

# USING YOLO IN DETECTING OBJECTS AT MIXFARM OELAMASI

ALBERT LAMBA<sup>1</sup>, TUGA MAURITSUS, DRS., MT. PHD<sup>2</sup>

<sup>1</sup>Information System Management Department, BINUS Graduate Program, Master of Information Systems Management, Bina Nusantara University, Jakarta 11480, Indonesia

<sup>2</sup>Information System Management Department, BINUS Graduate Program, Master of Information Systems Management, Bina Nusantara University, Jakarta 11480, Indonesia

E-mail: <sup>1</sup>albert009@binus.ac.id, <sup>2</sup>tmauritsus@binus.edu

## ABSTRACT

Supervise farm field can be challenging. It takes too much time and effort just to explore the land. So, we suggest the farmer to start using Solar CCTV. Adding AI to detect activity on what happened in live recording monitoring easier. This paper discuss subtopic in artificial intelligence which is Convolutional Neural Network (CNN) to detect object on farm environment using YOLO. This research was carried out in a plantation located in Oelamasi, Kupang Regency, NTT with the aim of identifying object in the farm. This research purpose is to supervise ongoing process on the field.

This research in conduct on papaya field with recorded CCTV video. This research used google collab as training platform, Yolov3 as reference model. The available GPU on google collab is Nvidia Tesla K80. We also predesign integrated system using cloud services so the data can be access via mobile apps or web browser. We hope from what we are doing can give others insight about the uses of artificial intelligence then can be applied on industries.

**Keywords:** CNN, YOLO, CCTV, CLOUD, Collab

## 1. INTRODUCTION

Latest technological development can be turning point for the change in history. With faster, better, cheaper product and services, old school method can't compete with automation. With the help from AI, we conducted an experiment from CCTV recording to detect object.

## 2. EASE OF USE

Here are some of basic concept of this paper. From farm monitoring, machine learning and using tools like google collab.

### 2.1. Farm Monitoring

First, why farmer should monitor their farm. In some research conduct in rabbit farm based on entrepreneur values[1], [2]. Said, monitoring can prevent unexpected activities. Several things that need to be monitored in farm activities are ensuring that land workers have carried out their duties according to the specified schedule, land clearing activities by workers, fertilizing activities, monitoring of land by workers also detecting activities that are unexpected happening on plantation land.



Figure 1: Farm we use to conduct research

### 2.2. Related Works

As in the paper "Fire Hotspots Detection System on CCTV Videos Using You Only Look Once (YOLO) Method and Tiny YOLO Model for High Buildings Evacuation". In this paper, we discuss the use of tiny yolo on CCTV to detect fire. This paper discusses the accuracy of yolo in detecting the presence of fire using binary classification (TP, NP, TN, FN). The results of this study found that the average success of yolo in determining the object was 90%.[3]

Other research conduct with YOLO on pig farm [4], stated that many constrain can impact YOLOv2 to detect object such as: environment, the backgrounds, camera angels, and colour of objects.

### 2.3. Machine Learning

Machine learning is the process of training computers to find patterns from feature data based on algorithms created.[5]

#### 2.3.1. Supervised

Supervised learning Is a method to determine the pattern of data that has been labelled. The level of accuracy is a measure of the success of the model. Examples of algorithms used are regression and classification.

Regression is used to see the relationship between variables. Usually, to predict. The way this algorithm works is to map the data in coordinates and generate a function that can represent the location of the data in the curve. Popular examples of regression are linear regression, logarithmic regression and polynomial regression

Classification aims to classify data based on categories. Popular algorithms are support vector machine (SVM), decision tree, naive bayes. Bayes itself is an algorithm that is the basis of various types of machine learning. The way this algorithm works is to map the probability of the data based on the features that the data has.

#### 2.3.2. Unsupervised

Unsupervised learning is a method for determining patterns from data to generate new structured information. Clustering itself is a data mining technique which produces structured data groups such as exclusive, overlapping, hierarchical, and probabilistic data. An unsupervised example is K-mean. K-mean itself is a classification method based on the centre of data groups that are close together CNN object detection.

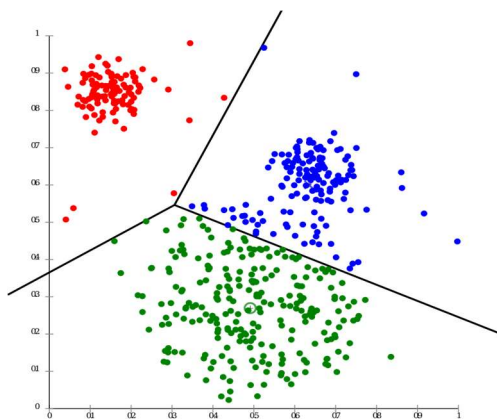


Figure 2: K-Mean

#### 2.3.3. Reinforcement

Reinforcement learning is a computer learning method by providing feedback in the process. Like unsupervised, reinforcement learning does not initially need labels. In the process, the algorithm

will optimize the results based on the feedback provided. Reinforcement learning models such as Markov decision process (MDP), Monte Carlo, etc.

#### 2.3.4. Assessment

There are several tools that are used to assess the results of machine learning. Namely as follows Binary classification, commonly used to monitor the success rate of supervised learning. Where the results of the algorithm are compared with the initial data. The number of classified data is compared with the number of initial data. Grouped into 4 categories, namely true positive, true negative, false positive and false negative. The success ratio of an algorithm is:

- The true positive (TP) rate is the number of true positives compared to the number of true positives and false negatives.
- The true negative (TN) rate is the number of true negatives compared to the number of true negatives and false positives
- Positive predicted value is the ratio of true positives to the number of true positives and false positives
- Negative predicted value is the ratio of true negatives to the number of true negatives and false negatives

### 2.4. CNN

There are 3 research contributions that support the detection process in images. Namely the image storage format, pixel classification and the faster computing process on the computer to process it. It takes a type of image storage that is easy to classify pixels. As well as a fast-computing process to reach real-time. The computational process in classifying pixels in images must be able to reach a minimum of 30 images every second (30fps). [6]

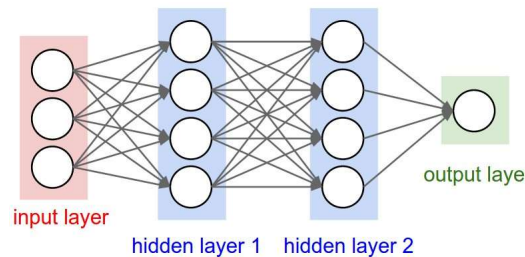


Figure 3: CNN

With the help of CNN (convolutional neural network), the object detection process becomes easier. Basically, CNN has lines in filtering input to produce output which becomes material for processing on convolve lines. There are several types of filters on CNN in detecting objects, namely line filters, colours, shapes, etc. The result of the first filter process, becomes the input for the next filter.

And finally, it can be data to be analysed using classification. Usually use SVM.

In the CNN architecture there are several types of layers that have certain functions, namely convolutional layer, pooling layer and fully connected layer. Convolutional layer is a layer that functions to process data and produce data that can only be processed by a specific layer behind it. Pooling layer functions to combine and simplify input and produce output to the 1 layer behind it. A fully connected layer is connected to all the layers in front of it.

R-CNN (regional convolutional neural network) is an advanced method of CNN. Where R-CNN forms the region on the input image first. After that proceed with the CNN process. Specifies the Region to be separated from the image using the method. That is by making square pieces of the image and running CNN from each square.[7]

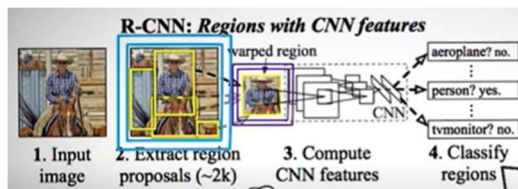


Figure 4: R-CNN

Fast R-CNN is an advanced method of R-CNN. The R-CNN method requires a lot of (enormous) time to process because in one image, CNN must run for each box that is formed. Can occur thousands of times CNN processes in 1 image. Fast R-CNN speeds up and simplifies the R-CNN method. That is using ROI pooling (Region of Interest pooling) and combining CNN results from other CNN results. Roi Pooling itself is a method to form a region based on the potential presence of an object. [8]

The drawback in fast R-CNN is that the Roi pooling process takes a long time. Makes the time needed to process not real-time. The Faster R-CNN method speeds up the Regional Proposal Network process. Where the best result of the box formed is labelled as a value. This value will be used as a benchmark for the priority of the CNN process sequence.

Mask R-CNN continued development of Faster R-CNN. That is, make a segment of the object down to the pixel level.



Figure 5: Mask R-CNN

## 2.5. Tools

In using machine learning, several devices are needed to perform the required tasks, for example libraries and functions that have been designed by developers so that users do not need to create new algorithms. machines to run software when training data also require machines with special specifications, in order to shorten the time when training data. The following tools are needed in this research:

### 2.5.1. OPEN CV

This is a programming functions that focuses on computer image processing.

### 2.5.2. YOLO

This is a tool that runs in the Darknet (open-source neural network). YOLO can be used to detect various objects quickly using the CNN method and is claimed to be even faster than faster RCNN.

### 2.5.3. Google Collab

Google Collab is a training platform for machine learning provided by google to execute python code. Also providing Nvidia GPU Tesla K80 GPU engine

### 2.5.4. Kaggle

This google platform is a community of machine learning researchers to discuss the results of developing and training machine learning.

### 2.5.5. CUDA

Software from Nvidia. Nvidia is a company that produces processors where the chip architecture is designed and adapted to the needs of the task to be run. In this case, the processing unit needed is one that supports the computational needs of machine learning

### 2.5.6. Solar CCTV

Useful tool for recording and providing data that is used as research material. With the solar panel as a power source, CCTV can be placed anywhere as needed. Using connectivity with 4G facilities that can upload recorded data to the specified cloud

### 3. METHODOLOGY

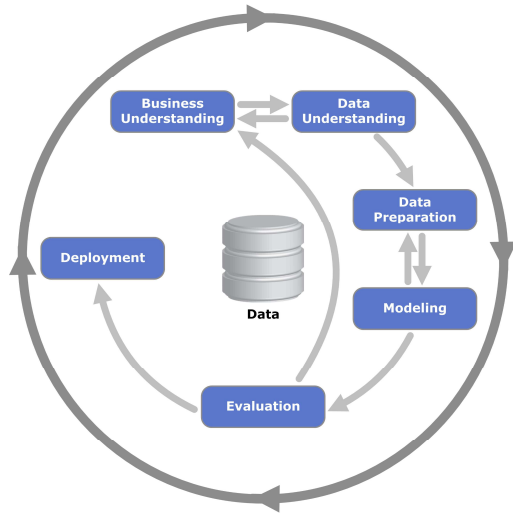


Figure 6: CRISP-DM

#### 3.1. Cross industry standard process for data mining (CRISP-DM)

This is a standard method for developing data mining. There are 6 phases in developing data mining, namely: Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation and Deployment.

##### 3.1.1. Business Understanding

This study aims to develop an application system using machine learning assistance. This research is expected to be able to assist the monitoring process on plantation land. Activities that are expected to be monitored include land clearing activities by officers, fertilization activities, monitoring by officers to detect unpleasant activities occurring on plantation land.

##### 3.1.2. Data Understanding.

Because the monitoring process requires image data as processed data, to detect image data in general using the CNN (convolutional neural network) method. The image data used were selected based on the research objectives, namely on plantation land. The video format used is .mp4

##### 3.1.3. Data Preparation

This is a phase where researchers collect data used in research. Image data taken is sourced from CCTV recordings in oil palm plantations and video recordings in plantations obtained from the internet. Another source is obtained from live streaming YouTube on plantation land.

##### 3.1.4. Modelling

This phase of machine learning is to train data and generate models from algorithms. This study uses google collab as a platform to process the python

code created. Imported libraries are TensorFlow and OpenCV. Data storage on Google Drive.

##### 3.1.5. Evaluation

Evaluation is the phase of the researcher to evaluate the model created and adjust it to the research objectives. By showing the results of research at Kaggle in order to get input and feedback from other researchers

##### 3.1.6. Deployment

Using the established model. In the form of a simple web view.

#### 3.2. Steps on conduct in this research

- Google collab is a platform for executing code for research machine learning purposes. How to use this application is to create a google account and register at [colab.research.google.com](https://colab.research.google.com).
- TensorFlow is a library provided by Google to support machine learning research needs. Use it by importing the above line of code:

```
#using tensorflow
import tensorflow as tf
```

- OpenCV (yolo) is a library commonly used to process image data. How to use this library is to import the above line of code:

```
#using opencv
import cv2 as cv
```

- Kaggle is a platform for machine learning researchers to discuss machine learning research. How to use this platform is to see the process of researchers in stages regarding the topic under study.

### 4. RESEARCH

#### 4.1. Prepare tools and accounts for research

- Prepare a google account and open the link <http://colab.research.google.com>.
- Make a "new notebook" and connect to the available google compute engine.
- Mount drive to import YOLO and videos as storage.

```
from google.colab import drive
drive.mount('/content/gdrive')
```



- Make sure to use Tesla K80 GPU (12GB) GPU

```
!nvidia-smi
```

- Download and install the required github repository and libraries

```
import os
os.environ['PATH'] += ':/usr/local/cuda/bin'
!rm -fr darknet
!git clone https://github.com/AlexeyAB/darknet/
!apt install gcc-5 g++-5 -y
!ln -s /usr/bin/gcc-5 /usr/local/cuda/bin/gcc-5
!ln -s /usr/bin/g++-5 /usr/local/cuda/bin/g++-5
%cd darknet
!sed -
i 's/GPU=0/GPU=1/g' Makefile
!sed -
```

- Prepare the Weight file as a model

```
!wget https://pjreddie.com/media/files/yolov3.weights
!chmod a+x ./darknet
```

- Get directory

```
!wget https://pjreddie.com/media/files/yolov3.weights
!chmod a+x ./darknet
!pwd
from google.colab import drive
drive.mount('/content/drive')
!apt install ffmpeg libopencv-dev libgtk-3-dev python-numpy python3-numpy libdc1394-22 libdc1394-22-dev libjpeg-dev libtiff5-dev libavcodec-dev libavformat-dev libswscale-dev libxine2-dev libgstreamer1.0-dev libgstreamer-plugins-base1.0-dev libv4l-dev libtbb-dev qtbase5-dev libfaac-dev libmp3lame-dev libopencore-amrnb-dev libopencore-amrwb-
```

## 4.2. Prepare Data

The recorded video on plantations has been place in our googles drive as (.mp4). The total number of videos are 55 files. The videos are collected randomly in June 2021. Prepared files must be stored on Google Drive so Google Collab can access easily. After that, access video files from google collab and run darknet using the weight file yolov3.weight

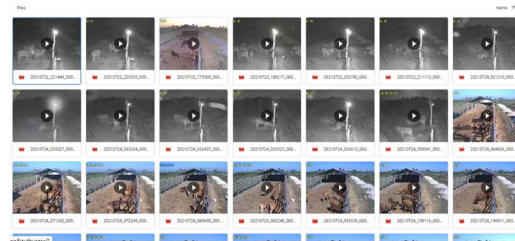


Figure 7: Drive Storage

```
from google.colab import files
uploaded = files.upload()

!./darknet detector demo cfg/coco.data cfg/yolov3.cfg yolov3.weights -dont_show 20210722_221440_000_10_P.mp4 -i 0 -out_filename 20210722_221440_000_10_PHDX.mp4 -thresh 0.7

from google.colab import files
files.download('20210722_221440_000_10_PHDX.mp4')
```

## 4.3. Executing and Result

The results in this object detection are in the form of images and descriptions of the detected objects in the form of logs. On overall object detected (3.293 objects), only “cow” (1.129 objects) and “person” (83 objects) are true. Other than that, is falsely identified. So, in conclusion Around 36% are true.

Object	Times it Occurs
Bear	4
Bird	21
Cow	1129
Dog	739
Horse	826
Person	83
Sheep	489

Figure 8: Object indentified

**4.3.1. Result when object detected is right**

```

input video stream closed.
closing...
closed!output_video_writer closed.
  CUDA-version: 11010 (11020), GPU
count: 1
  OpenCV version: 3.2.0
Demo
  0 : compute_capability = 370,
cudnn_half = 0, GPU: Tesla K80
net.optimized_memory = 0
mini_batch = 1, batch = 1,
time_steps = 1, train = 0
  layer  filters  size/strd(dil)
input          output
  0 Create CUDA-stream - 0

conv    32      3 x 3/ 1    416
x 416 x   3 -> 416 x 416 x  32
0.299 BF

FPS:17.3      AVG_FPS:16.9

  cvWriteFrame

Objects:

person: 92%

FPS:16.0      AVG_FPS:15.6

  cvWriteFrame
Objects:

cow: 83%
cow: 81%
    
```



Figure 9: When object detected is right

**4.3.2. Result when object detected is false**

```

Create CUDA-stream - 0
conv    32      3 x 3/ 1    416 x
416 x   3 -> 416 x 416 x  32 0.299
BF
  cvWriteFrame
Objects:
horse: 89%
FPS:16.0      AVG_FPS:16.1
  cvWriteFrame
Objects:
dog: 71%
FPS:16.4      AVG_FPS:16.1

  cvWriteFrame
Objects:
sheep: 73%
    
```



Figure 10: When Object detected is false

## 5. INTEGRATED DEVICE AND SYSTEM DESIGN

### 5.1. Hardware Requirement

- CCTV device used is an important part for recording, storing and sending recorded data to the cloud storage that has been provided. CCTV used must be connected to the internet or cloud ready.
- GPU is needed to execute the program so that the computation to detect objects can be done in real time. Examples of GPUs available in the market are NVIDIA Tesla K80, GeForce GTX 1080, GeForce RTX 3060, NVIDIA Tesla V100, etc. {GPU benchmarks for deep learning}

- Cloud is used for computing, storing and analysing data. Some of the services available are as follows: AWS (Amazon Web Service), Google Cloud Platform, Alibaba Cloud, Tencent Cloud, Huawei Cloud, Digital Ocean, Azure, ETC. Of the various cloud services mentioned, there are services for using GPUs for machine learning purposes.
- Smartphone or PC with a browser to monitor CCTV conditions and the results of detected objects as well as logs from CCTV recordings.

### 5.2. Software Requirement

- CUDA, Darknet can run on CPU, but will be very slow in the process. It takes a GPU to speed up the process and to use a GPU it takes CUDA. CUDA to use GPUs from NVIDIA.
- Flutter, to quickly create mobile applications and the GUI interface provided by the Flutter library looks native.
- SQL, to store logs and object detection activities and files stored in cloud storage. The most popular SQL now is NoSQL.
- Cloud Services, services from the Cloud that are needed are: cloud computing and cloud storage. On it required operating system such as Linux or windows server. As for the cloud that provides its own server. The operating system can run supporting software such as python, PHP, etc.

### 5.3. Design

- CCTV used to upload recordings into cloud storage which is provided periodically.
- Cloud storage stores video files sent from CCTV. Cloud storage also provides files to be accessed by cloud computing
- Cloud computing accesses files in cloud storage and with the help of GPU detects objects in video files and saves the results into SQL with the description of the reference file names.
- The available results can be accessed by a web browser or mobile apps connected to cloud computing.

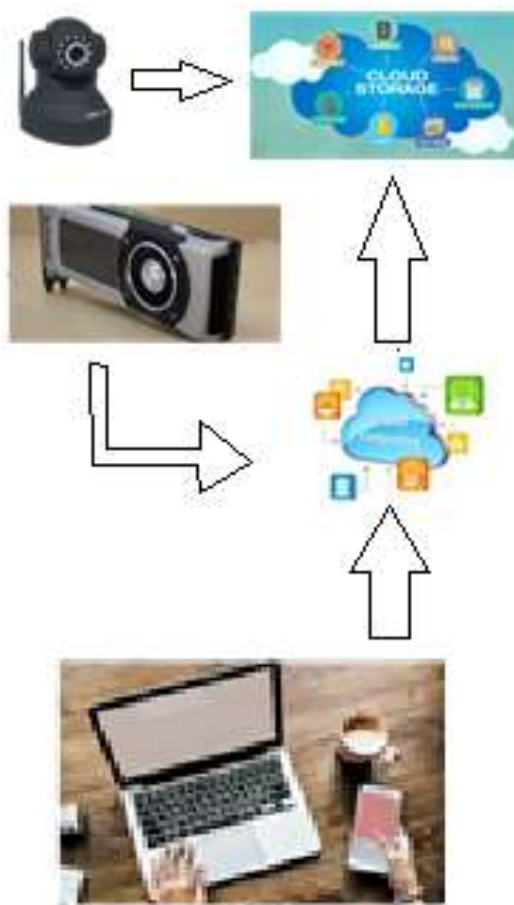


Figure 11: Design illustration

## 6. CONCLUSION

In this study, we got low accuracy rate of 36% in detecting objects. For example, a cow object identified as a horse, dog, or sheep. It is estimated that the quality of the camera used also affects the ability to detect objects and the background of objects that have similar colours. [4]

The suggestion from this research, is to try to use newer version of yolo, such as yolov5. Latest version of yolo is expected to increase the accuracy on detecting objects correctly. Also, besides the darknet, research can be conduct by using other artificial intelligence services that are more organized with software development. So, there is no need for computation to be run separately. These include artificial intelligence services from the cloud like open AI, vision AI, deep mind. Which provides access using pretrained data from a variety of trained sources.

Although, the result is low on accuracy. Monitoring activity can still be conduct with or

without AI. Due to the limitations of the researcher, it is recommended that there are researchers who continue to research the relationship between artificial intelligence and the Early Warning System. We also recommend using better CCTV quality. Other research that is expected to support is the capacity of CCTV cameras to record objects from a certain distance.

## REFERENCE

- [1] C. van Praag and P. Versloot, "What Is the Value of Entrepreneurship?," *Small Business Economics*, vol. 29, no. 4, pp. 351–382, 2007, doi: 10.1007/s11187-007-9074-x</SPAN>.
- [2] Y. Yun, H. Yueming, Z. Xiaojuan, J. Sheng, and S. Baoli, "Design and implementation of production supervising system for large scale rabbit farm," 2015.
- [3] D. P. Lestari, R. Kosasih, T. Handhika, Murni, I. Sari, and A. Fahrurrozi, "Fire Hotspots Detection System on CCTV Videos Using You only Look Once (YOLO) Method and Tiny YOLO Model for High Buildings Evacuation," *Proceedings - 2019 2nd International Conference of Computer and Informatics Engineering: Artificial Intelligence Roles in Industrial Revolution 4.0, IC2IE 2019*, pp. 87–92, 2019, doi: 10.1109/IC2IE47452.2019.8940842.
- [4] S. Khan and A. Q. Khan, "YOLOv2 for Pigs Detection in Industrial Farming," no. September, pp. 1–13, 2020, doi: 10.20944/preprints202009.0088.v1.
- [5] B. Rao, "Machine Learning Algorithms: A Review," *International Journal of Computer Science and Information Technologies*, vol. 7, no. 3, pp. 1174–1179, 2016, doi: 10.21275/ART20203995.
- [6] V. P. D.M. Gavrilă, "Real-Time object detection for 'Smart' Vehicles," *Proceedings of the Seventh IEEE International Conference on Computer Vision*, 1999.
- [7] J. Wang, L. C. Yu, K. R. Lai, and X. Zhang, "Dimensional sentiment analysis using a regional CNN-LSTM model," *54th Annual Meeting of the Association for Computational Linguistics, ACL 2016 - Short Papers*, pp. 225–230, 2016, doi: 10.18653/v1/p16-2037.
- [8] G. Plastiras, C. Kyrkou, and T. Theodoridis, "Efficient convnet-based object detection for unmanned aerial vehicles by selective tile processing," *ACM International Conference Proceeding Series*, 2018, doi: 10.1145/3243394.3243692.



File can be access through this link below or scan this barcode

Recorded file

<https://drive.google.com/drive/folders/1pxbbDpS8bez-tW8RoVHfL9eX7VrujtzV?usp=sharing>



YOLO video

<https://drive.google.com/drive/folders/14H3M61EqxhnrPrQe04IHkfLac4lPmBkt?usp=sharing>

