# Bi-Model Helmet Wearing Detection

Yongtai Yang(✉)

Computer Science, Wilfrid Laurier University, Waterloo, ON N2L0J2, Canada
Yangyongtai924@gmail.com

**Abstract.** The conflict between workers' personal will and the construction side's safety needs has existed for many years, until the idea of using algorithms to detect if workers are all wearing helmets came to fruition. This paper will use the combination of two mainstream computer visualization tools, Pifpaf and Yolo, to detect if the helmet is on the worker's head. Benefit from the existing powerful recognition tools, all the workers showing on the same screen can be detected altogether, which largely increases the efficiency. After combining two detection models, the precision is 13% higher than just using a single model, and the correctness rate of using the combination of yolo and pifpaf is more than 90%, with the processing speed unchanged.

**Keywords:** CV · Pifpaf · Yolov5 · data combination

## 1 Introduction

Civil engineering labor is facing the dangerous environment in the industry, as they are working in an environment that has so many unexpected safety risks. According to the research, millions of workers were injured during work every year in the period from 2012 to 2019. The number of non-fatal accidents has not decreased for almost a decade [1]. Among those cases of non-fatal accidents, head injuries are more dangerous than injuries to other parts of the body and cause a more serious sequence as well. The head injury in fatal accident cases occupies up to 30%, while it only has 7% in non-fatal cases, which means the head is more vulnerable than other parts of the human body.

To ensure every worker who is in the construction area wears a helmet, many people try multiple methods. Agnes Kelm in 2013 introduced a Radio Frequency Identification tag to check if everyone in the working area has safety protection equipment with them [2]. However, it has a drawback in that this tag cannot ensure the safety equipment is being used, not only being carried along. In 2015, Park firstly introduced the vision-based helmet and human body detection, and combine it with an on-site camera [3]. The vision-based algorithm was improved in 2018. Fang used a deep learning algorithm to detect helmets and the human body in real-time [4].

The mainstream method of helmet detection is to modify existing computer vision algorithms, such as YOLO. Alternatively, they can create their own frame recognition algorithm. One of the drawbacks is that a single model can only focus on one inspection, which depends on the strategy the model uses. Thus, noisy data in the real situation


G. Ali et al. (Eds.): ISEMSS 2022, ASSEHR 687, pp. 1947–1953, 2022.
https://doi.org/10.2991/978-2-494069-31-2_229

may largely affect the precision and performance of the model. Also, some models are built from scratch and have not been verified by a large number of datasets, so they have limitations in the general situation of use. So here is a question: why cannot the combination of models be used, and let each model do what it is good at? In the content following, the author will use the combination of two well-tested and mature models, YOLOv5 and Pifpaf, to detect helmets and the human body (head) separately, and find the relationship between the helmet and the head to justify if the helmet is on the head.

## 2 Related Work

Fang et al. have introduced a framework that uses deep learning to detect workers' performance. The framework has three components: video extraction, worker competency judgment, and trade recognition [5]. Fang et al. proposed an algorithm for detecting safety harnesses. They combined Faster R-CNN with their own CNN to detect workers and check if the safety harness is worn [6]. In specific of helmet detection [7], Fang et al. introduced Faster R-CNN [5]and also listed the results of different construction situations of visual conditions that will affect the performance of the model. Mneymneh et al. used the multi-staged method. In the first step, the model extracts the human in the video given and detects the head from the upper part of the body [8]. Liu et al. provided a dataset containing 3174 images with the benchmark, which can be used by YOLOv3 based PPE detection [9, 10]. However, they share some common drawbacks, such as being able to only work in certain conditions and that unexpected situations from the environment will affect their performance. Chen and Demachi for example, hips and shoulders are needed to calculate the distance from the helmet to the neck. Guo et al. are even worse in terms of flexibility because the distance threshold is set to be unchanged [11].

## 3 Bi-Model Helmet Worn Detection

### 3.1 Solution

The current solutions for helmet-wearing detection can be classified into two types: (1) extract the human body and helmet from the input frame, and determine if the helmet is worn in the following stages. (2) Separate the workers who wear helmets and those who do not wear helmets into two classes.

Both types have some flaws. For the first type, how to find the relationship between the helmet and the head becomes the key problem to solve. The solution based on the distance between two bounding boxes [11] does not always work in every case, as it depends on the worker's gesture being in an expected range.

The second type is bothered by inter-class similarity problems. Two classes, the worker with a helmet and the worker without a helmet, share too many common features, as they are all classified as human. It is well known that distinguishing the details of two subcategories is hard, so the author is not going to challenge it. Instead, the author is interested in the location of a human's head and the location of the helmet, because there are many well-tested and open-sourced computer vision tools that can make the job

done precisely and efficiently. This paper uses the Pifpaf, a tool that can recognize human poses, as head detection. The Pifpaf will return the location of each key point's location of human body parts, and this paper only uses the head point. Combining with the helmet location that YOLOv5 returns to, the relationship between the head and helmet can be compared. If the head point is in the area of the helmet bounding box, it can be said that the worker is wearing the helmet.

### 3.2 Implementation

The proposed solution is based on PIFPAF, which the Part Intensity Field is used to capture each parts of human body and a Part Association Field is used to find association between each parts and make prediction of human gesture [12]. Also, EfficientNet-lite, MixNet, GhostNet and MobileNetV3 are used to recognize the location of the helmet.

| Algorithm 1 | |
|---|---|
| Input: | |
| | Frames from video |
| Output: | |
| | Labeled image indicating helmet wearing status of each person |
| 1. | for each frame in the input frames: |
| 2. | Input frame to yolo to get the helmet location |
| 3. | Input frame to pifpaf to get the head location |
| 4. | For each head_location in yolo output: |
| 5. | if head_location matched helmet location in the same frame: |
| 6. | Label the person on that location as safe |
| 7. | Else: |
| 8. | Label the person on that location as unsafe |

### 3.3 Architecture

The two models are implemented independently of each other, and the output of the two models in the final stage is aggregated to make the prediction. Figure 1 demonstrates the Pifpaf architecture. It is a shared ResNet with two heads. One head, PIF (Part Intensity Field), is responsible for extracting the precise locations of each joint. The other head, PAF(Part Association Field) predicts the association of each joint. Figure 2 demonstrates YOLOv4 object detector architecture. The input are image, image pyramid, and patches. This model uses VGG16, ResNet-50, SpineNet, EfficientNet-B0/B7, CSPResNeXt50, and CSPDarknet53 as backbones. There are two blocks in the neck part: Additional blocks and Path-aggregation blocks. Additional blocks involves SPP, ASPP, RFB, SAM, while Path-aggregation blocks involves FPN, PAN, NAS-FPN, Fully-connected FPN, BiFPN, ASFF, SFAM.

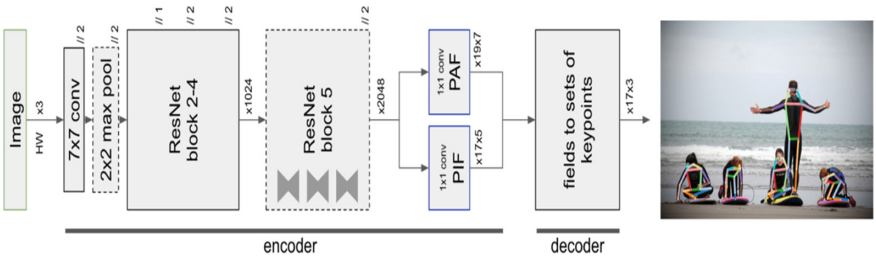


**Fig. 1.** Pifpaf Model

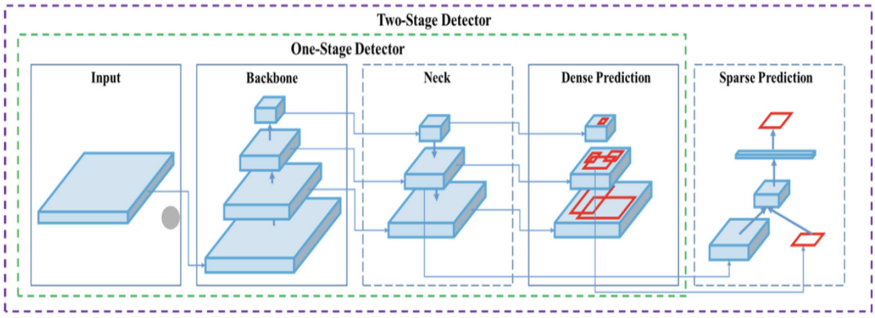


**Fig. 2.** YOLOv4 Architecture (Object detector) [13]

An image of height H and weight W will be inputted and divided into R,G,B channels. Then pass the processed data into the neural network with stride of two, and produce two fields of PIF($17 \times 15$ channels) and PAF($19 \times 7$ channels). The PIF and PAF fields will be converted into 17 joints of x and y coordinates by the decoder, also provide each confidential score of each joint.

### 3.4 Dataset

An open sourced dataset, Safety Helmet Wearing-Dataset(SHWD) is used for training and testing the model. This dataset provides the dataset used for both safety helmet wearing and human head detection. It includes 7581 images with 9044 human safety helmet wearing objects(positive) and 111514 normal head objects(not wearing or negative). Some of the negative objects obtained from SCUT-HEAD. Some bugs are fixed from the original SCUT-HEAD and the data can be directly loaded as normal Pascal VOC format [14].

## 4 Training and Result

The test set contains 334 pictures, and all of the pictures are correctly labeled as the image shows above (Fig. 3).


**Fig. 3.** One frame of output video

## 5 Discussion

During the research, many solutions of helmet wearing detection are founded. In comparison with the solution mentioned above, the advantage of the solution proposed in this article is higher accuracy. This solution can adapt to more human gestures and more complex environmental situations, which can be used in many different working environments. However, this solution has some flaws, such as a long running time. This model has a large potential to be improved. The first reason is that the accuracy can be higher if the dataset is larger. Also, the merging of two models is not smooth enough, which wastes some of the computational capacity. In the future, better solutions will be developed based on those two directions of improvement.

## 6 Conclusion

The solution stated in this paper points out the issue observed in the current solutions for helmet detection in the literature. The solution proposed involves two well developed and open-sourced models. One is for detecting the head, and the other is for detecting the helmet. This means that the proposed solution does not counter the inter-class problem, which many detection models suffer from. More importantly, this solution is much more flexible. The solution allows the human detection structure and helmet detection structure to be changed when the better detection model is introduced. Also, the data imputed is more concise than the current solutions, as it only focuses on the head location instead of considering the neck and shoulders, and there is no distance threshold. All these attributes make the model adaptable to more situations of use while maintaining high precision and efficiency.

In the testing, the solution proposed has higher precision than using only YOLOv5 in detecting if a human is wearing a helmet or not. Benefiting from the PIFPAF model, the

head location is more accurate than the YOLO of the bounding box to locate the head, and that makes the program have simpler implementation logic and the same processing time as PIFPAF. The model introduced in this paper still has some drawbacks. The major drawback is the high calculation intensity. Each individual model in the combination needs to use large computer capacity and the integration of two models is also capacity causing. This problem can be solved by optimizing the data transition from one model to another model, and pruning the models themselves to reduce the calculations. In the future, research will be mainly focused on increasing efficiency and reducing resource costs.

## References


1. United States Bureau of Labor Statistics, 2020. National census of fatal occupational injuries in 2019. https://www.bls.gov/news.release/cfoi.nr0.htm.
2. Kelm, A., Laußat, L., Meins-Becker, A., Platz, D., Khazaee, M., Costin, A., Helmus, M., Teizer, J., Mobile passive radio frequency identification (RFID) portal for automated and rapid control of personal protective equipment (PPE) on construction sites. Automation in Construction 36, 2013. pp. 38–52. https://doi.org/10.1016/j.autcon.2013.08.00
3. Park, M.W., Elsafty, N., Zhu, Z., Hardhat-wearing detection for enhancing on-site safety of construction workers. Journal of Construction Engineering and Management 141, 2015. 04015024. https://doi.org/10.1061/(ASCE)CO. 1943- 7862.0000974.
4. Fang, Q., Li, H., Luo, X., Ding, L., Luo, H., Rose, T.M., An, W., Detecting non-hardhat use by a deep learning method from far-field surveillance videos. Automation in Construction 85, 2018a. pp. 1–9. https://doi.org/10.1016/j.autcon.2017.09.018.
5. Fang, Q., Li, H., Luo, X., Ding, L., Rose, T.M., An, W., Yu, Y., A deep learning-based method for detecting non-certified work on construction sites. Advanced Engineering Informatics 35, 2018b. pp. 56–68. https://doi.org/10.1016/j.aei.2018.01.001.
6. Fang, W., Ding, L., Luo, H., Love, P.E., Falls from heights: A computer vision-based approach for safety harness detection. Automation in Construction 91, 2018c. pp.53–61. https://doi.org/10.1016/j.autcon.2018.02.018.
7. Ren, S., He, K., Girshick, R., Sun, J., Faster r-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence 39, 2017. pp.1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031.
8. Mneymneh, B.E., Abbas, M., Khoury, H., Vision-based framework for intelligent monitoring of hardhat wearing on construction sites. Journal of Computing in Civil Engineering 33, 2019. 04018066. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000813.
9. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C., SSD: Single shot multibox detector, in: Computer Vision – ECCV 2016, 2016. pp. 21–37. https://doi.org/10.1007/978-3-319-46448-0_2
10. Nath, N.D., Behzadan, A.H., Paal, S.G., Deep learning for site safety: Real-time detection of personal protective equipment. Automation in Construction 112, 103085. 2020. https://doi.org/10.1016/j.autcon.2020.103085.
11. Guo, S., Li, D., Wang, Z., Zhou, X., Safety helmet detection method based on Faster R-CNN, in: Sun, X., Wang, J., Bertino, E. (Eds.), ICAIS 2020: Artificial Intelligence and Security, 2020. pp. 423–434. https://doi.org/10.1007/978-981-15-8086-4_40
12. Kreiss, S., Bertoni, L., & Alahi, A. PifPaf: Composite Fields for Human Pose Estimation. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 11969-11978.

13. Alexey Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao. YOLOv4: Optimal Speed and Accuracy of Object Detection. 2020. https://doi.org/10.48550/arXiv.2004.10934
14. Dataset: https://github.com/njvisionpower/Safety-Helmet-Wearing-Dataset