TITLE IN CAPITAL LETTERS

TITLE IN CAPITAL LETTERS

TITLE IN CAPITAL LETTERS

TITLE IN CAPITAL LETTERS

WAN ZUKI AZMAN WAN MUHAMAD

UNIVERSITI TEKNOLOGI MALAYSIA

PSZ 19:16 (Pind. 1/13)

UNIVERSITI TEKNOLOGI MALAYSIA

DECLARATION OF THESIS / UNDERGRADUATE PROJECT REPORT AND COPYRIGHT

Author’s full name :

Date of Birth :

Title :

Academic Session :

I declare that this thesis is classified as:

|  |  |  |
| --- | --- | --- |
|  | CONFIDENTIAL | (Contains confidential information under the Official Secret Act 1972)\* |
|  |  |  |
|  | RESTRICTED | (Contains restricted information as specified by the organization where research was done)\* |
|  |  |  |
| ✓ | OPEN ACCESS | I agree that my thesis to be published as online open access (full text) |

1. I acknowledged that Universiti Teknologi Malaysia reserves the right as follows:
2. The thesis is the property of Universiti Teknologi Malaysia
3. The Library of Universiti Teknologi Malaysia has the right to make copies for the purpose of research only.
4. The Library has the right to make copies of the thesis for academic exchange.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  | Certified by: |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  | SIGNATURE OF STUDENT |  | SIGNATURE OF SUPERVISOR |  |
|  |  |  |  |  |
|  | 721601-01-8890 |  | AP. DR. KHAIRUR RIJAL JAMALUDIN |  |
|  | **MATRIX NUMBER** |  | **NAME OF SUPERVISOR** |  |
|  |  |  |  |  |
|  | Date: 20 JUNE 2015 |  | Date: 20 JUNE 2015 |  |

NOTES : If the thesis is CONFIDENTIAL or RESTRICTED, please attach with the letter from the organization with period and reasons for confidentiality or restriction

“Choose an item. hereby declare that we have read this thesis and in Choose an item.

opinion this thesis is suffcient in term of scope and quality for the

award of the degree of Doctor of Philosophy (Specialization)”

|  |  |  |
| --- | --- | --- |
| Signature | : | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| Name of Supervisor I | : | KHAIRUR RIJAL JAMALUDIN |
| Date | : | 9 May 2017 |
|  |  |  |
|  |  |  |
| Signature | : | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| Name of Supervisor II | : | NOOR HAZARINA HASHIM |
| Date | : | 9 May 2017 |
|  |  |  |
|  |  |  |
| Signature | : | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| Name of Supervisor II | : | MOHD ZULI JAAFAR |
| Date | : | 9 May 2017 |

**BAHAGIAN A - Pengesahan Kerjasama\***

Adalah disahkan bahawa projek penyelidikan tesis ini telah dilaksanakan melalui

kerjasama antara Click or tap here to enter text. dengan Click or tap here to enter text.

Disahkan oleh:

Tandatangan : Tarikh :

Nama :

Jawatan :

(Cop rasmi)

*\* Jika penyediaan tesis atau projek melibatkan kerjasama.*

**BAHAGIAN B - Untuk Kegunaan Pejabat Sekolah Pengajian Siswazah**

Tesis ini telah diperiksa dan diakui oleh:

|  |  |  |
| --- | --- | --- |
| Nama dan Alamat Pcmeriksa Luar | **:** |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| Nama dan Alamat Pcmeriksa Dalam | **:** |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| Nama Penyelia Lain (jika ada) | **:** |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

Disahkan oleh Timbalan Pendaftar di SPS:

|  |  |  |  |
| --- | --- | --- | --- |
| Tandatangan | : |  | Tarikh : 15JULAI 2018 |
| Nama | : |  | |

Final Year Project 1

POLE DETECTION USING DEEP LEARNING

KENTO ADACHI

XM18MJ0002

A thesissubmitted in partial fulfilment of the

requirements for the award of the degree of

Bachelor of Electronic Systems Engineering

School of Electrical Engineering

Malaysia-Japan International Institute of Technology

Universiti Teknologi Malaysia

14 December 2018





**Malaysia-Japan International Institute of Technology**

**Universiti Teknologi Malaysia**

**FINAL YEAR PROJECT**

**REPORT DRAFT**

**POLE DETECTION USING DEEP LEARNING**

**NAME OF STUDENT**

Kento Adachi

**MATRIC NO**

X18MJ0002

**SUPERVISOR:**

Dr. Mohd. Ibrahim bin Shapiai @ Abdul Razak

TABLE OF CONTENTS



[TABLE OF CONTENTS 2](#_Toc532516612)

[LIST OF TABLES 4](#_Toc532516613)

[LIST OF FIGURES 5](#_Toc532516614)

[CHAPTER 1 INTRODUCTION 6](#_Toc532516817)

[1.1 Problem Background 6](#_Toc532516818)

[1.2 Problem Statement 7](#_Toc532516819)

[1.3 Research Objectives 7](#_Toc532516820)

[1.4 Scope of Study 7](#_Toc532516821)

[1.5 Significance of the Study 8](#_Toc532516822)

[LITERATURE REVIEW 9](#_Toc532516823)

[CHAPTER 2 9](#_Toc532516824)

[2.1 Introduction of chapter 2 9](#_Toc532516825)

[2.2 Overview 9](#_Toc532516826)

[2.3 Object Detection 10](#_Toc532516827)

[2.4 Artificial Neural Network 10](#_Toc532516828)

[2.5 Convolutional Neural Network 11](#_Toc532516829)

[2.6 Optical Character Recognition 11](#_Toc532516830)

[2.7 Deep Learning for Object Detection 11](#_Toc532516832)

[2.7.1 YOLO 12](#_Toc532516833)

[2.7.2 R-CNN, Fast R-CNN, Faster R-CNN 12](#_Toc532516834)

[RESEARCH METHODOLOGY 13](#_Toc532516835)

[CHAPTER 3 13](#_Toc532516836)

[3.1 Introduction 13](#_Toc532516837)

[3.2 Overview of the System – Flow Chart 13](#_Toc532516838)

[3.3 Hardware 14](#_Toc532516839)

[3.4 Software Development 14](#_Toc532516840)

[3.4.1 Find Label from Picture 15](#_Toc532516841)

[3.4.2 Image Processing 15](#_Toc532516842)

[3.4.3 Text Recognition 16](#_Toc532516843)

[3.5 Data Labeling 16](#_Toc532516844)

[3.5.1 Data Augmentation 16](#_Toc532516845)

[3.6 Yolo Model 16](#_Toc532516846)

[3.7 Gannt Chart 17](#_Toc532516847)

[RESULTS 19](#_Toc532516848)

[CHAPTER 4 19](#_Toc532516849)

[4.1 Preliminary Results 19](#_Toc532516850)

[4.2 Expected Results 19](#_Toc532516851)

[CONCLUSION AND RECOMMENDATIONS 20](#_Toc532516852)

[CHAPTER 5 20](#_Toc532516853)

[5.1 Conclusion 20](#_Toc532516854)

[5.2 Future Work 20](#_Toc532516855)

[REFERENCES 21](#_Toc532516856)

LIST OF TABLES

|  |  |  |
| --- | --- | --- |
| TABLE NO. | TITLE | PAGE |

[Table 1 Mean Average Precision(mAP-50) and time of current methods [1] 6](#_Toc532510836)

[Table 2 - Workflow 17](#_Toc532510837)

LIST OF FIGURES

|  |  |  |
| --- | --- | --- |
| FIGURE NO. | TITLE | PAGE |

[Figure 1The overview of the liteleture review 9](#_Toc532516587)

[Figure 2 Example of object detection 10](#_Toc532516588)

[Figure 3 Flow chart of the proposed method 13](#_Toc532516589)

[Figure 4 Overview image of proposed method 14](#_Toc532516590)

[Figure 5 The Architecture of YOLO [1] 17](#_Toc532516591)







# INTRODUCTION

## Problem Background

Automation technology is a big topic for us. In recent years, with the development of AI technology, we can solve various tasks. Especially character recognition is expected to be applied in various fields from data input to automatic driving, it is strongly desired to improve character recognition accuracy. The applications of character recognition in the real world can roughly be classified into two types. The first is digitization of documents. The other is recognizing signs designed for people, such as billboards. Recognition of characters on general objects depends on shape, rotation, objects around them, recognition accuracy is still far from practical use. The purpose of this research is to improve these characters recognition accuracy.

Table 1 Mean Average Precision(mAP-50) and time of current methods [1]



## Problem Statement

In this research, we recognize characters on objects, especially on cylinder. Recent studies mainly use deep learning, and this research also uses deep learning to do character recognition.

## Research Objectives

The objectives of the research are as follows:

1. To recognize character strings on the cylindrical curved surface such as streetlights.
2. To investigate the accuracy of the proposed method.
3. To attempt object detection method to the text recognition.

## Scope of Study

In this research, we use darknet machine learning framework and YOLO algorithm for image recognition. YOLO shows high accuracy as well as Faster R-CNN in object detection and much faster. Machine learning also requires large resources. Therefore, it is often done with desktop PC with high GPU and CPU power, in this research, Nvidia Jetson Xavier is used for real time detection.

## Significance of the Study

By accomplishing this research, we can detect the text on the cylinder surface. That means that we can leave the maintenance of the streetlights to robot or drone, which people had previously done manually.

# LITERATURE REVIEW

## Introduction of chapter 2

In this chapter, we review the related study, starts from overview, overview of deep learning, finally we review two major object detecting algorithm, YOLO and Faster R-CNN.

## Overview

The overview for the literature review can be seen in the figure below.

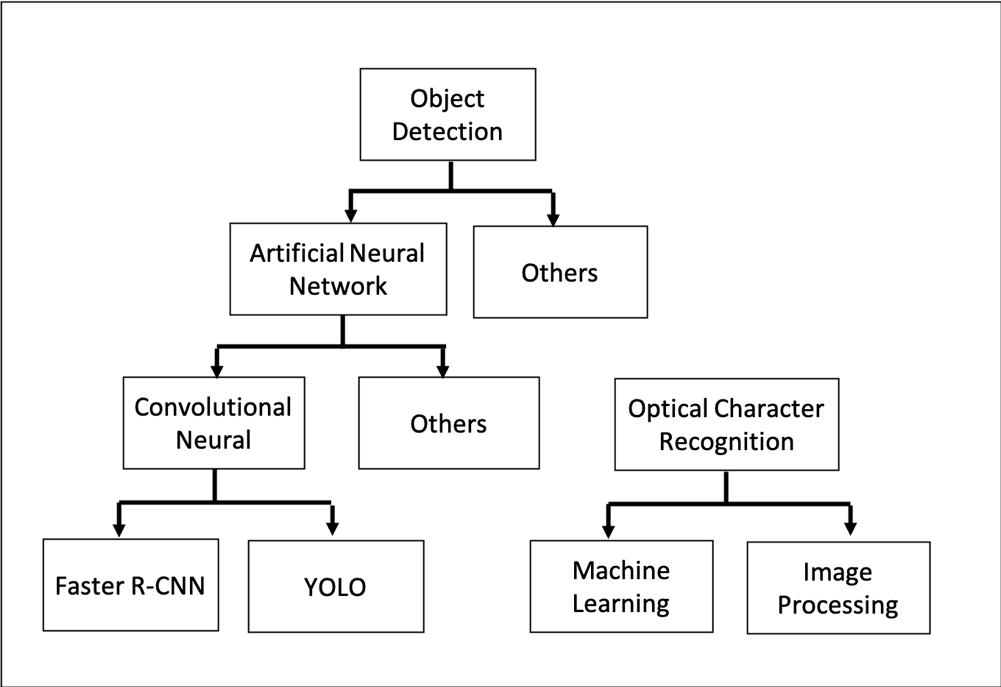


Figure 1The overview of the liteleture review

## Object Detection

Object detection is to capture an image and detect the position and category (class) of the object determined from the image. As shown in the figure below, we identify the position of the rectangle called the bounding box and its category from the image. There is "handwritten characters recognition" in the implementation example of image recognition, but in actual images, there is not always only one object. Normally, we have to find the various sized of bounding boxes in one image. After that, we can classify each image.

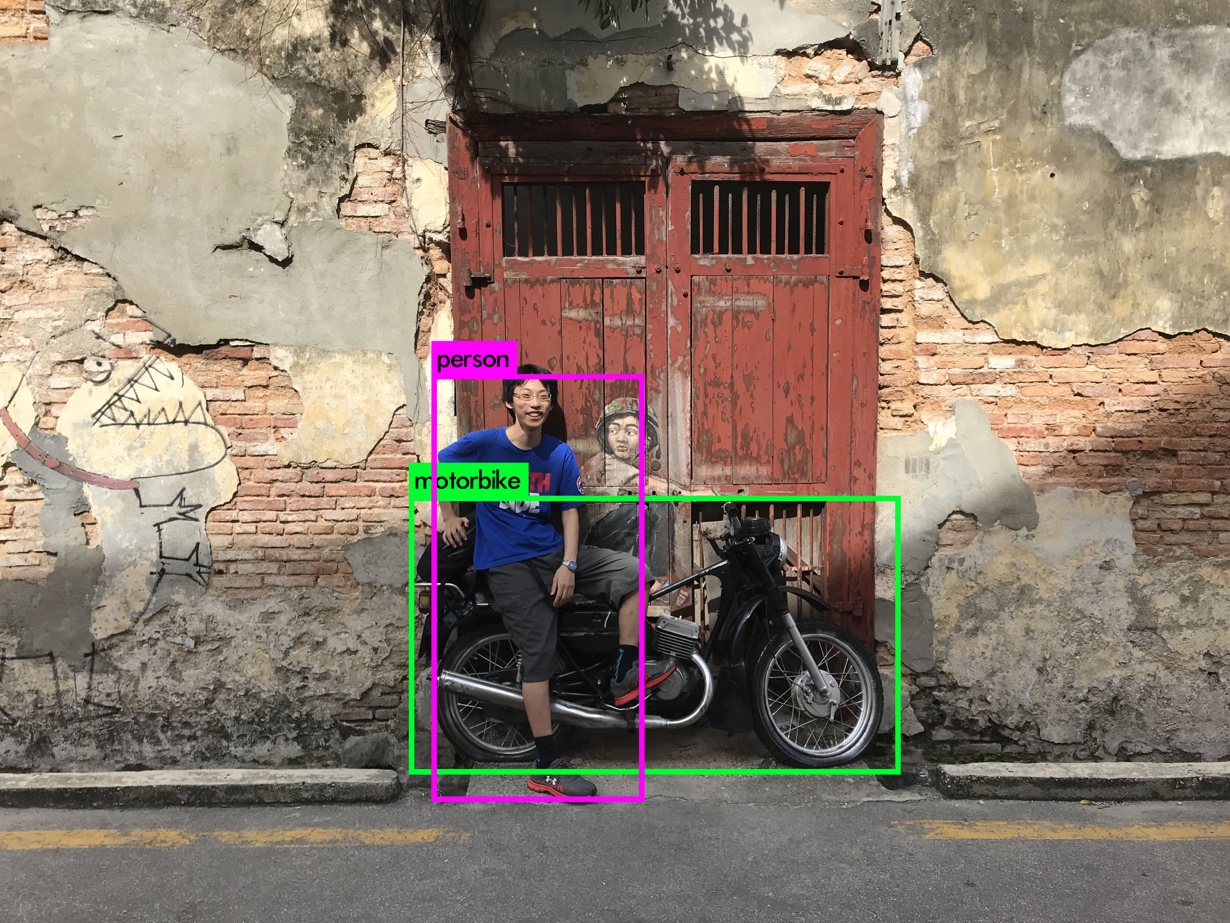


Figure 2 Example of object detection

## Artificial Neural Network

Artificial Neural Network is a mathematical model aimed at expressing some characteristics found in brain function by computer simulation. The origin of the research is modelling of the brain of the living body, but by revision of the knowledge of the neuroscience, it become more different. In the ANN, the artificial neurons forming the network aim to obtain the target output by changing the strength of the synaptic connection by training.

## Convolutional Neural Network

CNN is a multi-layered feedforward neural network which is an artificial neural network model with a simple structure devised earlier and does not have a looped connection in the network and signal propagates only in a single direction. It is a widely used model to recognize images and videos.

## Optical Character Recognition

Optical character recognition is to recognize characters on an image and convert them into character data. Conventional image recognition has been performed by template matching for comparing bit differences between recognition target images and font images. This method is weak against the inclination and rotation of the image. It shows high accuracy for the image scanned by the scanner, but recognition of characters from the photograph is far from practical use. Recent studies have shown that using machine learning can achieve higher accuracy than this. Therefore, we aim at recognition of character strings in photos using deep learning which is a type of machine learning.

## Deep Learning for Object Detection

In object detection using deep learning, instead of designing a feature amount by a person, it presents a large number of images to DNN. Then, DNN finds the feature from the labelled learning data. Then change the network by itself to respond to patterns close to the given image.

### YOLO

YOLO is a real-time object detection algorithm, which was announced in 2016. Regarding existing image detection algorithms such as "DPM" and "R - CNN", region estimation and classification of images are separated, and therefore processing tends to be complicated and processing time tends to be long. In "YOLO", we perform that "image area estimation" and "classification" at the same time. The algorithm of "YOLO" is simple because it is completed with one CNN, and it has merits such as quick processing as compared with the existing method, less background error detection.

### R-CNN, Fast R-CNN, Faster R-CNN

R-CNN [2] was invented based on CNN extracting features from the entire image. It can detect features of each region and that is close to human's recognition process. First, in R-CNN, region proposal is extracted from the input image. Then, we do recognition for the proposed region. Fast R-CNN [2] is an improvement of R-CNN model to reduce the execution count of the CNN Detector. Faster R-CNN [3] is an improvement of Faster R-CNN model to make the region proposal which is a bottleneck in Faster R-CNN better. Faster R-CNN integrates small regions having similar features and chooses an area where an object is likely to exist.

# RESEARCH METHODOLOGY

## Introduction

In this chapter, There are descriptions of proposed method how to recognize the label on the pole.

## Overview of the System – Flow Chart

The overview is shown in Figure below. Our proposed method can be divided into three processes. A process of extracting a label from a photograph, a process of image processing on the cut of image, and a process of zing the output.

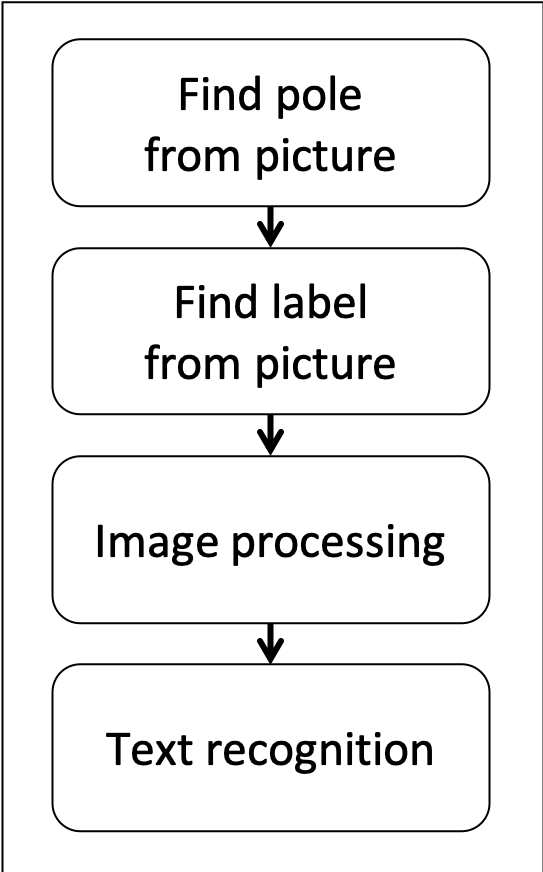


Figure 3 Flow chart of the proposed method

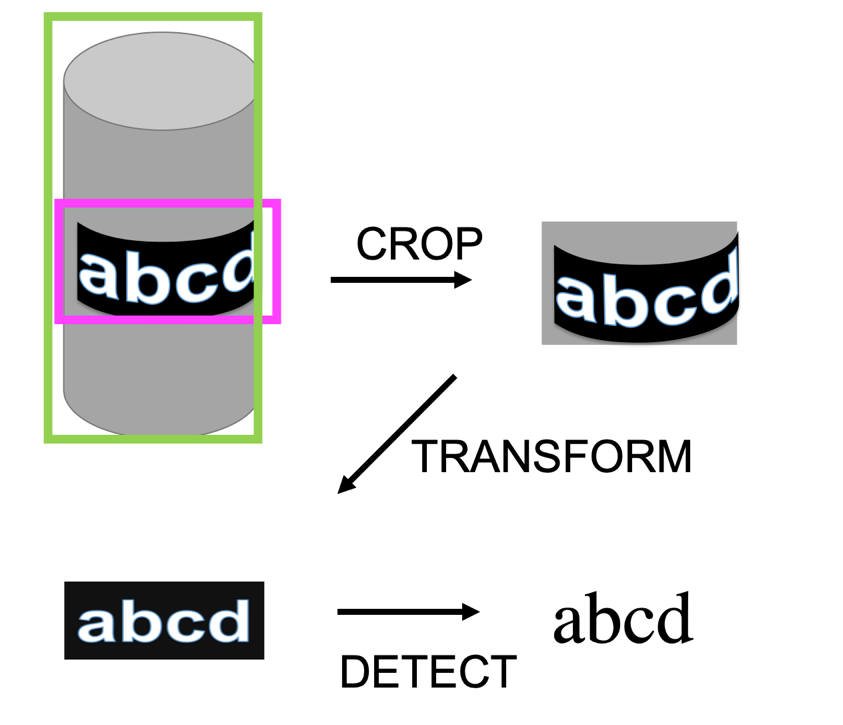


Figure 4 Overview image of proposed method

## Hardware

We use Jetson Xavier as hardware. The Jetson series is an AI computer board that can be mounted on relatively small mobile devices such as robots, drone, security cameras developed by Nvidia. The latest model "Jetson Xavier" has the 512 core Volta Tensor GPU and theoretically gained 20 times the performance of the conventional model TX 2.

## Software Development

There are three steps to detect pole and recognize the text of it.

### Find Label from Picture

Before recognizing characters, we extract the label region including character strings to be recognized from the photograph. By doing this, it is possible to suppress the detection of metabolism of characters and to reduce the calculation amount. To detect the label area, use YOLO as before. First of all, we label on a plain picture. Labeling is to give additional information such as classes and their coordinates included in images to original data, and machine learning learns based on this annotation data. For training data, we use photographs of streetlight poles shot from in-vehicle camera. The structure of YOLO's annotation data is simple and is expressed as {height of width object of center y coordinate object of center x coordinate object of category number object}. Since it is hard to edit the file directly, it is possible to visually perform the work by using the input support tool for labeling for labeling. We manually labeled a large number of photographs.

### Image Processing

Ordinary, image recognition technology is designed for a flatten image. If distorted images are used for recognition, high accuracy cannot be obtained. Therefore, it is necessary to transform label to a flat surface. To transform a curved surface image into a plane, use the following method.

1. First get the center of x-coordinate of detected pole.
2. Calculate the normal map of the pole.
3. Deform the image based on the given normal map.

This transformation reduces the distortion in the lateral direction, and recognition can be performed with higher accuracy.

### Text Recognition

Recognize characters from flatten image. The recognition process is shown below. We perform object detection of 52 classes of alphabet a - z A - Z using YOLO. For training, use padded font data by placing noise on the image or deforming it.

## Data Labeling

An image is necessary for the neural network to train. In our proposed method, in order to recognize the streetlights and its label area, 1000 images of the streetlight taken from the in-vehicle camera were prepared and each image was manually labelled. We used the labelling tool “labelimg” corresponding to format YOLO.

### Data Augmentation

A large amount of training data is necessary for deep learning. Therefore, in deep learning in recent years, researchers performed Gaussian noise on original data, adjust contrast, adjust brightness, perform smoothing, expand / shrink / rotate, invert by combining processing to augment the data. By applying data augmentation, it is possible to recognize with high accuracy with less data.

## Yolo Model

In YOLO, the image is divided into S × S grid. Each grid cell predicts B bounding boxes, detects one object, predicts C conditional class probabilities below.

P (what the object is | the probability that there is an object in the cell)

Each bounding box consists of the following five parameters.

1. Each bounding box consists of the following five parameters.
2. x: the coordinates of the center of the bounding box (x axis)
3. y: the coordinates of the center of the bounding box (y axis)
4. w "Breadth of bounding box
5. h: Height of bounding box
6. probability there is an object

The shape of YOLO’s prediction is (S, S, B×5 + C). In the paper, it used 7x7 bounding boxes to detect 20 classes. So, the shape of prediction should be (7,7,30). The Architecture of YOLO is below. This network consists of 24 layers of convolution and 2 dense layers.

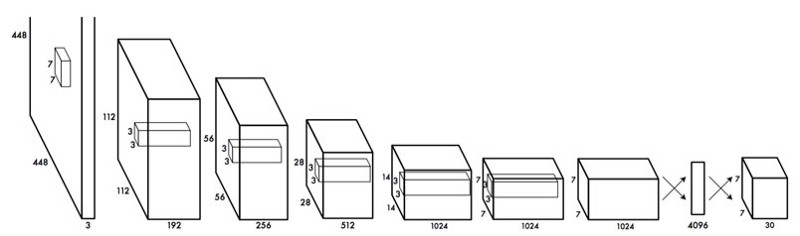


Figure 5 The Architecture of YOLO [1]

## Gannt Chart

We can see the Gantt chart of the workflow below.

2

# RESULTS

## Preliminary Results

We manually labeled streetlights and label areas from landscape images taken from the in-vehicle camera as training data. In order to analyze the recognition accuracy, we labeled three different types of streetlights.

.

## Expected Results

The expected result is to realize character detection with high precision so that robots can recognize signs and signs in the real world. This time we limited it to character detection on a cylinder, but in the future, we would like to consider a method to predict a three-dimensional shape from a photograph and reshape it on a plane.

# CONCLUSION AND RECOMMENDATIONS

## Conclusion

We have proposed a method to recognize character strings on a cylinder surface from photographs using deep learning. In particular, this study focused on detecting streetlight and read the character on it. We have labeled the images. Deep learning has the power to solve the problems that we have not accomplished so far and will continue to be applied in various fields from now on.

## Future Work

In the near future, the resolution of the photo will be improved and vividly reflect objects far away. Object detection technology will exceed the human cognitive ability. Then the object detection technology will exceed the human cognitive ability. I hope that this research contributes to improving those technology.

REFERENCES

|  |  |
| --- | --- |
| [1] | S. D. R. G. a. A. F. Joseph Redmon, You Only Look Once: Unified, Real-Time Object Detection, 2016. |
| [2] | J. D. T. D. J. M. Ross Girshick, Rich feature hierarchies for accurate object detection and semantic segmentation, 2013. |
| [3] | R. Girshick, Fast R-CNN, 2015. |
| [4] | K. H. R. G. J. S. Shaoqing Ren, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, 2015. |