JSCDM

http://penerbit.uthm.edu.my/ojs/index.php/jscdm

e-ISSN : 2716-621X

Journal of Soft Computing and Data Mining

Child Detection Model Using YOLOv5

Azrina Tahir[1,2*], Shamsul Kamal Ahmad Khalid[1], Lokman Mohd Fadzil[3]

[1]Faculty of Computer Science and Information Technology,
Universiti Tun Hussein Onn Malaysia, Parit Raja, 86400 Batu Pahat, Johor, MALAYSIA

[2]Department of Information Technology and Communication,
Politeknik Balik Pulau, Pinang Nirai, Mukim 6, 23000 Balik Pulau, Pulau Pinang, MALAYSIA

[3]National Advanced IPV6 Center of Excellence,
Universiti Sains Malaysia, 11800 USM, Pulau Pinang, MALAYSIA

*Corresponding Author

DOI: https://doi.org/10.30880/jscdm.2023.04.01.007
Received 14 April 2023; Accepted 11 May 2023; Available online 25 May 2023

Abstract: Closed-circuit television (CCTV) surveillance systems have been installed in public locations to search for missing children and fight crime. The Penang City Council has deployed a face recognition CCTV monitoring system. As a result, this research aims to identify children who were in the wrong area or at the wrong time and then notify authorities such as police and parents. According to child detection research, the average child loss rate is more significant due to lacking child detection features. Existing research employs machine learning and deep learning across several platforms, yielding inaccurate accuracy findings. Using the YOLOv5 algorithm, this study will categorize images based on children's detection in restricted locations. Coco, Coco128, and Pascal VOC were chosen because they are the standard datasets of YOLOv5 and the public dataset INRIA Person. Annotations and augmentation techniques are employed in the pre-processing phase to acquire labeling in text file format and offer data for any object position. The YOLOv5s model will then be designed to make the proposed detector model. After training using YOLOv5s, a child detector model is produced and evaluated on the dataset to acquire findings according to recall, precision, and mean average precision (mAP) performance metrics. Finally, the performance metrics obtained from all four datasets are compared. The INRIA Person dataset performed the best, with a recall of 0.995, an accuracy of 0.998, and a mean Average Accuracy of 0.995. Nevertheless, the findings for both YOLOv5s and the proposed model are pretty close. This demonstrates that the proposed model can detect as well as the YOLOv5s model.

Keywords: Image classification, object detection, deep learning, YOLOv5, child detector model

1. Introduction

With computer vision's growing potential, many companies have spent on object detection to interpret and analyze data primarily derived from visual sources for a wide range of applications, such as medical image classification, automated vehicle object recognition, face recognition for security reasons, and child detection in surveillance systems. Most research in the field of child detection involves prediction models [1], head and body ratio calculation [2], age group classification [3], and child versus adult classifier [4]. This distinction between children and adults improves social security since it makes it simpler to recognize criminals based on identifiable physical features. Several studies on the distinction among children and adults were presented. These investigations only apply to face characteristics or portions of the body and employed cameras with a near range of 20 to 50 cm [5]. Previous research has revealed why

*Corresponding author: azrinatahirofficial@gmail.com

most video monitoring systems employ CCTV cameras, which are insufficient for preventing criminal incidents. As a result, long distance classification study was carried out [4]. These investigations employ distances ranging from two to 10 metres from a CCTV camera coupled with biometric data. Adults are diagnosed at a 64.5% rate, while children are detected at a 100% rate. One of the primary reasons is that some people's physical features diverge from the typical ratio of their age group. A child is a young person who, as defined by the United Nations, is anybody under the age of 18. In computer vision, a best image to detect a child is between the ages of 0 and 12.

The main issue in this study is that the average child loss rate is greater based on the top 24 approaches of the Caltech Pedestrian Detection Benchmark due to the lack of child detection features. The second issue is that existing research using machine learning and deep learning to classify videos and images from different media such as the web, social media, CCTV cameras, speech, advertising, remote sensing, and disease diagnosis can lead to irrelevant accuracy results. A third issue is that most existing research presents results with accuracy metrics. However, there are other less published metrics. Based on the issues raised, the next research activities are related to child detection which involves investigating significant features of children, choosing classification algorithm design at appropriate platforms and performance metric classification is limited.

The body of this paper is structured as follows: Section 2 elaborates on the review of associated works. Section 3, will cover the research methodology. The experimental results, discussion and comparison with the existing model are presented in Section 4. Finally, Section 5 offers conclusions and recommendations for future works.

2. Related Works

Computer vision is a field of artificial intelligence (AI) that enables computers and systems to derive meaningful information from digital images, videos and other visual inputs. An image is represented on a computer by a matrix of pixel colors determined by integers. Numbers 0 to 255 represent pixel colors ranging from white to black. Mathematical calculations can tell something about an image by assigning numerical values to pixels. Then, computer vision will act or make recommendations based on that information. This is where image classifiers look at images and predict which images are assigned to a particular class. Object detection, on the other hand, uses image classification to identify certain image classes and then detects and locates the position of the object in the image or video. This section will discuss more on image classification, object detection, YOLOv5, annotation data and training data also results from previous research on YOLOv5.

2.1 Image Classification

In the study of image classification, there are various techniques and different classification models that have been used, from extracting image features to classifying them into different groups. Both unsupervised and supervised methods have been used in previous studies such as machine learning and deep learning algorithms. This method is used based on the selected dataset and algorithm. Some studies on child versus adult classification have been done in speech recognition [6], but only some in image classification [5]. Existing research uses machine learning and deep learning algorithms that provide solutions for classifications such as age, gender and emotion [7]. The summary of the previous research is given in Table 1.

Table 1 - Summary of previous research in image classification

No	References	Area	Dataset	Classification Algorithm
1.	Kumar & Kumar Agarwal [9]	Age classification	Dataset of University of Tuebingen, Germany	KNN & SVM
2.	Agarwal & Jain [3]	Child classification and underage detection	Benchmark dataset of face Photos from the Open University, Israel.	KNN, ANN and SVM
3.	Das et al. [13]	Gender, age, and ethnicity classification	UTKFace dataset and BEFA challenge dataset	Multi-Task Convolution Neural Net-work (MTCNN)
4.	Nagpal et al. [15]	Adult expression classification	Radboud faces dataset and CAFE	Mean Supervised Deep Boltzmann Machine classification
5.	Garcia et al. [16]	Children and adult classification	INRIA database	BHR, BLHR, BNLHR and GF

2.2 Object Detection

The object detection model architecture is monopolized by R-CNN family, YOLO family and SSD family. Surya Remanan [17] compares these three models in several aspects such as speed, accuracy and prediction. Among the three models in terms of speed, YOLO is very fast and uses very little processing memory. SSD is also fast, but it slows down when many convolution layers are involved, While R-CNN is slower than YOLO and SSD because it requires

large storage and processing power for detection. From the aspect of accuracy, R-CNN is the most accurate in its detection results compared to SSD and YOLO. SSD got more accurate results when compared to YOLOv1 however, YOLOv3 and YOLOv5 outperformed SSD in accuracy and speed. The prediction aspect shows that YOLO can predict only 1 class per grid and faces difficulties when it comes to multiple objects and small object detection. SSD, on the other hand, can handle multi-scale objects because it uses feature maps from all convolutional layers, and each layer operates at a different scale, but it also has difficulty detecting small objects. R-CNN is the most accurate in terms of detection. Each model has advantages and disadvantages as discussed. YOLO is the best bet if it has an object that is easy to detect. If object detection involves complex images such as cancer detection from X-rays, then R-CNN is the most suitable.

2.3 YOLOv5

YOLO stand for "You Only Look Once" is one of computer vision model that employ from Convolutional Neural Network (CNN) use to detect objects in real-time. YOLO algorithm is a single-stage deep learning object detection technique [18]. While faster-RCNN is a two-stage network [19]. The first YOLO is developed by Joseph Redmon in a darknet framework. Three key points of YOLO is extremely fast, sees the entire image to make predictions and learns generalizable representation of objects.

Yolov5 using PyTorch framework and python libraries. Yolov5 uses input images to create features. Following the feeding of these features through a prediction, a bounding box is drawn around the objects to predict their classes. YOLOv5 is the most recent version of the YOLO object detection system, which has excellent detection accuracy and is fast and efficient in real-time. YOLOv5 contains four models of YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x, with the smallest volume being YOLOv5s.

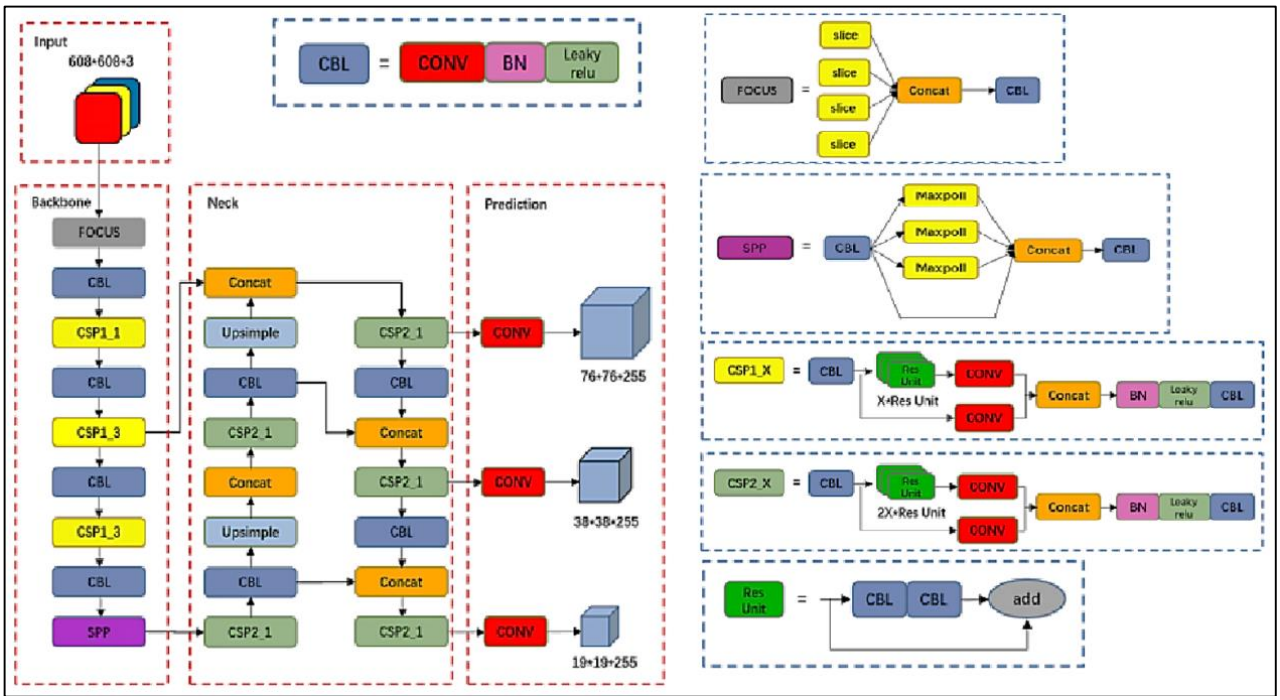

Fig. 1 - Yolov5 Architecture [18]

From Fig. 1, the Yolov5 model consists of four general modules specifically including input side, backbone, neck and head (prediction). The architecture also includes six components like CBL, Res Unit, CSP1_x, CSP2_x, Focus and SPP. The input is an image with a size of 608 * 608 that will go through the pre-processing phase. YOLOv5 effectively improves the detection of small targets on the input side by using the Mosaic data enhancement and CutMix approach. The purpose is to increase the network's data processing capacity and add adaptive scaling processing. The image is then uniformly scaled to a uniform size and fed to the learning network. Backbone module is used to extract some generic feature representations and it is a classifier network with excellent performance. The backbone consists of Focus structures, CSP networks, etc. The Focus structure converts the original 608 * 608 * 3 image into 304 * 304 * 32 feature maps by four-slice operations and one convolution of 32 convolutional cores. This structure first connects multiple slice results and then sends them to the CBL module. CBL module consists of convolutional, bayes naïve and leaky relu activation function. It will Utilizing the feature data from many layers, CSPNet performs local cross-layer fusion to produce richer feature maps. CSPNet module consists of CBL, Res unit, convolutional layer, concat, bayes naïve and leaky relu. Res unit module is borrowed from the residual structure in the Res network. It used to build a

deep network and CBM is a sub-module in the residual module. Both PANet and SPP are present in the Neck region. To properly integrate the image features from various layers, PANet (PathAggregation) aggregates the output features from various CSP networks in top-down order, followed by the shallow features from the bottom-up. For maximum pooling, SPP (space pyramid pooling) employs four different-sized nuclei, followed by tensor splicing. The maximum pooling of 1 * 1, 5 * 5, 9 * 9 and 13 * 13 is used for multi-scale feature fusion. The model head is utilized to complete the target detection results by performing the final stage operation. It makes use of the anchor box on the feature map to generate the final output of class, object score, and bounding box.

2.4 Annotation and Training Data

Data annotation is the process of labelling predetermined classes to images or videos. There are several annotations used for object detection such as bounding boxes, 3D cuboids, polygons, lines and splines as well as semantic segmentation. YOLOv5 uses bounding boxes annotation. The output to this annotation is one object annotation file for each image or video dataset. This file follows a format that YOLOv5 can use. For example, COCO uses JSON format while Pascal VOC uses XML for default annotation file format. However, data annotation can be changed by re-annotating using software such as Roboflow.

Training data is data used to train object detection algorithms. Training, validation and testing refers to the initial set of data given to any model from where the model was created. Training data contains pairs of input images and annotation files to train the model to perform tasks to obtain accurate results.

2.5 Results and Discussion of Previous Research

There is previous research on object detection using Yolov5 such as flower image classification method [18], face mask recognition [19], breast tumor detection [20] and face detection [21].

Table 2 - Summary of previous research in image classification

No	Research	Model	Results
1.	Tian & Liao [18]	Yolov5s	mAP - 0.959
2.	Yang et al. [19]	Yolov5s	mAP - 0.979
3.	Mohiyuddin et al. [20]	Breast tumor detector	mAP - 0.965
4.	Fahad Majeed et al. [21]	Face detector	mAP - 0.94

Tian & Liao [18] conducted image classification research on flowers by detecting the type of flower in each picture using Yolov5s model as an object detection model. This research uses precision, recall and mAP performance metrics to measure performance. The results showed that precision was 0.942, recall was 0.933 and mAP was 0.959.

The second study develop a face mask recognition system using the Yolov5 model with the aim of replacing the manual inspection of the use of face masks by the Chinese government to its citizens by using Yolov5 [19]. This study compares the Yolov5 object detection model with three other models namely Faster R-CNN, R-FCN and SDD by using precision as a performance metric. The result was Faster R-CNN got 70.40%, R-FCN got 77.60%, SDD got 84.60% and Yolov5 got the highest percentage of 97.90%.

Mohiyuddin et al. [20] research, propose a modified model of Yolov5 to detect and classify breast tumour using public dataset Curated Breast Imaging Subset of DDSM (CBIS-DDSM). In the pre-processing phase, image enhancing techniques, removal of pectoral muscles and labels is performed. Then, dataset is annotated, augmented and divide into training, validation and testing ratio 60%, 30% and 10%. The experiment was conducted according to the cluster size of 8, the learning rate of 0.01, the momentum of 0.843, and the epoch value of 300. A comparison was made with the YOLOv3 and RCNN models using the mAP performance metric, it was found that the proposed model showed better performance with a score of 96%, while YOLOv3 and RCNN got 85 % and 84.2% respectively. Overall, the proposed model successfully identifies and classifies breast tumours and outperforms previous research results.

Research Fahad Majeed et al. [21] is one of the research projects that implemented face recognition in surveillance system applications based on deep learning algorithms. Identifying multiple occurrences in a real-time environment is very important because of the difficult and heterogeneous environmental conditions and their blocking effects. The YOLOv5 model was chosen to investigate the efficiency of the surveillance system with very limited experimental analysis. The datasets used are the Face Detection Dataset & Benchmark (FDDB), the Celebrity Face Recognition Dataset (CFR) and personal datasets taken from runtime video streams for training and testing data. Attempts were made on the proposed model YOLOv5 and two other models YOLOv3 and YOLOv4. A comparison of the results was made using mAP performance metrics with scores of YOLOv3 87%, YOLOv4 89% and YOLOv5 94%. The analysis shows that the YOLOv5 algorithm has produced better results than the previous studies YOLOv4 and YOLOv3 respectively.

Findings from studies that have been conducted found that most of the results shows that the existing or modified object detector model produced accurate results which is above 90 % using existing or modified model of YOLOv5.

3. Methodology

The child detector framework consists of four phases involved data collection, pre-processing, classification and detection also evaluation, see Fig. 2.

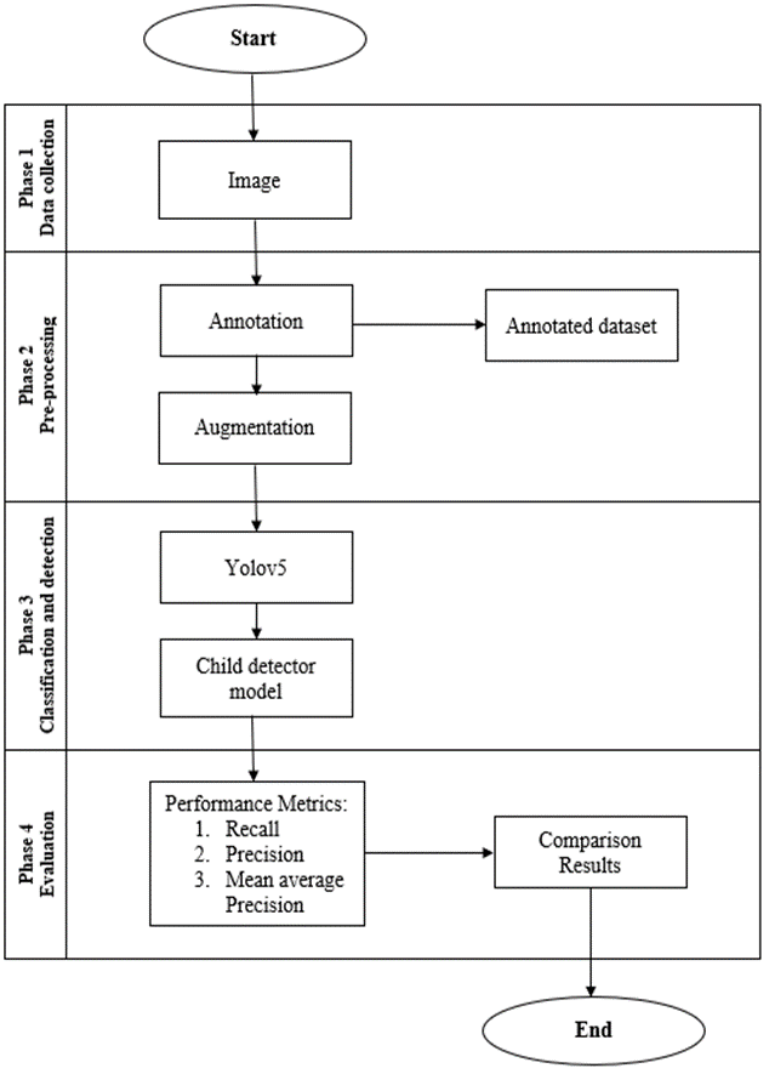

Fig. 2 - Child detector framework

3.1 Data Collection

Four dataset is used in this research COCO, COCO128, Pascal VOC and INRIA Person.

Table 3 - Datasets information

No.	Dataset	Dataset Size	No of Class
1.	COCO	330,000 images	80
2.	COCO128	128 images	80
3.	Pascal VOC	17,125 images	20
4.	INRIA Person	902 images	2

3.2 Pre-Processing

Pre-processing is an essential step in the preparation of a dataset. For the analysis to proceed properly, the dataset needs to be annotated. However, COCO, COCO128 and Pascal VOC has existing annotation and labelling file but still need to reannotate due to existing class do not consist of child and adult class. Annotation is the process of creating a bounding box and class labelling. Each dataset must be annotated to obtain five important variable that will be used to

get the location of the class. There is various online software that can be used to annotate the dataset and generate the label file automatically.

Then, augmentation process will be done automatically in YOLOv5 using default technique to prepare the image in any kind of situation such as scaling, blur image, object position rotation and flipping image horizontally and vertically.

3.3 Classification and Detection

The detection and classification process in involves training, validation and testing performed on datasets using the YOLOv5s model and the proposed child detector model. The first step is to divide the dataset into train, valid and test according to the ratio of 70%, 20% and 10%.

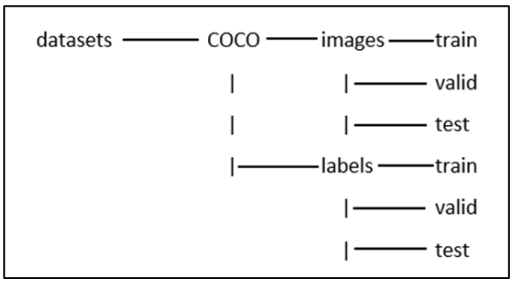

Fig. 3 - Directories structure for COCO dataset

Non-Max Suppression algorithm is used in detection and classification process as follow:

Algorithm 1: Non-Max Suppression

Input: Training images with bounding box and labels
Output: The best bounding box

```
function NMS (BB, cl)
    BB_nms ← ∅
    for bb_i ∈ BB
        discard ← false
        for bb_j ∈ BB
            if similar(bb_i, bb_j) > γ_nms
                if score(cl, bb_j) > score(cl, bb_i)
                    discard ← true
        if not discard
            BB_nms ← BB_nms ∪ bb_i
    return BB_nms
```

The Non-Max Suppression is the technique used to select the best bounding box process. Training images with bounding box and labels is the input. The bb_i and bb_j is box detected from bounding box. This box is selected based on the highest objectiveness score. If bb_i and bb_j box get the same IOU, score for each box will be compared. Bounding boxes that get IOUs more than 50% will be removed and the other box will be added to the list. Same process will be repeated for remaining boxes and the final list of BB_{nms} is the output.

3.4 Classification and Detection

The performance metrics used to measure the Yolov5 results are recall, precision and mean Average Precision (mAP). Precision is defined as the ratio of True Positives to all the positives predicted by the model, divided by total number of True Positives and False Positive as in Equation 1. The greater the number of false positives predicted by the model, the lower the accuracy and the lower the precision.

$$P = \frac{TP}{(TP+FP)} \tag{1}$$

Recall is the ratio of True Positives to all the positives predicted divided by total of True Positive and False Negative as shown in Equation 2. Low recall means the higher the number of False Negatives predicted by the model.

$$R = \frac{TP}{(TP+FN)} \quad (2)$$

The mAP (mean Average Precision) is the total of Average precision for each class divide by total class as shown in Equotion 3.

$$mAP = \frac{1}{N} \sum_{i=1}^{N} AP_i \quad (3)$$

Each score obtained based on the performance metric accuracy, recall and mAP results of the four datasets will be evaluated and compared to obtain the best detection results.

4. Results and Discussion

The result of this research is describing in this section.

4.1 Implementation Tools

The experiments are performed on a machine with a 12th Gen Intel® Core™ i7-12700H CPU @ 2.30Hz processor, 16 GB RAM and a NVDIA GeForce RTX 3050 Laptop GPU 4GB RAM. Visual Studio Code 1.74.1 and Python 3.9 has been used for the complete implementation of the proposed model. Yolov5 use Python as the programming language because its offers extended functionality with a lot of libraries such as NumPy and PyTorch. Another requirement that has been installed is NVDIA CUDA 11.7 used with PyTorch to optimize the GPU.

4.2 Data Collection

This research needs to identify child and adult which is not in 80 classes; therefore, these two classes must be added from existing number of classes in yaml file as number 80 and 81. The annotation of child and adult must be done before the classification and detection to replace the existing annotation file. Roboflow will be used to annotate the image and generate the dataset version. The version consists of dataset image and a text file with annotation that automatically divide into training, validation and testing. However, due to smaller size of Coco128, dataset partitioning is ignored by using the entire dataset for training and validation. Yolov5 obtained the results from training, validation and detection in real-time. The process is repeatedly for Coco and Pascal VOC dataset.

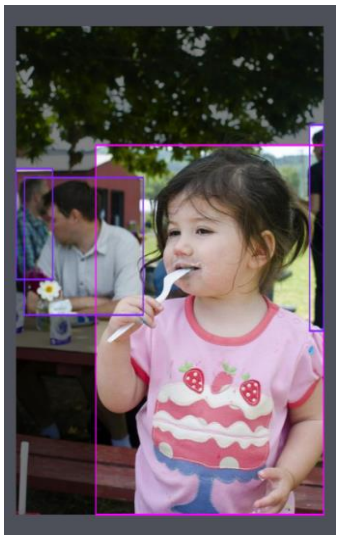

Fig. 4 - Annotated image

```
81 0.48478 0.39912 0.03503999999999997 0.1884533333333334
80 0.29564 0.49074666666666666 0.21339999999999998 0.6028533333333334
80 0.87788 0.43984 0.17122000000000004 0.5588000000000001
80 0.60226 0.382 0.06666000000000008 0.24754666666666664
80 0.57394 0.37666666666666665 0.04333999999999992 0.21909333333333328
80 0.45462 0.37666666666666665 0.044700000000000045 0.2013066666666667
80 0.51278 0.3873333333333333 0.06435999999999996 0.20487999999999995
80 0.149 0.3624266666666667 0.053459999999999994 0.15506666666666663
80 0.2105 0.3535466666666667 0.039779999999999975 0.15506666666666663
80 0.5307799999999999 0.35218666666666665 0.033680000000000064 0.14351999999999998
```

Fig. 5 - List of labelling in text format

Each labelling consists of class, bounding box top left x coordinate, bounding box of top left y coordinate, bounding box width and bounding box height. Augmentation process will be automatically done when Yolov5 is running. Augmentation in Yolov5 involve scaling, transforming, rotating and flipping to predict the object in any kind of position.

4.3 Classification and Detection

Classification and detection process will train, validate and test the four datasets with YOLOv5s model and model proposed child detector model. The data from train is used to train the pretrained model with a few configurations.

4.4 Result

The dataset must typically be separated into parts, or batches, because it cannot usually be fed entirely into the neural network at once. The batch size describes how many training samples are included in a single batch. Epoch refers to the single forward and backward feed of the training data through the neural network. The file use to train is train.py with an image size 640, 16 batch of dataset for the whole dataset, epochs is 300 and weight is pretrained model selected yolov5s.pt.

Table 4 - Performance results of YOLOv5s model

Dataset	Precision	Recall	mAP_0.5
COCO128	0.981	0.965	0.993
COCO	0.996	0.976	0.994
Pascal VOC	0.992	0.983	0.994
INRIA Person	0.998	0.995	0.995

The results from training four dataset are shown in Table 4 performance results of Yolov5s model. The COCO128 dataset obtained precision 0.981, recall 0.965 and mAP is 0.993. The result for COCO dataset is 0.996 for precision, 0.976 for recall and mAP 0.994. Pascal VOC has precision 0.992, recall 0.983 and mAP 0.994. The INRIA Person dataset with precision, recall and mAP_0.5 is 0.998, 0.995 and 0.995 respectively was the best performance for YOLOv5s model. Each dataset train generate the proposed model which child detector model in best.pt.

Table 5 - Performance results of child detector model

Dataset	Precision	Recall	mAP_0.5
COCO128	0.983	0.992	0.995
COCO	0.996	0.976	0.994
Pascal VOC	0.992	0.983	0.994
INRIA Person	0.998	0.995	0.995

Table 5 show the performance results of child detector model for each dataset. The proposed model child detector model maintains the same result as YOLOv5s for COCO, Pascal VOC and INRIA Person dataset except for COCO128 had improve the result by achieve precision 0.983, recall 0.992 and mAP 0.995.

5. Conclusion and Future Works

This study aims to detect children in restricted areas using an object detection model generated from YOLOv5. Four datasets were trained, validated and tested with the YOLOv5s model. During training, the proposed detector

object model was designed, which is the child detector model. Again, the same dataset is trained, validated and tested using the child detector model. The performance evaluation results show that the child detector model is able to achieve performance comparable to YOLOv5s for the COCO, Pascal VOC and INRIA Person datasets. However, the child detector model outperformed YOLOv5s with the COCO128 dataset. In the future, this study may be directed to the extension of the child tracking model to track missing children and crimes.

Acknowledgement

This research was supported by Universiti Tun Hussein Onn Malaysia (UTHM) and Politeknik Balik Pulau (PBU).

References

[1] A. González-Briones, G. Villarrubia, J. F. de Paz, and J. M. Corchado, "A multi-agent system for the classification of gender and age from images," *Computer Vision and Image Understanding*, vol. 172, pp. 98-106, Jul. 2018, doi: 10.1016/j.cviu.2018.01.012.

[2] O. F. Ince, M. E. Yildirim, J. S. Park, J. Song, and B. W. Yoon, "Video based adult and child classification by using body proportion," *2015 The 5th International Workshop on Computer Science and Engineering*, pp. 220-223, 2015, doi: 10.18178/wcse.2015.04.036.

[3] M. Agarwal and S. Jain, "Image Classification for Underage Detection in Restricted Public Zone," *Proceedings of the 8th International Advance Computing Conference, IACC 2018*, pp. 355-359, 2018, doi: 10.1109/IADCC.2018.8692093.

[4] O. F. Ince, I. F. Ince, J. S. Park, J.-K. Song, and B.-W. Yoon, "Child and adult classification using biometric features based on video analytics," *ICIC International*, vol. 8, no. 5, pp. 819-825, 2017, doi: 10.5281/zenodo.890713.

[5] V. Kamble and M. Dale, "Face recognition of children using AI classification approaches," in *2021 International Conference on Emerging Smart Computing and Informatics, ESCI 2021*, Mar. 2021, pp. 248-251. doi: 10.1109/ESCI50559.2021.9396891.

[6] R. Lahiri, M. Kumar, S. Bishop, and S. Narayanan, "Learning Domain Invariant Representations for Child-Adult Classification from Speech," *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, vol. 2020-May, pp. 6749-6753, 2020, doi: 10.1109/ICASSP40776.2020.9054276.

[7] A. Dehghan, E. G. Ortiz, G. Shu, and S. Z. Masood, "DAGER: Deep Age, Gender and Emotion Recognition Using Convolutional Neural Network," Feb. 2017, [Online]. Available: http://arxiv.org/abs/1702.04280

[8] A. Bansal, K. Mehta, and S. Arora, "Face recognition using PCA and LDA algorithm," in *Proceedings - 2012 2nd International Conference on Advanced Computing and Communication Technologies, ACCT 2012*, 2012, pp. 251-254. doi: 10.1109/ACCT.2012.52.

[9] R. Kumar and D. KumarAgarwal, "A Review on Age Group Classification using Facial Features," *International Journal of Engineering Research & Technology (IJERT)*, vol. 7, no. 05, pp. 528-531, 2018.

[10] L. Du, H. Hu, and Y. Wu, "Cycle Age-Adversarial Model Based on Identity Preserving Network and Transfer Learning for Cross-Age Face Recognition," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 2241-2252, 2020, doi: 10.1109/TIFS.2019.2960585.

[11] S. Lim, "Estimation of gender and age using CNN-based face recognition algorithm," *International Journal of Advanced Smart Convergence*, vol. 9, no. 2, pp. 203-211, 2020, doi: 10.7236/IJASC.2020.9.2.203.

[12] A. Venugopal, Y. O. Yadukrishnan, and R. N. Nair, "A SVM based Gender Classification from Children Facial Images using Local Binary and Non-Binary Descriptors," in *Proceedings of the 4th International Conference on Computing Methodologies and Communication, ICCMC 2020*, Mar. 2020, pp. 631-634. doi: 10.1109/ICCMC48092.2020.ICCMC-000117.

[13] A. Das, A. Dantcheva, and F. Bremond, "Mitigating bias in gender, age and ethnicity classification: A multi-task convolution neural network approach," *ECCV 2018*, vol. 11129 LNCS, pp. 573-585, 2018, doi: 10.1007/978-3-030-11009-3_35.

[14] R. Kumar and D. KumarAgarwal, "A Review on Age Group Classification using Facial Features," vol. 7, no. 05, pp. 528-531, 2018.

[15] S. Nagpal, M. Singh, M. Vatsa, R. Singh, and A. Noore, "Expression classification in children using mean supervised deep boltzmann machine," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Jun. 2019, vol. 2019-June, pp. 236-245. doi: 10.1109/CVPRW.2019.00033.

[16] C. A. Reyes-García, E. Morales-Vargas, H. Peregrina-Barreto, and C. Manfredi, "Discrimination between children and adult faces using body and head ratio and geometric features," *Models and Analysis of Vocal Emissions for Biomedical Applications - 11th International Workshop, MAVEBA 2019*, vol. 5846, pp. 253-256, 2019.

[17] Surya Remanan, "Beginner's Guide to Object Detection Algorithms," *Analytics Vidhya*, Apr. 28, 2019. https://medium.com/analytics-vidhya/beginners-guide-to-object-detection-algorithms-6620fb31c375 (accessed Feb. 01, 2023).

[18] M. Tian and Z. Liao, "Research on Flower Image Classification Method Based on YOLOv5," *J Phys Conf Ser*, vol. 2024, no. 1, p. 012022, Sep. 2021, doi: 10.1088/1742-6596/2024/1/012022.

[19] G. Yang *et al.*, "Face Mask Recognition System with YOLOV5 Based on Image Recognition," in *2020 IEEE 6th International Conference on Computer and Communications, ICCC 2020*, Dec. 2020, pp. 1398-1404. doi: 10.1109/ICCC51575.2020.9345042.

[20] A. Mohiyuddin *et al.*, "Breast Tumor Detection and Classification in Mammogram Images Using Modified YOLOv5 Network," *Comput Math Methods Med*, vol. 2022, 2022, doi: 10.1155/2022/1359019.

[21] Fahad Majeed *et al.*, "Investigating the efficiency of deep learning based security system in a real-time environment using YOLOv5," *Sustainable Energy Technologies and Assessments*, vol. 53, p. 102603, 2022, doi: 10.1016/j.seta.2022.102603Get.