Knife Detection using YOLOv5: A Deep Learning Approach

Huynh Phuoc Truong Sinh
sinhhptse173032@fpt.edu.vn
Ho Chi Minh, VietNam

Dang Quoc Thang
thangdqse170249@fpt.edu.vn
Ho Chi Minh, Viet Nam

Nguyen Duc Hien
hienndse173053@fpt.edu.vn
Ho Chi Minh, Viet Nam

Tran Nguyen Huu Phuc
phuctnhse171658@fpt.edu.vn
Ho Chi Minh, Viet Nam

Nguyen Quang Vinh Nguyen
nguyennqvse173698@fpt.edu.vn
Ho Chi Minh, Viet Nam

ABSTRACT

Automatic object detection inside photos and videos has grown in importance in the field of computer vision as their application in the current era increases. In this report, we present a technique for detecting knives in pictures and videos that makes use of the YOLOv5 model. In order to maximize performance and speed, the YOLOv5 model was created based on the YOLO (You Only Look Once) architecture. The study offers details on the backbone, neck, and head structures of YOLOv5, as well as the model's training procedure and performance assessment. Additionally, we stress the need to choose particular layers within the model's head to ensure alignment with our particular goal, this involves finding knives. The study discusses the benefits and drawbacks of the YOLOv5 model for object recognition, specifically for knife detection, and makes recommendations for future work to further enhance performance and lower false positives.

CCS CONCEPTS

• **Computing methodologies** → **Object detection**.

KEYWORDS

Knife Detection, Object Detection, YOLOv5, Computer Vision, Deep Learning

ACM Reference Format:

Huynh Phuoc Truong Sinh, Dang Quoc Thang, Nguyen Duc Hien, Tran Nguyen Huu Phuc, and Nguyen Quang Vinh Nguyen. 2018. Knife Detection using YOLOv5: A Deep Learning Approach. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (Conference acronym 'XX)*. ACM, New York, NY, USA, 6 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

In the field of computer vision, particularly in the broader domain of artificial intelligence, the application of computer vision technology for object detection in images and videos has evolved into an indispensable component across numerous real-world applications. Specifically, in the case of knife detection, it holds substantial potential for critical applications in areas such as security, safety inspections, and data management. This report centers on the introduction of the YOLOv5 model (You Only Look Once), an advanced and notable model within the realm of computer vision for object detection. YOLOv5 represents a substantial advancement over its predecessor, known for its outstanding performance and faster processing speed.

Within this introduction, we elucidate the structure of YOLOv5, encompassing the following core components: Backbone, Neck, and Head. Furthermore, we delve into a crucial decision regarding the utilization of only the P3 and P4 layers within the "Head" section for research purposes, particularly in the context of knife detection. We elaborate on the testing and evaluation processes of the model on knife datasets, which include the incorporation of evaluation metrics such as precision, recall, mAP (mean Average Precision).

In the subsequent sections of this report, we will expound further on the training process, the attained results, and an analysis of false positives. Additionally, we will present directions for potential future development aimed at augmenting the performance and real-world applicability of the YOLOv5 model in the realm of knife detection.

2 RELATED WORK

We will present current approaches to the challenge of weapon identification in this section. Several researchers, using diverse strategies and methodologies, have made substantial contributions to this topic. These major studies are summarized below.

This paper [4] introduces a novel application of Active Appearance Models (AAMs) for detecting knives in images. Unlike its common use in face segmentation and medical image analysis, AAMs can identify the presence of a knife using a characteristic point typical of knives. The objective is to create a robust visual knife detection system for security applications. The study focuses on the automatic detection of knives in images, which is crucial for security personnel due to restrictions on carrying knives in public places. This software-based knife detection idea can be applied in public surveillance using closed-circuit television (CCTV). An alarm is triggered when a knife is detected, allowing human operators to focus on the scene for immediate confirmation or rejection. Automation can significantly assist CCTV operators in dealing with multiple video feeds for extended periods. Another potential application is the computer-aided analysis of luggage X-ray scans, making it a novel research area in visual detection for security applications. The paper introduces a new research area of knife detection and uses AAMs for true object detection, not just object localization. The primary goal is to answer whether a knife exists

Conference acronym 'XX, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

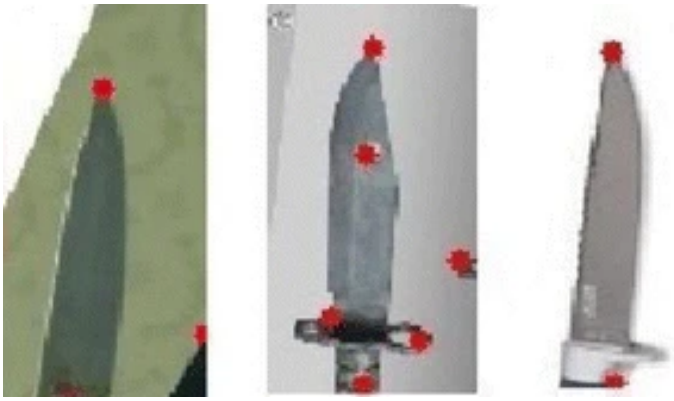

Figure 1: Interest points in knife images detected by the Harris corner detector

in a given image using AAMs as the detection method. It is based on the assumption that if a knife exists in the analyzed image, its tip will be designated as a corner by the Harris corner detector. All the designated points will be used to initialize AAMs trained to locate knives. The results of running the Harris corner detector on knife images have been presented in Fig. 1. We can see that the tip of the knife is highly likely to be designated as a corner.

This paper [16] introduces a new application of the Viola-Jones object detection framework implemented in OpenCV. Specifically, it focuses on the detection of dangerous tools, particularly knives, using Haar cascades, a method that has not been explored extensively in existing literature. The primary goal of this article is to highlight the potential advantages of this approach and assess the feasibility of further research in this field. The paper's main focus is on knife detection due to the prevalence of knife-related incidents. The objective is to provide the capability to detect potentially dangerous situations in real-life environments, such as when a person equipped with a knife poses a threat to others. While Haar cascades are commonly employed in face detection techniques, this paper proposes an alternative approach for object detection. It suggests that cascades can be trained using a suitable set of images, making it possible to use these cascades to detect objects similar to those present in the training dataset. Consequently, separate cascades can be prepared for detecting various objects of interest.

This paper [14] presents a review of various algorithms used in detecting handguns and knives. The detection of handguns and knives algorithms is classified into two major categories namely Non-deep learning and Deep learning algorithms. Non-deep algorithms are heavily depending on the quality of the image. Noise and occlusion impact the algorithms used for edge detection and color segmentation. Hence, they are suitable for images like X-ray and Terahertz. One of the major problems with all non-deep learning algorithms and some deep learning algorithms for handgun and knife detection is the use of different custom datasets. It makes the comparison of results unreliable as they do not share the same dataset. Some deep learning algorithms use using Imagenet dataset and provide accuracy and performance reports. Taking the Imagenet dataset as a benchmark and the available results, Faster RCNN has the best speed of 0.2 frames per second whereas the Overfeat algorithm has shown a more accurate result of 89% mAP.

Olmos et al [8]. developed a deep learning method for firearm recognition, achieving a 91.43% accuracy on a 6000-image dataset. However, they encountered occasional misclassifications, associating guns with similar objects like smartphones, knives, and pocketbooks. Verma et al. [13] developed an automated gun identification system using Convolutional Neural Networks for crowded scenes, using transfer learning, Deep Convolutional Networks, and a Faster R-CNN model. The system achieved an accuracy of 89.9% using Support Vector Machines. (SVM)

Luvizon et al. [7] developed a multitasking framework to improve the identification of movements and activities in a scene by incorporating human posture. This approach estimates 2D and 3D positions from still images and recognizes human actions from video sequences. While body position information is commonly used for activity and gesture recognition, its application to weapon detection has been limited.

Elmir et al. [1]developed a multi-stage framework for image collection, motion analysis, and weapons detection using Faster R-CNN, MobileNet, and CNN models. They trained with 9261 images, 3000 images for region tasks, and 608 for detection and classification. Their models achieved varying accuracy levels, with CNN at 55%, Faster R-CNN at 80%, and MobileNet at 90%. Lai et al. [5]utilized a TensorFlow-based Overfeat3 framework for classification, detection, and localization, achieving 89% test and 93% training accuracy in surveillance videos, films, and homemade movies.

Darknet YOLO, a cutting-edge convolutional neural network-based object recognition system, was introduced [11]. The Darknet framework, an open-source neural network platform written in CUDA and C, was used to create YOLO. Unlike standard classifiers that employ sliding window techniques or selective search, YOLO uses a single-pass approach to detection and does not repeat a classifier. It analyses the picture just once, segmenting it into subregions for detection, and achieves detection speeds that are up to a hundred times quicker than Fast R-CNN.[11]

Warsi et al. [15] suggested a real-time video surveillance system for visually identifying the presence of a weapon. The proposed method leverages the YOLO-V3 algorithm and compares the number of false-positive and false-negative predictions achieved with the Faster RCNN algorithm. They improved the findings by collecting guns from all possible angles and merging them with the ImageNet dataset. The dataset was trained and evaluated using the YOLO-V3 model. They used four different movies to validate the findings of YOLO-V3 against Faster RCNN. With an assessed F1 score of 75%, the detector fared poorly in identifying handguns in scenarios with diverse shapes, sizes, and rotations.

Before 2012, earlier efforts to construct and test strong knife detectors depended on conventional computer vision approaches, before deep learning began to dominate image analysis competitions and attain expert-level results. Computer vision technologies included edge detection, picture skeletonization, thresholding, and other techniques for pre-processing characteristics that might relate form, size, or posture to a judgment regarding danger intents. Zywick et al. (2011), for example, used Haar filters (as popularized for face feature identification) for multi-scale detection. The scientists highlighted in their results that the algorithm worked slowly (3 s) in

iterations that recognized knives in only 45% of true cases and 84% of negative cases without actual knives in the image.Their success in compiling a large, labeled dataset (>10,000 instances) prompted an updated review utilizing deep learning approaches. Some detection methods have now been adapted to certain regions of the spectrum (X-ray, visible, black and white), cameras (video, noisy CCTV), and motion tracking. With the most recent end-to-end detection systems that can infer both likely classifications, detection localizations, segmentation of multiple connected objects (hand and knife), and finally position or pose to infer contextual information like threatening intent, marshaling the correct datasets for size and diversity remains a significant challenge.

Since 2012, researchers have created a variety of object recognition algorithms and architectures, including the R-CNN and various versions [2, 3, 12] . In 2016, Joseph Redmon [9] suggested the "YOLO" (You Only Look Once) strategy. In contrast to traditional region-based algorithms, YOLO is a one-stage technique that uses FCNN to conduct a single pass over the image, making it comparatively efficient in comparison to rivals. YOLOv2 [10] uses batch normalization and higher resolution classifiers to tackle the problem of significantly decreased accuracy owing to imprecise localization. YOLOv3 [11] was released with incremental improvements. In 2020, YOLOv4, which stands for Optimal Speed and Accuracy of Object Detection, was released. Faster-RCNN and R-FCN are two-stage object detection networks, as is a Region Proposal Network (RPN). This type of network detects objects slowly. As a result, a single-stage object detection network, comparable to YOLOv3 and Single Shot Detector SSD, is proposed. A single forward CNN is used to identify the object's location and class [6].

Using YOLOv5, the most recent version of the YOLO (You Only Look Once) technology, provides numerous critical advantages over its predecessors, indicating a considerable advancement in object recognition and detection. YOLOv5 does object recognition in a single pass across the full image, making it very efficient and astonishingly quick. Furthermore, it improves precision, which improves object localization. The lightweight model size of YOLOv5 decreases computational resource needs, making it suited for deployment on resource-constrained devices. It supports a wide range of applications, everything from individuals and animals to sophisticated objects and cars. With an active and thriving development community, YOLOv5 continues to improve, ensuring users have access to the most recent updates and bug fixes. These benefits have propelled YOLOv5 to the forefront of object recognition and detection, providing a seamless balance of performance and computing economy.

3 METHOD

3.1 YOLOv5

In this section, we present detailed information about the YOLOv5 model architecture, one of the most famous models in the object detection domain. The YOLOv5 architecture is developed based on YOLO (You Only Look Once) and has been improved to achieve higher performance along with consistency and faster speed.

3.1.1 Overview Structure. YOLOv5 consists of three main components: backbone, neck, and head. The backbone is used to extract features from the input image. The neck transforms these features into multi-resolution representations, and the head is used to perform object detections.

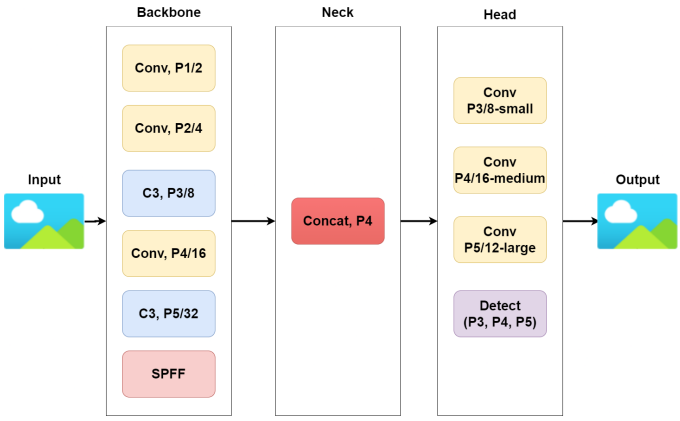

Figure 2: Structure of YOLOv5s

3.1.2 Backbone. The YOLOv5 backbone consists of multiple convolutional layers and C3 (3x3 convolutional) layers to extract features from the image.

An important point is that the backbone doesn't generate just one feature layer; it creates feature layers at multiple different resolutions, represented as P1, P2, P3, P4, and P5.

- `P1/2`: The P1/2 layer has the highest resolution and is used to detect large objects in the image. This is essential to ensuring accurate detection of large objects.
- `P2/4`: The P2/4 layer has a lower resolution than P1/2 and is used to detect objects of medium size in the image. This helps the model focus on detecting objects of medium and large sizes.
- `P3/8`: The P3/8 layer has a lower resolution than P2/4 and is used to detect small-sized objects in the image. This helps the model focus on detecting objects of small to medium sizes.
- `P4/16`: The P4/16 layer has a lower resolution than P3/8 and is used to detect very small objects in the image. This helps the model focus on detecting very small and small-sized objects.
- `P5/32`: The P5/32 layer has the lowest resolution and is used to detect extremely small objects in the image. This helps the model focus on detecting objects of extremely small and very small sizes.

3.1.3 Head. The head part of YOLOv5 is used to make predictions about the positions and classes of objects in the image. In the default model, the head uses the P3, P4, and P5 layers for predictions.

However, in this research, we made a significant decision to use only the P3 and P4 layers in the head.

3.2 Decision only using P3 and P4 on head

The decision to use only the P3 and P4 layers in the YOLOv5 head is driven by two main considerations:

- **Relevance to the Specific Task:** Our primary objective is knife detection. For this task, we are primarily interested in detecting objects of medium and small sizes. The P3 and P4 layers are well-suited for detecting such objects, as they offer a balance between resolution and object size. These layers allow the model to focus on detecting objects of sizes that are most relevant to our task.
- **Efficiency:** By using only the P3 and P4 layers in the head, we can significantly reduce the computational burden of the model. This reduction in complexity not only saves valuable training time but also makes the model more efficient during inference. The exclusion of the P5 layer further streamlines the model, aligning it with the specific requirements of our task.

4 EXPERIMENT

4.1 Data

The data used for this research was obtained from https://github.com/ari-dasci/OD-WeaponDetection/tree/master/Knife_detection, comprising 2078 images of knives and 2,155 class knife. The dataset includes images of varying sizes, with the average image dimensions close to 1280x720 pixels, making it suitable for training a robust object detection model. The dataset was divided into two subsets: 70% for training, 30% for validation. Each image was labeled to indicate the presence of a knife object.

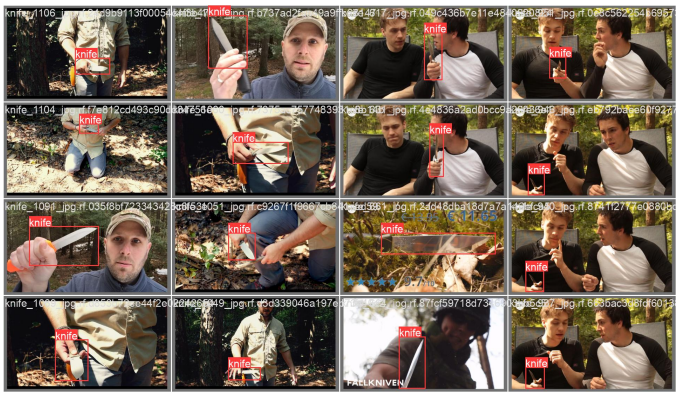

Figure 3: Examples about labeled data

4.2 Model Parameters

During the training process, we set the batch size to 32 and trained the model for 85 epochs. The model's learning rate and other hyperparameters were set to their default values.

4.3 Training and Testing Process

The model was trained on the training dataset with ground truth labels for knife objects. We used the validation set to monitor the model's performance and prevent overfitting.

4.4 Evaluation Metrics

To assess the performance of the model comprehensively, we used a range of evaluation metrics, including precision (P), recall (R), mean average precision (mAP). The specific formulas are as follows:

Precision (P) measures the proportion of true positive predictions (TP) in relation to the total positive predictions. It assesses the model's ability to avoid false positives, indicating how well it maintains accuracy when classifying objects as knives. The precision formula is defined as:

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

Recall (R), also known as sensitivity or true positive rate, quantifies the proportion of true positive predictions in relation to the total actual positive instances. It gauges the model's ability to capture all instances of knives present in the dataset, minimizing false negatives. The recall formula is given by:

$$R = \frac{TP}{TP + FN} \tag{2}$$

Mean Average Precision (mAP) is an aggregate metric that provides an overall assessment of the model's performance across multiple categories or confidence thresholds. It quantifies the precision-recall trade-off and considers how well the model performs at different levels of confidence. The mAP formula involves integrating precision and recall values across a range of confidence thresholds, and then taking the mean over all categories (n):

$$mAP = \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{1} P(R)\, dR \tag{3}$$

4.5 Results

We will begin by showcasing the overall results of our model through each epoch during the training process. Figures 4 and 5 depict the graphs of loss and some key metrics, including precision, recall, and mean average precision (mAP).

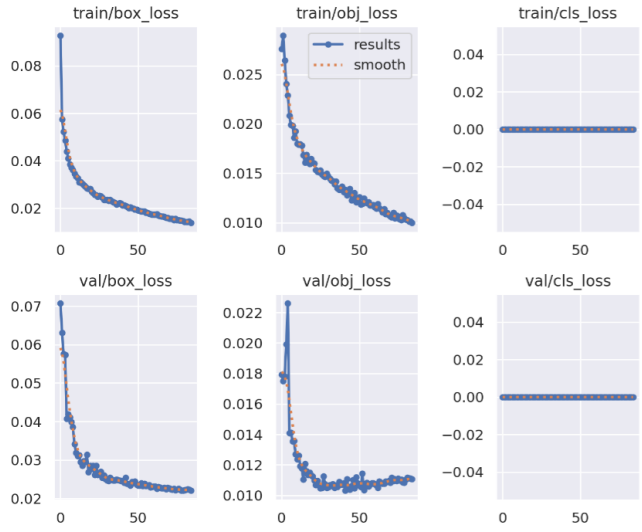

Figure 4: Loss

Figure 4 illustrates the gradual decrease in loss over each epoch. We observe that the total loss continuously decreases and stabilizes across the initial epochs, especially from epoch 0 to epoch 20. Subsequently, the total loss continues to decrease slightly but remains stable.

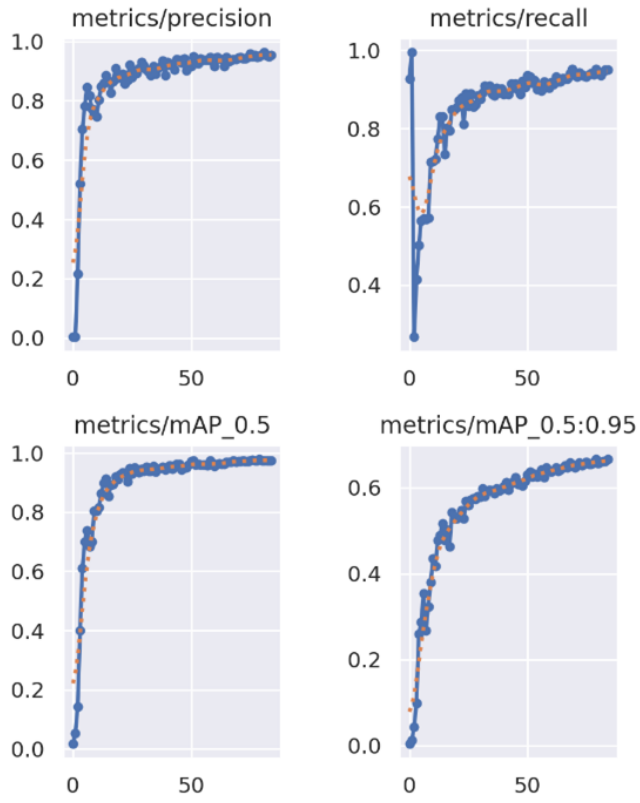

Figure 5: Precision, Recall, mAP

Figure 5 presents the necessary metrics, including precision, recall, and mAP over each epoch. In general, precision and recall both achieve quite high and stable values. This shows that our model achieves good performance in object detection.

4.6 Analysis of False Positives

When we tested on a dataset consisting of 3,194 images without knife objects, the model demonstrated varying numbers of false positives (FP) at different confidence thresholds. This test evaluates the model's ability to control and quantify false positive detections, which is crucial for its real-world application.

In Figure 6, we illustrate the model's performance in controlling the percentage of false positives at various confidence thresholds. The model demonstrates a remarkable ability to distinguish knives from other objects when operating at high confidence levels. This suggests that the model has learned how to accurately classify objects as knives.

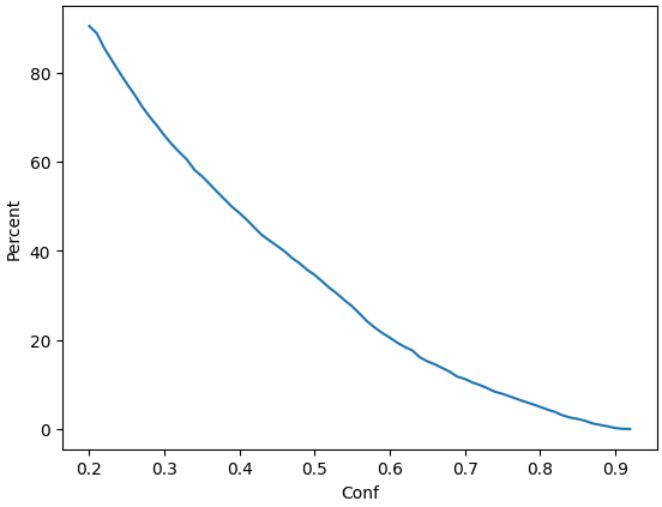

Figure 6: Percentage of False Positives at Different Confidence Thresholds

4.7 Performance Comparison

We conducted a performance evaluation of our custom model compared to the YOLOv5s architecture. The table below summarizes key metrics and model specifications.

Model	Layers	Parameters	Precision	Recall	mAP50	Size (MB)
Our Model	140	5,232,324	0.942	0.948	0.976	10.59
YOLOv5s	157	7,012,822	0.977	0.94	0.976	13.8

Table 1: Performance Comparison between Our Model and YOLOv5s.

As shown in the table, our model, despite having fewer layers and parameters, our model obtains high precision, recall, and mAP50 scores while keeping the model size minimal. This implies that our model is efficient in terms of computational resources as well as model storage.

These results demonstrate our model's usefulness in the setting of knife detection, demonstrating its capacity to attain comparable performance to a widely used architecture such as YOLOv5s.

5 DISCUSSION

5.1 Advantages

In the area of knife detection, our study showed some major advantages of the YOLOv5s model.

- `High Recall and Precision Scores`: The model's high recall rate and precision score demonstrate its accuracy in recognizing knives. This achievement is especially useful in situations where safety and security are critical, since it reduces the possibility of missing sharp items.
- `Designed for small to medium-sized items`: By selectively exploiting the P3 and P4 layers inside the "Head," our model shows an exceptional capacity in the identification

of small to medium-sized objects, which is critical in knife detection. This purposeful emphasis on task-relevant object sizes considerably improves the model's accuracy and applicability.
- `Lighter Model and Faster Training`: The architectural changes in our model not only boost its performance but also make it more efficient. The simplified model not only processes faster but also has a lower computational overhead, making it a great alternative for real-time applications.
- `Applicability`: While our research focuses on knife identification, the insights acquired from this work can be extended to a variety of item detection tasks, broadening the model's applicability to a broader range of domains.

5.2 Limitations

Despite the obvious benefits, our research reveals some limits that must be addressed for continued improvement.

- `False Positives at Low Confidence Thresholds`: The possibility of false positives in the absence of knife items is a problem, particularly at lower confidence levels. This problem highlights the necessity for continual improvements to reduce false positives and improve model precision, especially in difficult settings.
- `Future Research and Refinement`: Additional research and development are required to fine-tune the model and reduce false positives. This includes investigating advanced post-processing techniques and maybe incorporating more contextual information to improve the model's decision-making process.
- `Expanded Dataset and Real-World Deployment`: Increasing the dataset's diversity to include a broader range of situations and environmental variables will be a vital step toward further improving the model's performance. Furthermore, testing and implementing the model in real-world applications will be required to demonstrate its practical value and dependability.

In conclusion, our study represents a promising step forward in the field of item detection, with a particular emphasis on knife detection. While there have been numerous advantages proven, we recognize the problems posed by false positives and the need for continued development. As we move forward, we are committed to refining our model, experimenting with new approaches, and eventually assuring its beneficial contribution to real-world security and safety applications.

6 CONCLUSION AND FUTURE WORK

6.1 Summary

In conclusion, this paper has presented a study on knife detection using the YOLOv5s model in the context of Computer Vision. The model has demonstrated high accuracy in detecting knives, making it a promising solution for this specific task. The architectural adjustments have not only improved accuracy but also contributed to a lighter model that trains more rapidly, enhancing its practical utility in real-world applications.

6.2 Future Works

Future work includes further refining the model to minimize false positives in knife detection, exploring advanced post-processing techniques, and expanding the dataset to enhance performance. Additionally, investigating real-world applications and deploying the model in practical scenarios will be essential steps in demonstrating its utility.

7 ACKNOWLEDGEMENT

We would like to express our gratitude to the lecturer and students of the CPV301 class for their invaluable guidance and support throughout the process of writing this paper.

We also extend our appreciation to those who provided data support, which was instrumental in conducting this research.

REFERENCES

[1] Youssef Elmir, Sid Ahmed Laouar, and Larbi Hamdaoui. Deep learning for automatic detection of handguns in video sequences. In *JERI*, 2019.
[2] Ross B. Girshick. Fast r-cnn. 2015.
[3] Ross B. Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 580–587, 2013.
[4] Andrzej Glowacz, Marcin Kmieć, and Andrzej Dziech. Visual detection of knives in security applications using active appearance models. *Multimedia Tools and Applications*, 74(12):4253–4267, jun 2013.
[5] Justin Lai and Sydney Maples. Developing a real-time gun detection classifier. *Course: CS231n, Stanford University*, 2017.
[6] Pu Li and Wangda Zhao. Image fire detection algorithms based on convolutional neural networks. *Case Studies in Thermal Engineering*, 19:100625, 2020.
[7] Diogo C. Luvizon, David Picard, and Hedi Tabia. 2d/3d pose estimation and action recognition using multitask deep learning. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, jun 2018.
[8] Roberto Olmos, Siham Tabik, and Francisco Herrera. Automatic handgun detection alarm in videos using deep learning. *Neurocomputing*, 275:66–72, jan 2018.
[9] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.
[10] Joseph Redmon and Ali Farhadi. Yolo9000: Better, faster, stronger. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6517–6525, 2016.
[11] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.
[12] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 2015.
[13] Gyanendra K. Verma and Anamika Dhillon. A handheld gun detection using faster r-CNN deep learning. In *Proceedings of the 7th International Conference on Computer and Communication Technology*. ACM, nov 2017.
[14] Arif Warsi, Munaisyah Abdullah, Mohd Nizam Husen, and Muhammad Yahya. Automatic handgun and knife detection algorithms: A review. In *2020 14th International Conference on Ubiquitous Information Management and Communication (IMCOM)*. IEEE, jan 2020.
[15] Arif Warsi, Munaisyah Abdullah, Mohd Nizam Husen, Muhammad Yahya, Sheroz Khan, and Nasreen Jawaid. Gun detection system using yolov3. In *2019 IEEE International Conference on Smart Instrumentation, Measurement and Application (ICSIMA)*, pages 1–4. IEEE, 2019.
[16] Marek Żywicki, Andrzej Matiolański, Tomasz M Orzechowski, and Andrzej Dziech. Knife detection as a subset of object detection approach based on haar cascades. In *Proceedings of 11th International Conference "Pattern recognition and information processing*, pages 139–142, 2011.