POULTRY ESOPHAGUS DETECTION USING RETINANET AND MASK REGION-BASED CONVOLUTIONAL NEURAL NETWORK OBJECT DETECTION MODEL

NOR AZIAH AMIRAH NOR MUHAMMAD

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Master of Philosophy

Malaysia-Japan International Institute of Technology
Universiti Teknologi Malaysia

MARCH 2022

DEDICATION

This thesis is dedicated to my family, who have supported me on my journey in completing this study. To my friends and lab mates who have kept me motivated and given words of encouragement along the way. Thank you to all the important people in my life.

ACKNOWLEDGEMENT

In preparing this thesis, I was in contact with many people, researchers, academicians, and practitioners. They have contributed towards my understanding and thoughts. In particular, I wish to express my sincere appreciation to my main thesis supervisor, Dr. Uswah Khairuddin, for encouragement, guidance, critics, and friendship. I am also very thankful to Professor Dr. Rubiyah Yusof for her guidance, advice, and motivation. Without their continued support and interest, this thesis would not have been the same as presented here.

I am also indebted to MJIIT, Universiti Teknologi Malaysia (UTM) for funding part of my Master’s study. Librarians at UTM and the staff of CAIRO also deserve special thanks for their assistance in supplying the relevant pieces of literature and providing a working space.

My fellow postgraduate student should also be recognized for their support. My sincere appreciation also extends to all my colleagues and others who have provided assistance on various occasions. Their views and tips are useful indeed. Unfortunately, it is not possible to list all of them in this limited space. I am grateful to all my family members.

ABSTRACT

Syariah Compliance Automated Chicken Processing System (SYCUT) is a system for monitoring the slaughtering process to ensure that chickens are slaughtered in accordance with sharia of Islam. SYCUT uses vision inspection technology to determine whether slaughtered chickens are halal or otherwise. The vision inspection technology consists of a detection module to detect whether the esophagus of chickens is cut accordingly. The researcher employed Viola-Jones object detection framework to train the detection module. The detection module had problems due to images of esophagus that were bloodied, blurred, or occluded. This resulted in a low detection rate in the system. Besides, a conventional method requires image preprocessing tool like low-pass filter and Otsu's thresholding to improve the conditions of the images before detection which adds to the computational cost. In this study, the researcher divided image inputs into categories to reduce misclassification and aid in data annotation. Then, the researcher proposed a poultry esophagus detection system based on deep learning to improve the current algorithm in SYCUT. The researcher combined the deep learning method with the RetinaNet and Mask R-CNN models, which could perform segmentation and object detection in a single image. The researcher then compared the proposed method with the previous conventional SYCUT algorithm. The proposed method could detect bloodied and occluded images more accurately. The developed algorithm improves overall esophageal detection performance from 68.65 to 92.77 per cent. The SYCUT performs efficiently even in uncontrolled working environments due to the effectiveness of the developed deep learning method. However, the limitation of this deep learning method is it needs huge data for training. This research only improves the detection of certain image types, like bloodied and occluded. Future work should include improving the precision-recall value of the system and its real-time implementation for esophageal detection in real or simulated environments.

ABSTRAK

Sistem Pemprosesan Ayam Automatik Patuh Syariah (SYCUT) adalah sistem yang digunakan untuk memantau proses penyembelihan untuk memastikan bahawa ayam yang disembelih menepati Syariah Islam. SYCUT menggunakan Teknologi Pemeriksaan Visi untuk mengesan dan mengklasifikasikan sama ada ayam yang disembelih adalah halal atau tidak. Teknologi Pemeriksaan Visi mengandungi modul pengesan untuk mengesan sama ada esofagus ayam dipotong dengan betul. Penyelidik menggunakan rangka kerja pengesanan objek Viola-Jones untuk melatih modul pengesan. Modul pengesan berhadapan dengan masalah hasil daripada imej-imej esofagus seperti imej yang berdarah, kabur dan terhalang. Ini mengakibatkan kadar pengesanan yang rendah dalam sistem. Selain itu, kaedah konvensional memerlukan alat prapemprosesan imej seperti penapis laluan-rendah dan pengambangan Otsu untuk menambah baik keadaan imej sebelum pengesanan, menyebabkan kos pengiraan bertambah. Di dalam kajian ini, penyelidik membahagikan pemasukan imej kepada kategori masing-masing bagi mengurangkan klasifikasi salah dan membantu dalam anotasi data. Kemudian, penyelidik mencadangkan sebuah sistem pengesan esofagus unggas yang menggunakan pembelajaran dalam bagi menambah baik algoritma sedia ada dalam SYCUT. Penyelidik menggabungkan kaedah pembelajaran dalam dengan model RetinaNet dan Mask R-CNN yang boleh melakukan segmentasi dan pengesanan objek dalam satu imej. Penyelidik membandingkan kaedah yang dicadang dengan algoritma konvensional yang digunakan SYCUT sebelum ini. Kaedah yang dicadangkan ini dapat mengesan imej yang berdarah dan terhalang dengan lebih tepat. Algoritma yang dibangunkan dapat meningkatkan prestasi pengesan esofagus daripada 68.65 ke 92.77 peratus. SYCUT dapat berfungsi dengan cekap walaupun dalam persekitaran kerja yang tidak terkawal kerana keberkesanan kaedah pembelajaran dalam yang dibangunkan. Walau bagaimanapun, kaedah pembelajaran dalam ini mempunyai batasan di mana ia memerlukan sejumlah data yang besar. Penyelidikan ini terbatas kepada penambahbaikan pengesan untuk jenis pemasukan imej tertentu sahaja iaitu imej jenis berdarah atau terhalang. Kerja di masa hadapan perlu melibatkan peningkatan kadar “precision-recall” dalam sistem dan pelaksanaan sistem masa nyata untuk pengesanan esofagus dalam persekitaran sebenar atau tiruan.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AP	-	Average Precision
CEIEC	-	Contrast Enhancement based on Intensity Expansion-Compression
CHE	-	Conventional Histogram Equalization
CLAHE	-	Contrast-Limited Adaptive Histogram Equalization
CNN	-	Convolutional Neural Network
COCO	-	Common Objects in Context
CPU	-	Central Processing Unit
CSV	-	Comma-Separated Values
DR	-	Detection Rate
ELIIOHTP	-	Enhancement of Low Illumination Images based on an Optimal Hyperbolic Tangent Profile
FCN	-	Fully Convolutional Network
FN	-	False Negative
FP	-	False Positive
FPN	-	Feature Pyramid Network
FPR	-	False Positive Rate
GPU	-	Graphics Processing Unit
GUI	-	Graphical User Interface
HDF5	-	Hierarchical Data Format version 5
HE	-	Histogram Equalization
HSI	-	Hue, Saturation, Intensity
IDE	-	Integrated Development Environment
IoU	-	Intersection over Union
IPOL	-	Image Processing On Line
mAP	-	Mean Average Precision
MD	-	Multi-Detection
P	-	Precision
PCA	-	Principal Component Analysis
R	-	Recall

R-CNN	-	Region-Based Convolutional Neural Network
ResNet	-	Deep Residual Neural Network
RGB	-	Red Green Blue
RM	-	Ringgit Malaysia
ROI	-	Region of Interest
SfM	-	Structure-from-Motion
SSD	-	Single Shot MultiBox Detector
SYCUT	-	Syariah Compliance Automated Chicken Processing System
TN	-	True Negative
TP	-	True Positive
USD	-	United States Dollar
UTM	-	Universiti Teknologi Malaysia
YCbCr	-	Luminance; Chroma: Blue; Chroma: Red
YOLO	-	You Only Look Once

LIST OF SYMBOLS

$	-	Dollar
I	-	Intensity of all pixels
c	-	Intensity transformation formula
α	-	Scaling constant for each pixel
n	-	Number of thresholds/classes
m	-	Mean
$\sum$	-	Sum

CHAPTER 1

INTRODUCTION

1.1 Introduction

1.2 Problem Background

The global Muslim population was 1.8 billion in 2011 and has been steadily increasing since then [1,2]. According to Zubairi and Abdul Ghani [3–5], the global halal market for food and non-food products is estimated to be worth USD 2.1 trillion per year, with USD 150 billion for halal food trade, and it is growing as a result of global trends and international initiatives. The halal market piques the interest of both Muslim and non-Muslim producers and consumers worldwide. In 2017, the Malaysian halal industry was valued at $30 billion, and it is expected to grow by 25% over the next five years [6]. Chicken is consumed at a rate of about 50 kg per capita per year, and it is the second-most popular meat among Malaysians [7]. Broiler farms produce approximately 840 million chickens per year [8]. Annual production of poultry eggs and meat ranges from RM 1.78 billion to RM 6.03 billion [9]. According to one of the SYCUT reports [10], a slaughterhouse can handle up to 80,000 chickens per day.

According to Md. Yusof [11], halal or lawful meat is meat obtained by slaughtering using Islamic-compliant methods. Consuming meat that has not been slaughtered in the name of Allah is considered unlawful and prohibited in Islam. The Islamic way of slaughtering an animal includes cutting the trachea, esophagus, carotid arteries, and jugular veins; using sharp tools; draining the blood; not causing pain to the animal. Not all slaughterhouses confirmed the use of halal slaughter methods, and the cost of highly automated chicken processing machines is prohibitively expensive for small and medium-sized businesses. To address these issues, a Syariah Compliance Automated Chicken Processing System (SYCUT) was developed [12]. The SYCUT vision inspection system, on the other hand, goes through several challenges. Previous research employs a low-pass filter (LPF) [13] and Otsu's thresholding technique [14] for image preprocessing and segmentation and uses the Viola-Jones object detection framework [15] to detect chicken esophagus in an image.

Previous studies show an unsatisfactory overall detection performance from the system, which is mainly caused by the occlusion of chicken blood or feathers that partially or completely cover the esophagus and the chicken neck is not fully facing the camera, making detection impossible. The overall performance of the system, which reflects the loss function in which an error of 1 is added to the overall error if the system fails to detect the esophagus in a chicken, is 68.65% on average across two experimental sites. The esophageal invisibility rate, which is mainly caused by occlusion and position problems, is an average of 27.55% at both experimental sites. Some images are problematic due to poor illumination, motion blur, and other forms of visual noise that affect the overall system accuracy.

Based on these previous findings, the improvements made in this study are on the software side to improve detection accuracy and overall performance. This study does not address potential solutions to the esophageal invisibility rate, which would necessitate hardware configurations such as multiple camera angles.

1.3 Problem Statement

The previous work report [16] addressed the main problems of poor performance for detecting partially occluded esophagus in images, as well as lighting variation and noise in the images. The existing method, which implements image preprocessing and conventional object detection methods, is unsuitable for system robustness, which is important for input images with varying quality and conditions in a real-world environment. As a result, shifting SYCUT from a conventional machine learning method to a deep learning method can solve the system's robustness requirement by eliminating the need to preprocess the image before detection. Object detection systems using deep learning methods have been shown to achieve good accuracy, especially concerning image datasets of varying conditions [17–20].

The SYCUT project does not have a properly organized dataset for system analysis and evaluation. The image and video datasets from two sites (Az-Zain and 3P) should be compiled and filtered. The images should then be categorized according to their problems, which include bloodied, blurred, dark, occlusion, and position problems. The different categories of problematic images aid in the result analysis in determining which problem in the images leads to poor detection results. This also addresses the low detection accuracy caused by certain problematic images by employing the right methods for resolution.

One of the current challenges faced by the system is the poor condition of the captured images, which leads to poor detection results [16]. This study focuses on and resolves problematic images such as bloodied and occluded images. As a result, a more accurate detection method for these problematic images is required. The main approach of deep learning can be used to improve accuracy.

The conventional method of detecting the esophagus in an image, as used in the previous work of SYCUT [21], is not the most accurate or computationally fast. Therefore, comparison and analysis of the two methods would shed more light on which method is better suited for detection. The implementation of deep learning is expected to be more effective in combating the accuracy and computational challenges.

1.4 Research Objectives

The following are the research objectives of this study:

a) To improve dataset organization and include image annotation for data leveling in the system.

b) To develop a more effective method for esophageal detection using RetinaNet and Mask R-CNN deep learning.

c) To compare and analyze the use of conventional and deep learning methods for esophageal detection.

1.5 Research Scope

This study examines the challenges encountered when applying esophageal detection to a real-world dataset. Some of the examples of problematic images in the acquisition step are dark, bloodied, and occluded images. Clear, bloodied, and occluded images were used to test the robustness of the deep learning method. The potential solutions to these challenges were tested and discussed. Due to the lack of a suitable GPU for real-time system testing, the study was conducted using still images. The implementation in this study can be translated into real-time object detection in a poultry processing plant environment.

The research is limited to using pre-existing datasets from previous work where a site or new slaughtered chicken samples are not currently accessible. The experiments were conducted using software methods that, while unable to solve some esophageal invisibility problems, focused on the partially occluded dataset. The dataset was divided into two parts: the control dataset, which contains clear images; and the bloodied and occluded dataset, which was used to test the robustness and ability of the deep learning method to detect occluded esophagus.

1.6 Research Significance

This study focuses on problem-solving when dealing with images taken in a challenging environment. It emphasizes the use of deep learning to improve esophageal detection and seeks an alternative to using conventional methods to detect a specific object. The research is significant in terms of developing a good detection platform for the detection of the esophagus or other similar-looking objects. The research also contributes to the improvement of SYCUT. The findings of this study may be worth further investigation and trigger better ideas and implementations in the field of image processing.

1.7 Thesis Organization

This chapter describes the previous work on the SYCUT project. The significance of this research is to improve esophageal detection and to discover methods to reduce computational cost. The conventional method applied in the system uses Haar features to detect objects. The system’s issues are classified as bloodied, blurred, dark, occlusion, position, and invisible esophagus. The research objectives were formed in response to the problems’ requirements. This research is valuable in the image processing and object detection fields, and it has the potential to be implemented in similar areas to SYCUT.

The thesis is organized as follows:

a) Chapter 2 reviews the literature on certain topics related to the scope of this study. Topics such as deep learning, occluded image detection, and problematic image preprocessing are discussed in this chapter.

b) Chapter 3 describes the methodology and research flow that were used throughout the whole study. It explains how the research was conducted, the information on deep learning, how the dataset was prepared, and what valuables were observed in the evaluation and analysis of the system.

c) Chapter 4 presents the results and discussion based on the experiments conducted in the study. The analysis and findings are shown in this chapter.

d) Chapter 5 summarizes and concludes the research, as well as discusses the study’s limitations and potential future work based on the findings.

REFERENCES

[1] Bohari AM, Hin CW, Fuad N. The competitiveness of halal food industry in Malaysia: A SWOT - ICT analysis. *Geogr Online Malaysia J Soc Sp,* 2013. 9: 1–9.

[2] Dierks. Market Watch 2010 – The Food Industry. *Malaysian-German Chamber of Commerce & Industry*. 2011.

[3] Zubairi SI, Abdul Ghani M. Halal Food : Scenario and Current Status in Malaysia (PROCEEDING). *10th ASEAN Food Conf. 2007,* 2007.

[4] Abdul Latif M. *Application for Halal certifications*. 2007.

[5] Abdul Latif M. *Current Issues on Halal Food*. 2006.

[6] Ab Hamid N 'Adha, Mohd Shahwahid F, Othman N, Saidpudin W. Challenges and Ways Improving Malaysia Halal Food Industry. *Spec Issue SciInt(Lahore)* 2017. 29: 149–153.

[7] Eddie C, Bahri H, Jamil S. Do you know ... about Malaysia's poultry industry? *The Star*. 2020.

[8] Ferlito C. *The Poultry Industry and Its Supply Chain in Malaysia: Challenges from the Covid-19 Emergency*. 2020.

[9] Ashraf A, Abd Rahman F, Abdullah N. Poultry Feed in Malaysia: An Insight into the Halalan Toyyiban Issues. *Proc. 3rd Int. Halal Conf. (INHAC 2016),* Singapore: Springer Singapore. 511–531; 2016.

[10] Ahmad Sayuti MFZ, Muhd Yunus MR. *Technical Report - 3P Pandan Slaughter House*. 2014.

[11] Md.Yusof F. *The modern methods of animal's slaughtering in Malaysia : an Islamic perspective*. 2005.

[12] Selamat H. *Technofund Technical Report on Syariah-Compliance Automated*

Chicken Processing Machine for Small and Medium Sized Industries. Kuala Lumpur: 2011.

[13] Makandar A, Halalli B. Image Enhancement Techniques using Highpass and Lowpass Filters. *Int J Comput Appl,* 2015. 109: 21–27.

[14] Otsu N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans Syst Man Cybern,* 1979. 9: 62–66.

[15] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. *Proc. 2001 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognition. CVPR 2001*. IEEE Comput. Soc. 2001. p. I-511-I-518.

[16] Arfa R. *SYCUT Vision Inspection Technology (SVIT) Analysis and Future Works*. 2015.

[17] Selmi Z, Halima M Ben, Pal U, Alimi MA. DELP-DAR system for license plate detection and recognition. *Pattern Recognit Letters,* 2020. 129: 213–223.

[18] Namjoshi M, Khurana K. A Mask-RCNN based object detection and captioning framework for industrial videos. *Int J Adv Technol Eng Explor*, 2021. 8(84): 1466–1478.

[19] Dong W, Roy P, Peng C, Isler V. Ellipse R-CNN: Learning to Infer Elliptical Object from Clustering and Occlusion. *IEEE Trans Image Process*, 2020. 30: 2193–2206.

[20] Wu Q, Feng D, Cao C, Zeng X, Feng Z, Wu J, et al. Improved mask r-cnn for aircraft detection in remote sensing images. *Sensors*, 2021. 21(8).

[21] Arfa R. *SYCUT Vision Inspection Technology (SVIT) Technical Report*. 2015.

[22] Department of Islamic Development Malaysia (JAKIM). *Malaysian Protocol for the Halal Meat & Poultry Production*. 2011.

[23] Freund Y, Schapire RE. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *J Comput Syst Sci*, 1997. 55: 119–139.

[24] Cortes C, Vapnik V. Support-vector networks. *Mach Learn,* 1995. 20: 273–297.

[25] Lee SL, Tseng CC. Color image enhancement using histogram equalization method without changing hue and saturation. *2017 IEEE Int Conf Consum Electron - Taiwan, ICCE-TW 2017,* 2017. 305–306.

[26] Chi Liu S, Liu S, Wu H, Arifur Rahman M, Ching-Feng Lin S, Yeow Wong C, et al. Enhancement of Low Illumination Images based on an Optimal Hyperbolic Tangent Profile. *Comput Electr Eng,* 2018. 70: 538–550.

[27] Xu Z, Wu H, Yu X, Qiu B. Colour image enhancement by virtual histogram approach. *IEEE Trans Consum Electron,* 2010. 56: 704–712.

[28] Tian Q-C, Cohen LD. Global and Local Contrast Adaptive Enhancement for Non-uniform Illumination Color Images. *2017 IEEE Int. Conf. Comput. Vis. Work.*, IEEE, 2017. 3023–3030.

[29] Pizer SM, Amburn EP, Austin JD, Cromartie R, Geselowitz A, Greer T, et al. Adaptive histogram equalization and its variations. *Comput Vision, Graph Image Process,* 1987. 39: 355–368.

[30] Liu S, Rahman MA, Lin CF, Wong CY, Jiang G, Liu SC, et al. Image contrast enhancement based on intensity expansion-compression. *J Vis Commun Image Represent,* 2017. 48: 169–181.

[31] 2vin. *Color Correction.* 2018.

[32] Fernando S. *OpenCV Tutorial C++ Histogram Equalization*. 2018. https://www.opencv-srf.com/2018/02/histogram-equalization.html.

[33] HYPJUDY. *Histogram Equalization on Grayscale and Color Images*. 2017.

[34] HYPJUDY. [DIP] *Histogram Equalization on Grayscale and Color Image*. 2017. https://hypjudy.github.io/2017/03/19/dip-histogram-equalization/.

[35] Lisani J-L. An Analysis and Implementation of the Shape Preserving Local Histogram Modification Algorithm. *Image Process Line,* 2018. 8: 408–434.

[36] Zuiderveld K. Contrast Limited Adaptive Histogram Equalization. *Graph*

Gems, 1994. IV: 474–485.

[37] Shan Q, Jia J, Agarwala A. High-quality motion deblurring from a single image. *SIGGRAPH'08 Int Conf Comput Graph Interact Tech ACM SIGGRAPH 2008 Pap 2008*, 2008. 27(3).

[38] Oliveira JP, Figueiredo MAT, Bioucas-Dias JM. Parametric blur estimation for blind restoration of natural images: Linear motion and out-of-focus. *IEEE Trans Image Process*, 2014. 23: 466–477.

[39] Lee JE, Kwak N. Detection and recovery of occluded face images based on correlation between pixels. *ICPRAM 2012 - Proc 1st Int Conf Pattern Recognit Appl Methods*, 2012. 2: 568–571.

[40] Qiu S, Wen G, Fan Y. Occluded Object Detection in High-Resolution Remote Sensing Images Using Partial Configuration Object Model. *IEEE J Sel Top Appl Earth Obs Remote Sens*, 2017. 10: 1909–1925.

[41] LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-Based Learning Applied to Document Recognition. *Proc IEEE*, 1998. 1–46.

[42] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. *2016 IEEE Conf. Comput. Vis. Pattern Recognit.*, IEEE. 2016. 770–778.

[43] Tian H, Zheng Y, Jin Z. Improved RetinaNet model for the application of small target detection in the aerial images. *IOP Conf Ser Earth Environ Sci*, 2020. 585(1).

[44] Hoang T, Nguyen P, Truong N, Lee Y, Park K. Deep RetinaNet-Based Detection and Classification of Road Markings by Visible Light Camera Sensors. *Sensors*, 2019. 19(2): 281.

[45] Redmon J, Divvala S, Girshick R, Farhadi A. You Only Look Once: Unified, Real-Time Object Detection. *2016 IEEE Conf. Comput. Vis. Pattern Recognit.*, IEEE. 2016, 779–788.

[46] Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C-Y, et al. SSD: Single Shot MultiBox Detector. *Lect. Notes Comput. Sci. (including Subser. Lect.*

Notes Artif. Intell. Lect. Notes Bioinformatics). 21–37; 2016.

[47] Lin T-Y, Goyal P, Girshick R, He K, Dollar P. Focal Loss for Dense Object Detection. *2017 IEEE Int. Conf. Comput. Vis.*, IEEE. 2017. 2999–3007.

[48] Girshick R. Fast R-CNN. *Proc IEEE Int Conf Comput Vis*, 2015. 2015 Inter: 1440–1448.

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

[50] He K, Gkioxari G, Dollar P, Girshick R. Mask R-CNN. *2017 IEEE Int. Conf. Comput. Vis.*, IEEE. 2017. 2980–2988.

[51] Redmon J, Farhadi A. *YOLOv3: An Incremental Improvement.* 2018.

[52] Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger. *Proc - 30th IEEE Conf Comput Vis Pattern Recognition, CVPR 2017*, 2016. 2017-January: 6517–6525.

[53] Fu C-Y, Shvets M, Berg AC. RetinaMask: Learning to predict masks improves state-of-the-art single-shot detection for free. 2019.

[54] Lin T-Y, Dollár P, Girshick R, He K, Hariharan B, Belongie S. Feature Pyramid Networks for Object Detection. Proc. - *30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, 2017, 1–10.

[55] Lin T-Y, Maire M, Belongie S, Bourdev L, Girshick R, Hays J, et al. Microsoft COCO: Common Objects in Context. *Lect Notes Comput Sci (Including Subser Lect Notes Artif Intell Lect Notes Bioinformatics)*, 2014. 8693 LNCS(PART 5): 740–755.

[56] Gené-Mola J, Sanz-Cortiella R, Rosell-Polo JR, Morros J-R, Ruiz-Hidalgo J, Vilaplana V, et al. Fruit detection and 3D location using instance segmentation neural networks and structure-from-motion photogrammetry. *Comput Electron Agric*, 2019. 169(December 2019): 105165.

[57] Westoby MJ, Brasington J, Glasser NF, Hambrey MJ, Reynolds JM. 'Structure-from-Motion' photogrammetry: A low-cost, effective tool for geoscience applications. *Geomorphology,* 2012. 179: 300–314.

[58] Yu Y, Zhang K, Yang L, Zhang D. Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN. *Comput Electron Agric,* 2019. 163(February): 104846.

[59] Li Y, Kuang Z, Chen Y, Zhang W. Data-Driven Neuron Allocation for Scale Aggregation Networks. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognition,* 2019. 2019-June: 11518–11526.

[60] Chahyati D, Arymurthy AM. Multiple Human Tracking Using Retinanet Features, Siamese Neural Network, and Hungarian Algorithm. *IJMET_10_05_047 Int J Mech Eng Technol,* 2019. 10(5): 465–475.

[61] Yan M, Wang J, Yang Z, Li J. Improved real-time joint object detection and road segmentation multi-task network. *2021 6th Int. Conf. Autom. Control Robot. Eng.,* IEEE. 2021. 541–545.

[62] Kapania S, Saini D, Goyal S, Thakur N, Jain R, Nagrath P. Multi Object Tracking with UAVs using Deep SORT and YOLOv3 RetinaNet Detection Framework. *Proc. 1st ACM Work. Auton. Intell. Mob. Syst.,* New York, NY, USA: ACM. 2020, 1–6.

[63] Russell BC, Torralba A, Murphy KP, Freeman WT. LabelMe: a database and web-based tool for image annotation. *Int J Comput Vis,* 2008. 77(1-3): 157–173.

[64] Torralba A, Russell BC, Yuen J. LabelMe: online image annotation and applications. *Proc IEEE,* 2010. 98(8): 1467–1484.

[65] Everingham M, Zisserman A, Williams C, Van Gool L. *The PASCAL Visual Object Classes Challenge 2006 (VOC2006) Results*. vol. 2006. 2006.

[66] Wada K. *labelme: Image Polygonal Annotation with Python*. 2016. https://github.com/wkentaro/labelme.

[67] Deng J, Dong W, Socher R, Li L-J, Kai Li, Li Fei-Fei. ImageNet: A large-Scale Hierarchical Image Database. *2009 IEEE Conf. Comput. Vis. Pattern Recognit.*, IEEE. 2009. 248–255.

[68] Fizyr. *Keras MaskRCNN*. 2018. https://github.com/fizyr/keras-maskrcnn.

[69] Gaiser H, Liscio E, vcarpani. *fizyr/keras-maskrcnn: 0.2.0*. 2018.

[70] Agarap AF. Deep Learning using Rectified Linear Units (ReLU). *ArXiv Prepr ArXiv180308375*, 2018.

[71] Wasser LA, Webb E. Hierarchical Data Formats - What is HDF5. *NeonscienceOrg*. 2018. https://www.neonscience.org/about-hdf5.

[72] Kumar G. Evaluation Metrics for Intrusion Detection Systems-A Study. *Int J Comput Sci Mob Appl*, 2014. 2: 11–17.

[73] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. *J Mach Learn Res*, 2011. 12: 2825–2830.

[74] Manning CD, Raghavan P, Schütze H. Evaluation in information retrieval. *Introd. to Inf. Retr.*, Cambridge University Press. 2008. 151–175.

[75] Gad AF. *Evaluating Object Detection Models Using Mean Average Precision (mAP)*. 2020. https://blog.paperspace.com/mean-average-precision/.

[76] Hui J. *mAP (mean Average Precision) for Object Detection*. 2018. https://jonathan-hui.medium.com/map-mean-average-precision-for-object-detection-45c121a31173.

[77] Lin T-Y, Maire M, Belongie S, Bourdev L, Girshick R, Hays J, et al. *COCO dataset*. 2014. https://cocodataset.org/#detection-eval.

[78] Everingham M, Winn J. *The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Development Kit*. 2012.

[79] Google. Descending into Machine Learning. *Mach. Learn. Crash Course*. 2021st ed., n.d.