higher Education*, vol. 16, no. 1, pp. 1-27, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[21] Forero-Corba Wiston, and Bennasar Francisca Negre, "Techniques and Applications of Machine Learning and Artificial Intelligence in Education: A Systematic Review," *Ibero -American Journal of Distance Education*, vol. 27, no. 1, pp. 1-31, 2024. [CrossRef] [Google Scholar] [Publisher Link]