International Symposium on Applied Geoinformatics (ISAG2021)

VISION-BASED PROVISION OF OCCUPATIONAL SAFETY CONTROL FOR ELECTRICAL WORKERS BY DETECTING HARD-HATS AND GLOVES WITH YOLO ARCHITECTURES

Feyza Kocaturk[1], M. Taha Aydin[1], Ozan Metin[2], Hatice Sude Unal[2], Dilara Oz[2], Bulent Bayram[2], Tolga Bakirman[2]

[1]AYVOS Information Technologies, Istanbul, Turkey
(feyza.kocaturk/mtaha.aydin@ayvos.com); **ORCID 0000-0001-6066-0981, ORCID 0000-0002-2084-6340**
[2]Yildiz Technical University, Faculty of Civil Engineering, Department of Geomatics Engineering, Istanbul, Turkey
(l4617008/l4618090/l4619042@std.yildiz.edu.tr; bayram/bakirman@yildiz.edu.tr); **ORCID 0000-0002-4248-116X, ORCID 0000-0001-7828-9666**

Introduction

Electricity distribution companies are responsible for the safety of the personnel they employ, as well as the regions they are responsible for electricity distribution. The absence of adequate security measures can lead to various material and moral damages. In the USA, there are work and safety standards for workers determined by the Occupational Safety and Health Administration (OSHA). According to OSHA, the energy sector is in very dangerous class according to the hazard classification made among the existing business lines (OSHA, 2014). The physical and mental factors that may affect workers are also a source of moral damages. The electricity workers can be protected from impact and penetration hazards, electrical shock and burn hazards, by wearing hardhats and wearing insulated gloves (Park et al., 2015). Even if they have been previously taught and instructed, personnel may fail to follow the rules exactly due to exhaustion, distractions, or carelessness (Green et al., 2021). These seemingly tiny details enable workers to survive minor accidents undamaged, and to avoid major accidents with minimal damage without loss of life. In this context, controlling the safety equipment of the workers at the work site and automating the process constitute a solution to prevent possible negative results.

There are inspectors at work sites for workers' safety inspections. However, manual control of security measures with manpower will increase the workforce and bring negligence. The fact that the control process is independent of human observation will minimize the workforce. Considering the relevant studies on the subject, it is observed that basically two different methods are used to reach the conclusion namely sensor-based and vision-based methods. Dong et al. developed a sensor-based method that generates an alarm if the worker does not wear a hardhat while entering a dangerous area within the work area with a pressure-affected silicon sensor (Dong et al., 2015). Kelm et al. used a passive sensor with Radio Frequency Identification method to identify personal protective equipment (Kelm et al., 2013). Sensor-based methods are usually based on the data of a pressure sensor mounted inside the hardhat, which creates a condition that will disturb the worker in hot and humid weather and there may be false stimulation in the pressure sensor with any effect from outside the hardhat (Zhang et al., 2019). Thus, sensor-based methods are much more costly than vision-based methods. In vision-based methods, Rubaiyat et al. presented a solution with the combination of the Histogram of Oriented Gradient (HOG) method and the Circle Hough Transform (CHT) feature extraction method, which is based on color difference for human detection and then hardhat detection (Rubaiyat et al., 2016). Li et al. has studied background extraction from images, full-body and hardhat detection using computer vision, machine learning and image processing to perform safety surveillance of workers working in electrical substations (Li et al., 2017). Qi et al. studied with Faster-RCNN a method that detects the use of hardhats of remote workers over videos and observed that the results of the Faster-RCNN model are more accurate than YOLO in detecting distant objects, although the YOLO architecture will give faster results (Qi et al., 2018).

Classifiers have been repurposed to perform detection in previous work on object detection. Rather, we consider object detection to be a regression issue with spatially separated bounding boxes and associated class probabilities. In a single assessment, a single neural network predicts bounding boxes and class probabilities directly from entire images. Because the entire detection pipeline is a single network, it can be optimized end-to-end directly on detection performance (Redmon et al., 2016). In our research, we present the vision-based object detection using 4 different versions of the YOLO (v3, v4, v4x-Mish, v4-Tiny) artificial neural network architecture on the detection of hardhats and voltage resistant gloves, which are necessary equipment to ensure the personal safety of the workers working in electrical distribution works.

Materials and Methods

The dataset was created from the frames in ⅕ of a second from the videos collected with 1080P resolution Full HD digital IP cameras installed in the electricity distribution regions. The videos were collected in different locations, times and conditions in order to increase the diversity of the data set and to give the model optimum results in all conditions. The prepared dataset consists of 3260 images with geometric resolution of 1920 x 1080 pixels. The images in the dataset were saved in YOLO format by drawing rectangular bounding boxes on hardhats and gloves with the labelImg tool (Tzutalin, 2015). Figure 1 shows a sample labeled images. 316 images (10%) from the labeled dataset were reserved for testing the trained model and the dataset was separated into 76% train and 14% validation.

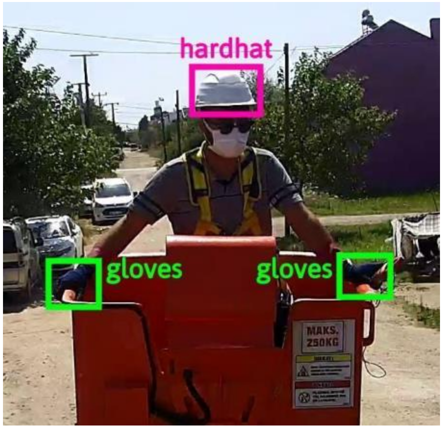

Figure1 Labeled sample image of gloves and hardhat worn by electrical worker

YOLO versions we compared in our research are:

- YOLOv3: Yolov3 Network is a hybrid approach between the network used in YOLOv2, Darknet-19, and that newfangled residual network stuff. This network uses successive 3x3 and 1x1 convolutional layers but now has some shortcut connections as well and is significantly larger. It has 53 convolutional layers so we call it Darknet-53 (Redmon et al., 2018). Additionally, this approach provides 106 layers of fully convolutional architecture using 53 more layers for detection.
- YOLOv4: YOLOv4's architecture consists of CSPDarknet53 as a backbone, SPP additional blocks, PANet path-aggregation neck and YOLOv3 head also compared to YOLOv3, it increased 10% in Average Precision (AP) and 12% in Frame Per Second (FPS) (Bochkovskiy et al., 2020).
- YOLOv4-Tiny: The biggest difference between YOLOv-Tiny and YOLOv4 is the network size. The number of convolutional layers in the CSP backbone has been compressed, the number of YOLO layers has been reduced from three to two, and there are fewer anchor boxes for prediction (Jiang et al., 2020).
- YOLOv4x-Mish: Leaky, Linear and Mish Activation functions are used for activation in YOLOv4 (Bochkovskiy et al., 2020), however fully Mish activation function is used in YOLOv4x-Mish As a result of the research using SqueezeNet in the CIFAR-10 dataset, it is seen that the Mish activation function gives better results than the Leaky activation function (Misra, 2019). YOLOv4x-Mish has more convolutions than YOLOv4 and has a deep network.

Results and Discussion

The system features of the dataset training are given below:

Table 1 The system features

Operating System:	Ubuntu 18.04.5 LTS
Processor:	Intel® Core™ i5-10400F CPU @ 2.90GHz × 12
Graphics Card:	Geforce RTX 3070
RAM:	32 GB

Anchors were found as 10, 13, 12, 20, 18, 21, 14, 32, 21, 28, 25, 36, 25, 46, 49, 81, 98,171 in the calculations made before the training to be used in the training of all YOLO versions. Hyperparameters, preprocess parameters and label counts used for training are given below:

Table 2 The parameters we used in our research

Number of Classes :	2
Label Counts (Hardhat - Gloves):	3389 - 2642
Batch:	64
Subdivisions:	32
Resize:	416 x 416 pixels
Max batches(Number of classes *2):	4000
Steps(80% & 90% of max batches):	3200, 3600
Filters ((Number of Classes + 5)*3):	21

Models trained with the same parameters were tested on untrained images. The mAP metrics obtained as a result of the trainings with different YOLO versions are shown in Table 1. As a result of the evaluations, the highest metric for the hardhat class was YOLOv4x-Mish, and the highest metric for the glove class was the YOLOv4 model. The detection results of different YOLO versions on the same images are shown in Figure 2.

When the image results tested in the trained models are compared, it is seen that the hardhat class can be detected more easily than the gloves class. It is thought that the detection rate has decreased due to the fact that the colors of the gloves are dark and they are absorbed on the background.

YOLOv3 **YOLOv4** **YOLOv4x-Mish** **YOLOv4-Tiny**

Figure 2 Detection results from trained YOLO models

Table 3 mAP metrics obtained as a result of the trainings

Model	Class	mAP (IoU Threshold .50)	mAP (IoU Threshold .75)	mAP
YOLOv3	Hardhat Gloves	98.19% 49.74%	32.64% 11.03%	47.67% 19.17%
YOLOv4	Hardhat Gloves	**93.93%** **62.07%**	44.33% 18.87%	50.80% **25.19%**
YOLOv4x-Mish	Hardhat Gloves	98.91% 59.07%	**45.75%** **13.52%**	**50.92%** 21.06%
YOLOv4-Tiny	Hardhat Gloves	97.30% 52.07%	34.74% 7.68%	46.22% 17.62%

Conclusion

It is vital that workers working in electricity distribution areas use personal safety equipment like hardhat and insulated gloves. In this study, 4 different YOLO architectures were trained and the architecture with the highest detection performance was determined in the dataset used, and it was ensured that the workers use the relevant equipment or not with a vision-based method**.** In the following stages, it is aimed to develop the research in a way to identify workers with or without equipment.

Keywords: *Personal Protective Equipment (PPE), Work Safety, Object Detection, Computer Vision*

Acknowledgements

We would like to thank AYVOS Information Technologies for their support and contributions to our research with the supported TUBITAK Project 3205079 and also we would like to thank the supported TUBITAK 2209-B Project 1139B412100393 for contributions.

References

Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020). Yolov4: Optimal Speed and accuracy of object detection. arXiv.org. Retrieved October 7, 2021, from https://arxiv.org/abs/2004.10934

Dong, S., He, Q., Li, H., & Yin, Q. (2015). Automated PPE Misuse Identification and Assessment for Safety Performance Enhancement. ICCREM 2015.

Fang, Q., Li, H., Luo, X., Ding, L., Luo, H., Rose, T. M., & An, W. (2018). Detecting non-hardhat-use by a deep learning method from far-field surveillance videos. Automation in Construction, 85, 1–9.

Jiang, Z., Zhao, L., Li, S., & Jia, Y. (2020). Real-time object detection method based on improved Yolov4-Tiny. arXiv.org. Retrieved October 7, 2021, from https://arxiv.org/abs/2011.04244.

Kelm A., Laußat L, Meins-Becker A, Platz D, Khazaee M, Costin A, Helmus M, Teizer J (2013). Mobile passive Radio Frequency Identification (RFID) portal for automated and rapid control of Personal Protective Equipment (PPE) on construction sites, Automation in Construction, Volume 36, pp. 38-52.

Li, J., Liu, H., Wang, T., Jiang, M., Wang, S., Li, K., & Zhao, X. (2017). Safety helmet wearing detection based on image processing and machine learning. 2017 Ninth International Conference on Advanced Computational Intelligence (ICACI).

Misra, D. (2019). Mish: A self-regularized non-monotonic neural activation function. arXiv.org. Retrieved October 7, 2021, from https://arxiv.org/abs/1908.08681v2.

OSHA. Commonly Used Statistics | Occupational Safety and Health Administration. (n.d.). Retrieved October 7, 2021, from https://www.osha.gov/data/commonstats.

Park, M.-W., Elsafty, N., & Zhu, Z. (2015). Hardhat-Wearing Detection for Enhancing On-Site Safety of Construction Workers. Journal of Construction Engineering and Management, 141(9)

Real-time proactive safety in construction. POWER Mag. Retrieved October 3, 2021, from https://www.powermag.com/real-time-proactive-safety-in-construction/ (2012, January 1).

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. arXiv.org. Retrieved October 7, 2021, from https://arxiv.org/abs/1506.02640.

Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv.org. Retrieved October 7, 2021, from https://arxiv.org/abs/1804.02767.

Rubaiyat, A. H. M., Toma, T. T., Kalantari-Khandani, M., Rahman, S. A., Chen, L., Ye, Y., & Pan, C. S. (2016). Automatic Detection of Helmet Uses for Construction Safety. 2016 IEEE/WIC/ACM International Conference on Web Intelligence Workshops (WIW).

Tzutalin. (n.d.). Tzutalin/labelimg, Git Code (2015): LabelImg is a graphical image annotation tool and label object bounding boxes in images. GitHub. Retrieved October 7, 2021, from https://github.com/tzutalin/labelImg.

Zhang, H., Yan, X., Li, H., Jin, R., & Fu, H. (2019). Real-Time Alarming, Monitoring, and Locating for Non-Hard-Hat Use in Construction. Journal of Construction Engineering and Management, 145(3)