Using CNN to Detect the Use of Mobile Phone in Examination Rooms

Reuben Moyo [1] Stanley Ndebvu [2] Michael Zimba [3] Jimmy Mbelwa [1]

[1]University of Dar es Salaam, Tanzania [2]Mzuzu University, Malawi [3]Malawi University of Science & Technology, Malawi

Motivation & Objective

Goal

To enable the detection of mobile phones as prohibited objects in an exam to assist a physical invigilator in continuously monitoring the students.

Motivation

The presence of security cameras in an exam room could significantly assist in monitoring the use of mobile phones.

Problem

- **Over-reliance on the invigilator** to detect and monitor all the students
- **Long recorded videos** to be effectively analysed by an invigilator after an exam as it is tedious and rigorous.
- **Easily miss a metric** during the analysis because of fatigue

Solution Design & Setup

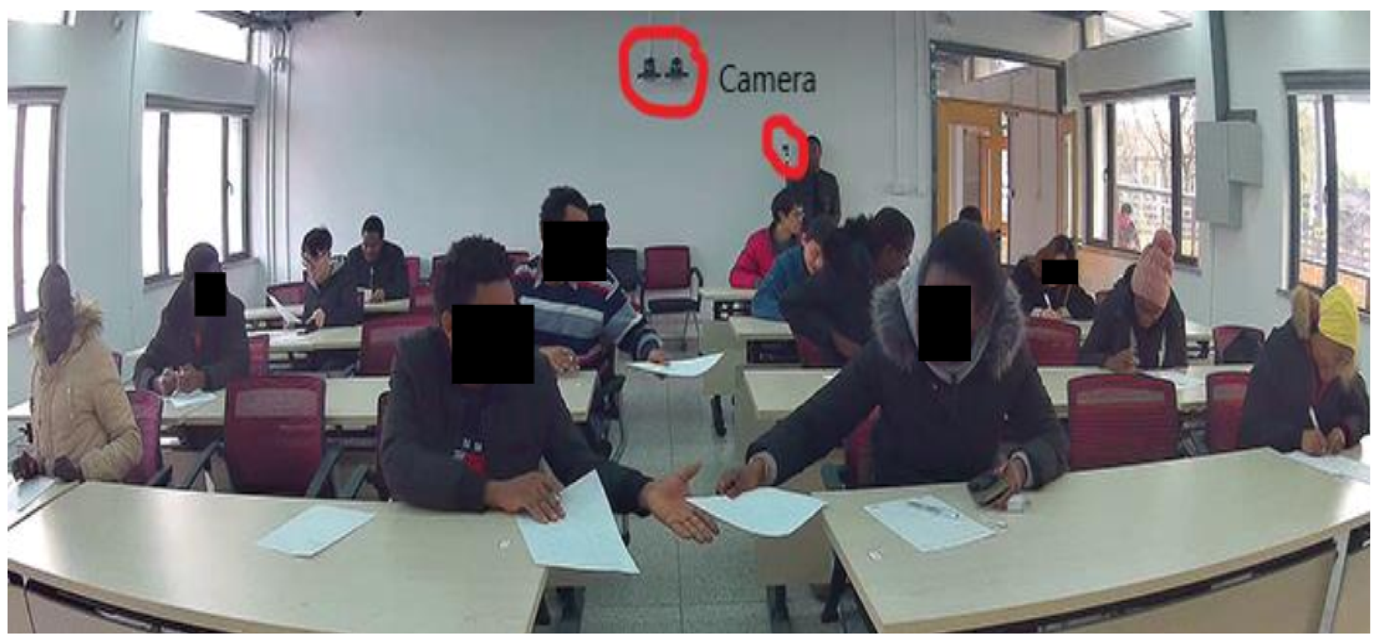


Figure 1. Student in an exam room with purposely placed cameras.

The Idea

Cameras were purposely and strategically fitted to monitor the activities of students taking an exam. The aim was to capture the presence of mobile phones as prohibited materials in the exam room.

10-minute videos were sent to the server for processing:

- **Image preprocessing**: sampling and resizing the width and height of the images as specified by the segmentation and the localisation model. Image pre-processing is very critical as it improves the performance and accuracy of the network
- **Segmentation and localization of objects**: marking the objects with the information we are interested in, like phone and person.With our Instance Segmentation, every person and every phone class object are identified and segmented from each other.
- **Object Classifier Fine-tuning**: employed a custom MMDetection library with Faster R-CNN. This CNN framework was pre-trained on the state-of-the-art COCO and Pascal VOC-Style dataset. However, the pre-trained network performed poorly on the videos because of the angle of capture for the phone and the small size of the phone object. During transfer learning, we add new FC layers to train on new data as we modify FC classifier layers. We replaced the last FC layer with a layer that will classify only two classes; Phone or No-Phone

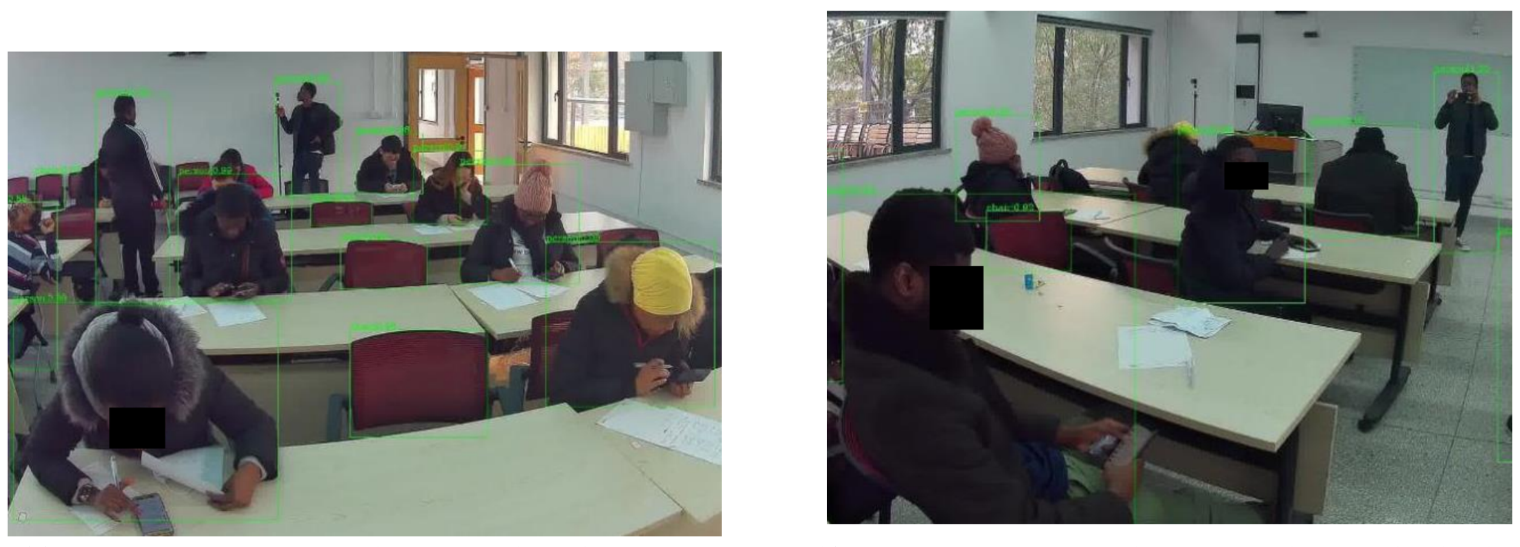
Dataset Creation

Since no existing dataset depicted people using phones in an exam room, we had to create one. To do this:

- we recorded videos of a simulated exam room setting. Sixteen students were randomly chosen to participate in a video recording experiment.
- Additional images were collected from the web
- 10,000 images were annotated using the open-source image and video labeller tool, OpenLabeling

Results and Discussion

Results from the generic model



(a) A front camera view of students using their phones in an exam room. The classification model missed and did not classify the cell phone. It classified the other objects.

(b) A Top-rear view of a student using a phone under the desk. However, the generic model did not classify the cell phone

Figure 2

The generic model performed poorly in the following ways:

- all objects in the 21 classes of the VOC dataset being classified
- high bounding-box classification loss
- Mean Average Precision (mAP) 0.057
- recall 0.368
- 800 misclassified detections out of 19 ground truths.

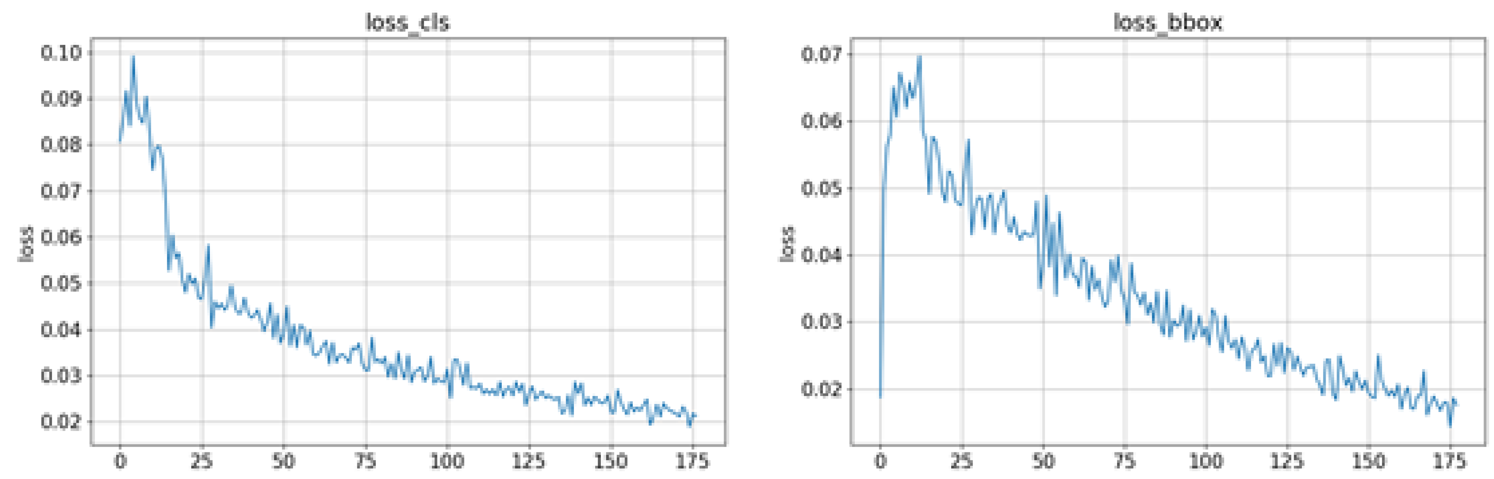
Results and Discussion Cont'd

Results from the fine-tuned model

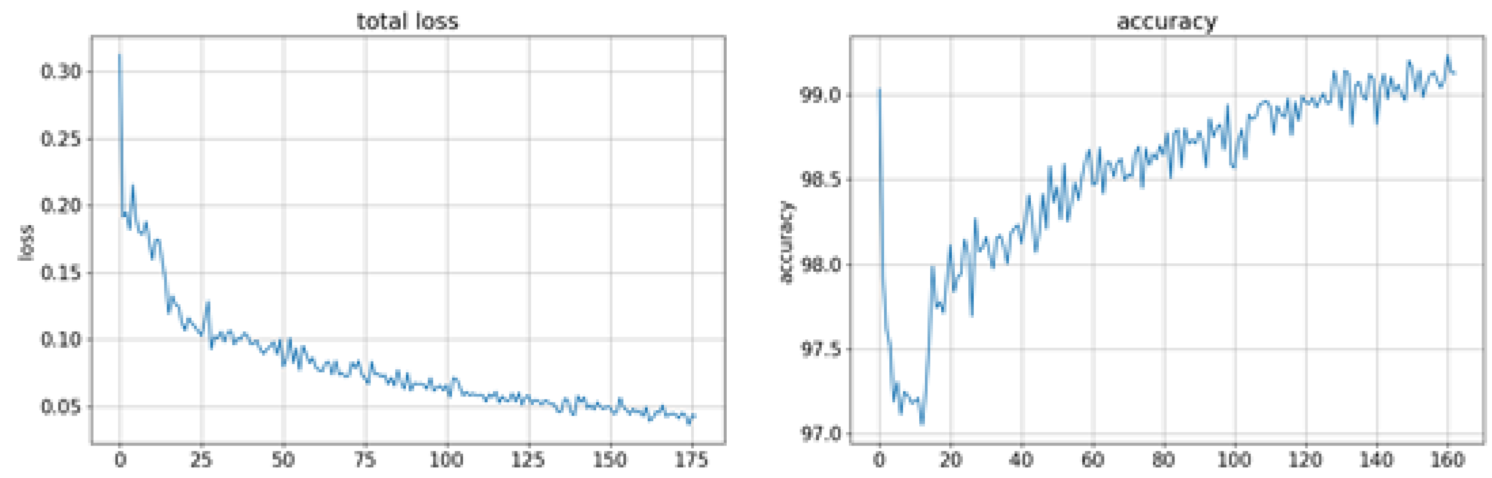
(a) Front view of the exam room with some cell phones classified by the fine-tuned model

(b) A top-rear view of a classified phone by the fine-tuned model

Figure 3



(a) Training results for accuracy and total loss



(b) Training results for classification loss, bounding box loss

Figure 4

The fine-tuned model

The fine-tuned model achieved the following:

- Accuracy 98.9 %
- Mean Average Precision (mAP) 0.783
- Recall 0.795
- F-measure 0.697
- 18 out of 19 ground truths detected

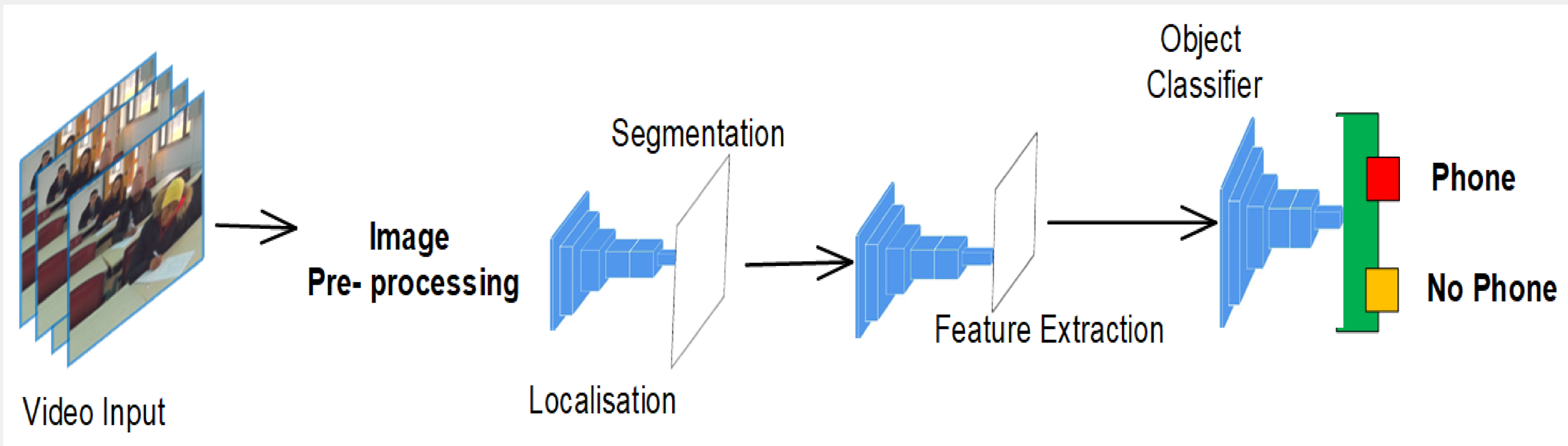


Figure 5. CNN Pipeline

Much as not all occurrences of phones in a video could be detected, the model achieves its objective.

References

[1] K. Chen et al., *"MMDetection: Open MMLab Detection Toolbox and Benchmark,"* arXiv preprint, 2019, arXiv:1906.07155.

[2] R. Girshick, "Fast R-CNN [J]," arXiv preprint, 2015, arXiv: 1504.08083v2-1.

[3] L. Jiao et al., *"A survey of deep learning-based object detection,"* IEEE Access, 2019, 7(3): 128837–128868.

[4] L. Liu et al., *"Deep Learning for Generic Object Detection: A Survey,"* International Journal of Computer Vision, 2020, 128(2): 261–318.

[5] R. Girshick, *"Fast R-CNN,"* in 2015 IEEE International Conference on Computer Vision (ICCV), 2015:1440–1448.

[6] S. Ren, K. He, R. Girshick, and J. Sun, *"Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,"* IEEE Trans. Pattern Anal. Mach. Intell., 2017, 39(6): 1137–1149.

[7] E. R. Cavalcanti, C. E. Pires, E. P. Cavalcanti, and V. F. Pires, *"Detection and evaluation of cheating on college exams using supervised classification,"* Informatics Educ., vol. 11, no. 2, pp. 169–190, 2012.

[8] K. A. D'Souza and D. V Siegfeldt, *"A Conceptual Framework for Detecting Cheating in Online and Take-Home Exams,"* Decis. Sci. J. Innov. Educ., vol. 15, no. 4, pp. 370–391, 2017, doi: 10.1111/dsji.12140.

[9] T. Liu, S. Fang, Y. Zhao, P. Wang, and J. Zhang, *"Implementation of Training Convolutional Neural Networks,"* arXiv pre-prints, 2015, arXiv:1506.01195.

[10] L. Liu et al., *"Deep Learning for Generic Object Detection: A Survey,"* International Journal of Computer Vision, 2020, 128(2): 261–318.