

# Comparison of Cellular Video Quality for Object Detection Using Neural Network Convolution

Kevin  
Universitas Prima Indonesia  
Medan, Indonesia  
[kevinntan191@gmail.com](mailto:kevinntan191@gmail.com)

Nico Gunawan  
Universitas Prima Indonesia  
Medan, Indonesia  
[nicogunawann@gmail.com](mailto:nicogunawann@gmail.com)

Mariana Erfan Kristiani Zagoto  
Universitas Prima Indonesia  
Medan, Indonesia  
[Marianazagoto12@gmail.com](mailto:Marianazagoto12@gmail.com)

Laurentius  
Universitas Prima Indonesia  
Medan, Indonesia  
[Laurentius9797@gmail.com](mailto:Laurentius9797@gmail.com)

Amir Mahmud Husein  
Universitas Prima Indonesia  
Medan, Indonesia  
[amirmahmud@unprimdn.ac.id](mailto:amirmahmud@unprimdn.ac.id)

**Abstract**— The purpose of this study is to compare the video quality between the Samsung HP camera and the Xiaomi HP camera. The object of study was UNPRI students who walked through the front yard of the UNPRI SEKIP campus. Here we test how accurate the camera's HP capture capacity is used to take the video. The method used to test this research is the Convolution Neural Network method. Object detection and recognition aim to detect and classify objects that can be applied to various fields such as face, human, pedestrian, vehicle detection, besides the ability to find, identify, track and stabilize objects in various poses and important backgrounds in many real-time video applications. Object detection, tracking, alignment and stabilization have become very interesting fields of research in the vision and recognition of computer patterns due to the challenging nature of several slightly different objects such as object detection, where the algorithm must be precise enough to identify, track and center an object from the others

**Keywords**— Camera, Video, Human, Object, & Convolution Neural Network

## I. INTRODUCTION

Object detection and recognition aim to detect and classify objects that can be applied to various fields such as face, human, pedestrian, vehicle detection (Huang, Pedoeem, & Chen, 2018), besides the ability to find, identify, track and stabilize objects in various poses and important backgrounds in many real-time video applications. Object detection, tracking, alignment and stabilization have become very interesting fields of research in the vision and recognition of computer patterns due to the challenging nature of several slightly different objects such as object detection, where the algorithm must be precise enough to identify, track and center an object from the others. Studying good video representation is the foundation of many computer vision tasks with many methods that have been proposed by researchers such as Faster R-CNN (Ren, He, & Girshick, 2017).

Mobilenet-SDD discretizes the output space of the bounding boxes into a set of default boxes above different aspect ratios and scales per feature map location (Li, Li, Lin, & Li, 2018). SSDs are relatively simple for methods that require object proposals because they completely eliminate proposal creation and subsequent pixels or resampling feature stages and summarize all calculations in one network. This makes it easy to be trained and directly integrated into systems that require detection components, using a model architecture called MobileNet based on convolution that can be separated in depth, suitable for use in embedded devices (Li, Li, Lin, & Li, 2018). MobileNets is based on a sleek architecture that uses convolution that can be separated in depth to build a lightweight neural network (Howard, et al., 2017)

In this paper, we propose the use of the CNN method for object detection using Single deep neural network with the Single Shot MultiBox Detector (SSD) approach to object data sourced from the recording of student activity using 2 (two) types of camera types of mobile devices namely Samsung and Xiomi. The analytical framework was tested

by the Mobilenet-SSD method to detect objects from 2 camera source records and then evaluated by comparing the effect of the recording quality of the mobile device on the accuracy of object recognition using the Mobilenet-SSD method. The results of the model classification will be evaluated every frame per second (FPS) on 2 types of video recordings and testing is done with the stratified cross validation method to measure accuracy based on the confusion matrix (Li, Li, Lin, & Li, 2018).

## II. LITERATURE REVIEW

Object detection is one of the fields of computer vision research that attracts much attention by researchers, it is seen from many new methods proposed with a high degree of accuracies such as Faster R-CNN, SSD, YOLO, and others. The R-CNN Faster method proposed by Shaoqing Ren (Ren, He, & Girshick, 2017) is a proposed method to improve accuracy and efficiency in object detection from the Fast R-CNN and R-CNN methods. The R-CNN Faster uses a feature map to identify regional proposals, a separate network is used to predict regional proposals. The predicted region proposal is then reshaped using the RoI pool layer which is then used to classify the image in the proposed area and predict the offset value for the bounding box. In addition, this method has two networks namely the network proposal area (RPN) to produce regional proposals and the network uses this proposal to detect objects. The main difference here with Fast R-CNN is that it will later use selective search to produce regional proposals. The time costs of producing regional proposals are much smaller in RPN than selective searching, when RPN shares most of the calculations with the object detection network. In short, the RPN ranks the area squares (called anchors) and suggests the objects that are most likely to contain objects.

Single Shot MultiBox Detector (SSD) method proposed by Wei Liu (Liu, et al., 2016), to detect objects in an image using a neural network in a single, discretizes the boundary box output space into a set of default boxes for various ratios and scales per location feature. At the time of prediction, the network generates a score for the existence of each object category in each default box and results in adjustments to the box to better match the shape of the object. In addition, the network combines predictions from several feature maps with different resolutions to naturally handle objects of various sizes and SSDs have much better accuracy, even with smaller input image sizes, while the YOLOv3 method proposed by Joseph Redmon (Redmon & Farhadi, 2018) The YOLO network divides images into regions and predicts bounding boxes and probabilities for each region. This bounding box is weighted by the predicted probability.

This model has several advantages over classifier-based systems. It is seen in all images at the time of testing so that the predictions are informed by the global context in the image.

## III. PROPOSED METHOD

In this study, the work procedures in this study are:

1. This dataset is in the form of video data of student activity recordings that are held together using 2 (two) types of cellular devices
2. Object detection itself has a role in determining the objects captured by the camera, stored in a database or not. So that it can conclude the results of objects that have been tracked.
3. Feature Extraction is the process of taking the characteristics of an object that can describe the characteristics of the object.
4. Classification is the process of grouping objects based on the characteristics of similarities and differences.
5. Object Tracking is able to detect moving objects, filter out noise, and other movements that are not needed.
6. The results of this study are the MobileNet-SSD method which is proven to be able to classify objects which at this stage the objects in the dataset are classified whether included in people, objects and others

Comparative testing of the video quality of mobile devices for object recognition using the MobileNet-SSD method is carried out in several stages, namely the training phase aimed at training the MobileNet-SSD network. Training data can be seen in the picture below. The process of data retrieval uses 2 (two) cellular devices, namely Samsung and Xiaomi by taking natural data around the location of the University of Prima Indonesia Campus with 9 video items totaling 412MB in size, the recording results of both cellular devices will be tested on the MobileNet-SSD method, such as shown in figure 2.

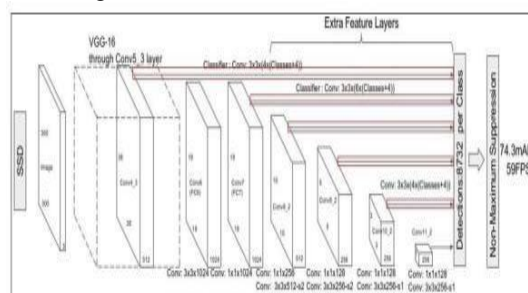


Figure 2.2 Architecture

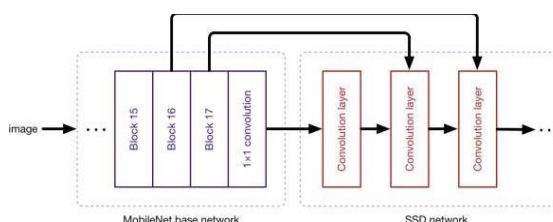
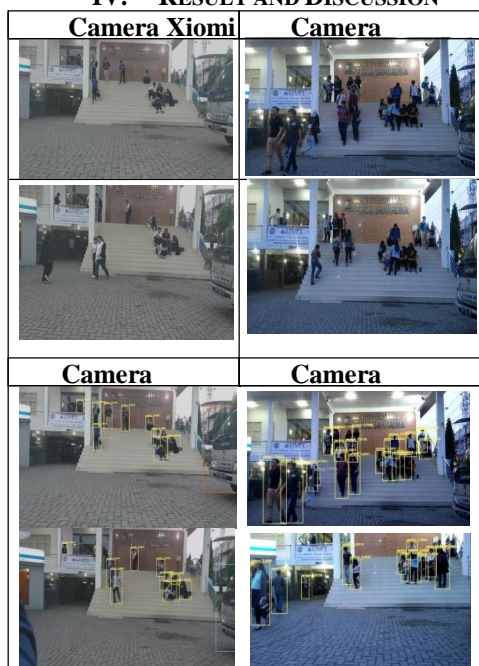


figure 2 Mobile-SSD

**IV. RESULT AND DISCUSSION**



Tabel 3.1 Classification Comparison Results

| No | Video          | Classification |         | Accuracy |         |
|----|----------------|----------------|---------|----------|---------|
|    |                | Xiomi          | Samsung | Xiomi    | Samsung |
| 1  | Video camera 1 | 91%            | 95%     | 90%      | 95%     |
| 2  | Video camera 2 | 90%            | 99%     | 92%      | 90%     |
| 3  | Video camera 3 | 90%            | 94%     | 93%      | 92%     |
| 4  | Video camera 4 | 89%            | 97%     | 92%      | 90%     |
| 5  | Video camera 5 | 92%            | 97%     | 90%      | 93%     |
| 6  | Video camera 6 | 90%            | 96%     | 90%      | 93%     |
| 7  | Video camera 7 | 90%            | 99%     | 91%      | 95%     |
|    |                | 90%            | 97%     | 92%      | 93%     |

In table 3.1 the results of the comparison of object recognition classifications using 2 (two) different camera sources, namely Xiaomi and Samsung cameras, for the use of Samsung have a classification accuracy of 96% higher than 91% Xiaomi cameras, but for the mAP value the two cameras have no difference where this indicates the accuracy of object recognition using the MobileNet-SSD method is not so big, but the results of the classification is very influential on video quality. The results in graphical form are shown in Figure 3.2

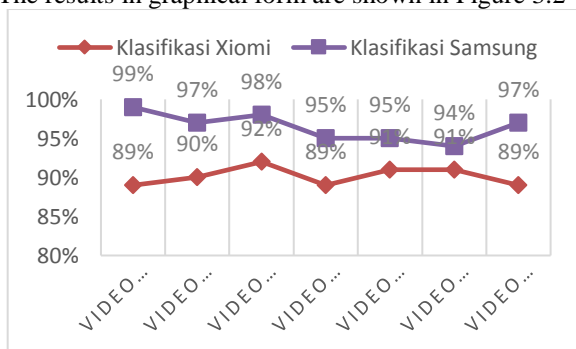


Figure 3.2 Graph Comparison of Classification Results

Based on the test results, there are differences in the size of the original video with the object classification results where the application of the proposed MobileNet-SSD method re-establishes the video to show the classification results, but the resulting size increases by 22%

**V. CONCLUSION AND SUGGESTION**

In this study the MobileNet-SSD method is proposed for object recognition with 2 (two) different data sources, namely Xiaomi and Samsung cameras, from the test results several conclusions can be drawn, namely, proposed MobileNet-SSD method has object recognition accuracy above 90% where Samsung cameras are higher by 97% compared to data sources from Xiaomi cameras by 94% but quality of video recording results has a significant effect on accuracy. object recognition, but for Xiaomi camera mAP value 93% while samsung camera 93% and object recognition classification results using the MobileNet-SSD method produce 22% larger video sizes compared to the original viode from both data sources.

**References**

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., . . . Adam, H. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications.

Huang, R., Padoem, J., & Chen, C. (2018). YOLO-LITE: A Real-Time Object Detection Algorithm Optimized for Non-GPU Computers. *IEEE International Conference on Big Data (Big Data)* (pp. 2503–2510). Seattle, WA, USA, USA: IEEE. doi:<https://doi.org/10.1109/BigData.2018.8621865>

Li, Y., Li, J., Lin, W., & Li, J. (2018). Tiny-DSOD: Lightweight Object Detection for Resource-Restricted Usages. *29TH BRITISH MACHINE VISION CONFERENCE* (pp. 1-12). Northumbria: Northumbria University.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2016). SSD: Single Shot MultiBox Detector. *European Conference on Computer Vision* (pp. 21-37). Amsterdam, The Netherlands: Springer.

Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. Washington, Washington, USA.

Ren, S., He, K., & Girshick, R. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (pp. 1137–1149). IEEE. doi:<https://doi.org/10.1109/TPAMI.2016.2577031>