Research and Application of Convolutional Neural Networks for Smartphone Recognition

Võ Hồng Sơn
Thai Nguyen University of Information and Communication Technology
Thai Nguyen University, Thai Nguyen, Vietnam
student657@ictu.edu.vn

***Abstract*—This bachelor thesis investigates the use of Convolutional Neural Networks (CNN) for the detection of smartphones in images. The study adopts the YOLOv8 architecture, fine-tuned on the publicly available Cell Phone Detection v2 dataset. Data preprocessing involved labeling with LabelImg and Makesense, followed by transfer learning from a pre-trained YOLOv8 model. Training was conducted on a Tesla T4 GPU with 15 GB VRAM, using 100 epochs, a batch size of 16, and an image resolution of 640 × 640 pixels, completing in approximately 1.7 hours. Evaluation on the validation split yielded a mean Average Precision at 50 % (mAP50) of 0.52 and a mean Average Precision across 50-95 % (mAP50-95) of 0.338. The confusion matrix indicates a 51 % correct classification rate for the smartphone class. The results demonstrate the feasibility of YOLOv8 for smartphone detection while highlighting the impact of limited, low-quality data on performance.**

***Index Terms*—Convolutional Neural Networks, YOLO, Smartphone Recognition, Transfer Learning, Object Detection, Deep Learning**

I. Introduction

Smartphones have become ubiquitous, serving as communication, entertainment, and productivity tools. Accurate detection of these devices in images is valuable for security, inventory management, and user-experience personalization. This thesis aims to develop a CNN-based detector that can reliably identify smartphones in varied visual contexts. The work builds on foundational concepts of deep learning and CNN architecture, introduces the YOLOv8 object-detection framework, and outlines the data preparation and labeling workflow. By integrating transfer learning, the study seeks to mitigate the challenges posed by a small, heterogeneous dataset while evaluating the model's performance on standard metrics [1]. Earlier ICTU student work on related designs is included here for context [2], [3], [4], [5], [6], [7], [8]. Published technical books provide background for the methods discussed in this summary [9], [10], [11]. These contextual works were selected for this English summary; they are not asserted to appear in the original thesis bibliography.

II. Methods

The dataset used is the Cell Phone Detection v2 collection, comprising images annotated with bounding boxes for smartphone instances. Labels were created using LabelImg and Makesense, ensuring consistency across the training and validation splits. The YOLOv8 architecture was selected for its balance of speed and accuracy; it contains 22 layers, including convolutional, C2F, Upsample, Concat, SPPF, and Detect modules. Training employed a pre-trained YOLOv8s backbone, fine-tuned with 100 epochs, batch size 16, and image size 640 × 640. The process ran on a Tesla T4 GPU (15 GB VRAM), completing in roughly 1.7 hours. The report describes transfer learning and the YOLOv8s training configuration; it does not document a separate layer-freezing experiment [1].

III. Results

After training, the model achieved a mean Average Precision at 50 % (mAP50) of 0.52 and a mean Average Precision across 50-95 % (mAP50-95) of 0.338 on the validation set. The confusion matrix shows that the smartphone class was correctly identified 51 % of the time, as reported for the smartphone label in the thesis. Training loss curves indicate a steady decline, while precision and recall metrics improved over epochs. These figures confirm that YOLOv8 can detect smartphones with moderate success, though performance is constrained by dataset size and image quality [1].

IV. Discussion

The reported mAP and accuracy values demonstrate that YOLOv8, when fine-tuned on a limited smartphone dataset, can achieve reasonable detection performance. However, the 51 % correct classification rate highlights the influence of data scarcity and low-resolution images on model robustness. The report does not present a controlled comparison against training from scratch, so the specific gain from transfer learning cannot be isolated. Its confusion matrix still shows substantial misclassification. Future work could expand the dataset, apply data augmentation, and experiment with newer YOLO variants or hybrid architectures to enhance precision and recall [1].

V. Conclusion

The thesis successfully implements a CNN-based smartphone detector using YOLOv8, achieving mAP50 of 0.52 and a 51 % classification accuracy. The report documents a model training exercise and its evaluation figures, without demonstrating deployment in an operational setting. While the current model demonstrates feasibility, its performance

Supervisor: Nguyễn Đình Dũng, Thai Nguyen University of Information and Communication Technology (nddung@ictu.edu.vn).

is limited by the small, low-quality dataset. Planned extensions include enlarging the training set, refining labeling, and exploring advanced training strategies to improve detection reliability in real-world scenarios [1].

References

[1] Võ Hồng Sơn, ”Research and Application of Convolutional Neural Networks for Smartphone Recognition,” internal undergraduate thesis (original in Vietnamese), Thai Nguyen University of Information and Communication Technology, 2024.

[2] ICTU Student 51, ”Character Recognition and Automated Grading of Multiple-Choice Tests,” internal undergraduate thesis, Thai Nguyen University of Information and Communication Technology, 2008.

[3] ICTU Student 44, ”Hopfield Neural Network and Its Application,” internal undergraduate thesis, Thai Nguyen University of Information and Communication Technology, 2008.

[4] ICTU Student 93, ”Artificial Neural Network and Fuzzy Logic,” internal undergraduate thesis, Thai Nguyen University of Information and Communication Technology, 2009.

[5] ICTU Student 177, ”Hand Gesture Recognition in Human-Computer Interaction,” internal undergraduate thesis, Thai Nguyen University of Information and Communication Technology, 2010.

[6] ICTU Student 178, ”Hand Gesture Recognition for Program Control in Video,” internal undergraduate thesis, Thai Nguyen University of Information and Communication Technology, 2010.

[7] Nguyễn Minh Tuấn, ”Construction of a Face Recognition Program for Family Robots,” internal undergraduate thesis, Thai Nguyen University of Information and Communication Technology, 2021.

[8] Nguyễn Thị Hồng, ”Construction of a Plant Species Recognition Program in Vietnam,” internal undergraduate thesis, Thai Nguyen University of Information and Communication Technology, 2023.

[9] F. Chollet, Deep Learning with Python, 2nd ed. Manning Publications, 2021. ISBN 9781617296864.

[10] A. Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 3rd ed. O’Reilly Media, 2022. ISBN 9781098125967.

[11] E. Stevens, L. Antiga, and T. Viehmann, Deep Learning with PyTorch. Manning Publications, 2020. ISBN 9781617295263.